Drawing Visual Query Images

Use, Users and Usability of Query by Drawing Interfaces for

Content Based Image Retrieval Systems

Lars-Jacob Hove

Dissertation for the degree philosophiae doctor (PhD)
at the University of Bergen

2010

Acknowledgements

I would like to thank my advisor Associate Professor Joan C. Nordbotten for her invaluable support throughout the work on this thesis. Without her encouragement, advice and belief in me this work would not have been completed. In addition to being an enthusiastic and involved advisor, she has provided invaluable support and motivation during the last stages of this project when

Tor Kristian Bjelland became my second advisor during the project, giving invaluable input. Our discussions helped me shape and refine the thesis.

Linn Katrine Erstad provided me excellent transcriptions of most of the interviews performed in the project. She provided quick and accurate transcriptions of the material.

I am also very grateful to the staff at the Bergen Academy of the Arts, particularly Associate Professor Ashley Booth. They provided me with both access to their students and facilities for performing part of my project. Without their assistance it would have been much more difficult to recruit relevant participants for the study.

Aleksander Krzywinski, Ingerid Rødseth, Magnhild Viste, Nina Svane-Mikkelsen, Sonja Wilhelmsen, Vibeke Vold and all my other fellow PhD students and other colleagues have provided me with an interesting academic and social environment for my work, and provided me with numerous encouraging conversations and lunches.

I would also like to thank the members of the CAIM project, particularly Svein Nordbotten and my fellow research fellow Christian Hartvedt, for all the interesting discussions and valuable input during the work.

I am very grateful to my friends and my family for always supporting and believing in me, particularly when my work got interrupted during the final stages of the project. Your support motivated me to finish this work. I am particularly grateful to my parents who have always done everything they could to support me.

At last, I would like to thank the person who is dearest to me, Anne. Without your understanding, support and endurance during the entire project, I could not have gone through this process. I look forward to being fully present in our lives again.

Abstract

Query by Drawing (QBD) is an approach to Content Based Image Retrieval (CBIR) systems where users express their image needs by drawing an image representative of the images they wish to retrieve. CBIR is based on comparison of the query image and images in an image collection. This approach to image retrieval has been an active field of research for over a decade, but despite this, few end-user applications are available. An often quoted reason for this is that CBIR systems are capable of retrieving images based on low-level image structures such as colours, textures and shapes, while users are primarily interested in the *semantic* content of the image.

The role of the user in image retrieval systems is a relatively unexplored area, and little empirical data has been collected on the expectations, needs and behaviour of these users. Literature in the field suggests that image retrieval based on low-level image structures is not very important for users, and consequently current CBIR systems may not be very useful for end-users.

The main motivation behind this research project has been to collect and analyze empirical data on the use and users of QBD CBIR systems. Four major goals were defined for the project:

- Understand how users *behave* when using QBD CBIR systems
- Understand how users *experience* using QBD CBIR systems
- Determine if QBD CBIR systems can be a *useful tool* for end users despite the current challenges related to these systems
- Identify potential improvements that can be made to QBD CBIR Systems

30 respondents were asked to perform a set of image retrieval tasks in two different QBD CBIR systems. The respondents represented two different groups of users. The first group represented “non-professional” users, and consisted of 17 information science students. The second group represented “professional” users, and consisted of 14 respondents with a background in visual arts, visual design and industrial design. The two QBD CBIR systems represented two different approaches to the QBD CBIR process. They were selected as representative systems based on an analysis of 59 past and current CBIR systems.

The respondents performed a total of 414 queries. The queries and the query sessions were analyzed using three different approaches:

- A protocol analysis of the QBD query process based on observation and interface videos

- A grounded-theory based approach based on questionnaires, structured interviews, the interface videos and observation
- An analysis of the query images drawn by the respondents based on a custom framework created for QBD query images

The evaluation indicated that the respondents preferred to keep the query drawings as simple as possible. They wanted to quickly sketch the query images using freehand drawing, and to limit the amount of details to the level they felt that they needed in order to express their image requests. They often created these drawings as *visual keywords*, i.e. very simple representations of the objects they wanted to retrieve images of.

The "non-professional" respondents found the drawing process difficult and challenging. They were frustrated that they were not able to draw the objects in a realistic manner, and felt that they would not be able to fully benefit from the QBD CBIR approach because of this. These respondents also felt that the time required creating QBD CBIR queries was a major obstacle, particularly when compared to creating text based queries. The "professional" respondents were positive towards the QBD CBIR process, and did not experience similar problems related to the drawing process, but they were not *willing* to spend time drawing realistic query images.

The "professional" respondents believed that they would use QBD CBIR systems on a regular basis if such systems were available and could be used on large scale image collections. They described several realistic scenarios where they would have benefited from using QBD CBIR over normal text based retrieval systems. The "non-professional" users were not so sure that they would use these systems for anything other than entertainment.

Based on the feedback from the respondents and the evaluation of the QBD CBIR process, a set of prioritized improvements to QBD CBIR systems have been identified. A four-step process for leveraging QBD CBIR systems from research prototypes to full-scale systems that can be of real benefit for real-world users is suggested.

These results indicate that the role of QBD CBIR systems may have been understated in literature. Even with the current challenges facing these systems, the feedback from the respondents in this study indicates that, given some changes, users may find QBD CBIR systems a very useful tool, particularly when combined with text based queries.

Table of Contents

Table of Tables

Table of Figures

1 Challenges Drawing Visual Queries

The digital computer and the World Wide Web have radically changed the way we store, manage, retrieve and use images. The ability to digitalize images has put almost infinite amounts of images at our fingertips. While our ancestors only had access to the images present in the scrolls, books or engravings in their immediate surroundings, we now have the possibility to store large amounts of images digitally on a single optical disk. This large quantity of available images poses some problems for efficient retrieval of desired images.

A fundamental prerequisite for image retrieval is that the users are able to express their image requests in a format that a retrieval system is capable of interpreting and processing. If the users are unable to express their image information needs, or the retrieval system is unable to interpret and process the query, the retrieval process will not yield satisfying results. Consequently, the vast amounts of stored images are of little benefit to users unless they have the tools required to access them. They need to be able to search, identify and retrieve images.

The scientific disciplines of *information retrieval* and *library science* have provided efficient tools, methods and algorithms for managing, indexing and retrieving information through textual descriptions. The success of these approaches is evident in tools such as Google, which is capable of indexing a large part of the World Wide Web and allows us to search and retrieve relevant information in mere seconds. However, these techniques are primarily based on textual indexing and retrieval. The textual content of documents, web pages, books and other sources of information is analyzed, and textual descriptors such as keywords are used to create metadata which is used for indexing and retrieval. While this is highly efficient for textual information sources, transferring these approaches to complex data structures such as images present some major challenges. Text based information is structured by basic semantic units, such as letters, sentences or paragraphs. These are easily parsed by automatic software, and it is possible to automatically create indexes based on these. However, images do not have a similar easily parsed basic structure. Manually creating textual annotations of images is time consuming and prone to subjectivity, and some visual structures may be difficult to precisely describe using text.

In response to some of these challenges *Content Based Image Retrieval (CBIR)* evolved from the fields of *Computer Vision*, *Signal Processing* and *Pattern Recognition*. In CBIR, images are described using mathematical and statistical representations of their visual structures. These structures are automatically extracted from the images in a collection, and used for retrieval based on similarity comparisons.

Figure 1 presents a simplified view of a CBIR image search. A user has an image request, e.g. “I need images of dolphins and dolphin caretakers interacting in a theme park”. He expresses this image request to a CBIR system through a *visual query interface*, using one or more methods for specifying the visual characteristics of the request. The resulting *visual query* is processed by the *image retrieval system*. This system compares the visual query to images in an image collection, and presents the user with images similar to the visual query, as defined by a set of query parameters.

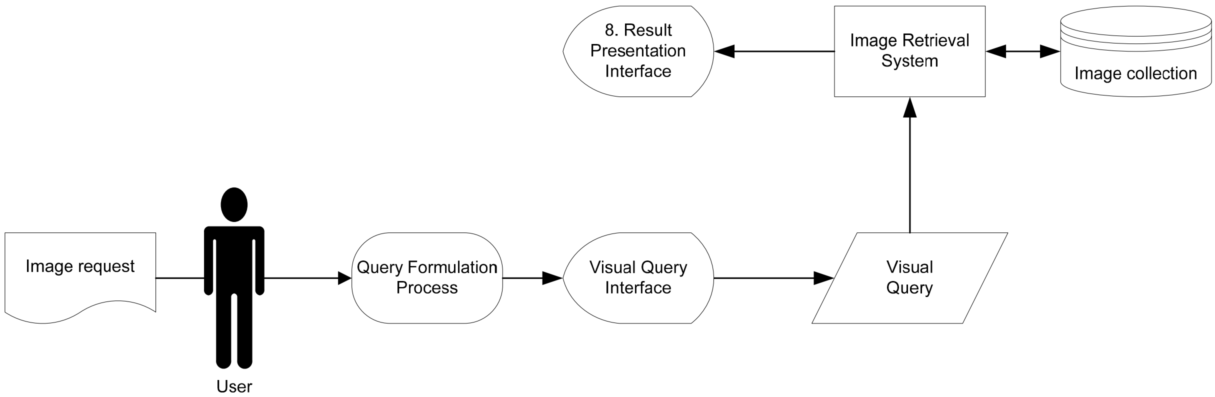


Figure 1 - Simplified view of CBIR query processing

A *Visual query* can be expressed through a number of different techniques. Common for these techniques is that they represent image requests based on *visual structures* in some manner. A *visual structure* represents the basic syntactical structure present in an image, such as the shapes, the textures, the colours and the spatial relationships between these structures. The most used techniques[1] for expressing image requests through visual structures include:

- *Drawing an image* representative of the image request
- *Presenting a representative example image* to the system by submitting a new image
- *Selecting a representative image* from the images existing in the collection
- *Presenting the system with a set of visual structures representative of the image request*, e.g. using colour histograms or texture samples

These techniques often also allow the user to *refine* their queries by *manipulating* the visual structures. Furthermore, the user is often offered the option to define *how* the retrieval should

[1] The currently available techniques for visual query formulation are discussed in detail in chapter 2.5

process the visual query by presenting *query parameters,* e.g. telling the system to focus primarily on colours and textures present in the visual query image. Based on this, we have the following definitions:

> *Visual Structures are the basic syntactical structures present in an image, such as shapes, colours, textures and the spatial relationships between these structures (Definition 1).* [2]
>
> *A Visual query is defined as a request for images based on submitting, manipulating or creating visual structures, expressed in a visual query interface (Definition 2).*
>
> *A Visual Query Interface is an interface for expressing visual queries (Definition 3).*

The different visual query specification techniques offer different options for creating visual queries, and have their own strengths and weaknesses. However, as will be discussed in detail in chapter 2.5, *Query by Drawing* (QBD) represents the approach which, theoretically, offers the user with the highest degree of freedom and flexibility when expressing their image needs to a CBIR system. *Query by drawing* represents the focal point of this work:

> *Query by Drawing is defined as expressing an image need by creating visual structures through drawing using either freehand sketching or using one or more of drawing tools (Definition 4).*

A large share of research in the field of CBIR has been aimed at the development and improvement of fast, reliable and working techniques for indexing all types of image content. The end users and their tasks, needs, requirements and expectations by contrast have received relatively little attention (Venters, Hartley et al. 2001; McDonald and Tait 2003). The currently available systems may not be very well adapted to the needs and behaviour of the human user (Lew, Sebe et al. 2006). And, as noted by Datta, Joshi et al (2008), there is a scarcity of user studies focusing on identifying scenarios in which a typical end-user might benefit from using the CBIR approach.

The work presented in this thesis represents an effort to gain a better understanding of the *expectations, experiences* and *challenges* of *users* using QBD CBIR systems by collecting empirical data on these issues.

1.1 User Challenges in Visual Image Retrieval

There are several user related challenges associated with image retrieval, particularly with regards to the visual queries. The following two scenarios highlight and describe some of these challenges.

[2] The definitions are also available in Appendix 1 - Definitions

1.1.1 The Novice User

Consider the case of a teacher preparing a lecture on dolphins and various aspects of dolphin life: anatomy, habitat, feeding habits, mating cycles, and the relationship between dolphins and humans. The teacher wishes to include images related to the subject: Dolphins, humans and other animals in various maritime environments.

The teacher has access to an image collection describing maritime life: marine mammals, fishing vessels, fishing tools and related activities. The images have been made available to the public through the internet, and the retrieval system supports text based and content based image retrieval.

He probably has a general idea of the type of images he wishes to retrieve. First of all, he wishes to find generic images containing dolphins and various aspects of dolphin life. Next, he might be interested in finding images of a well known dolphin, such as “Skippy”. Finally, he might have a detailed request, such as finding an image of a dolphin jumping out of the water, similar to Figure 2. The viewpoint of the image should be from the surface, the dolphin should be on the crescent of a jump from the left to the right of the image, with the high seas and the sky as a backdrop to the image.

Figure 2 - An image information need. Image retrieved from the VISI[3] system.

In order to retrieve such images, the teacher has to somehow express these information needs to the image retrieval system, in a manner that the system is capable of processing.

The teacher has some experience using Google, and may try to express these requests using keywords. For the generic queries (i.e. finding “Skippy” or generic images of dolphins), this might easily be expressed using keywords such as “Skippy, dolphin” or “Dolphin, feeding, surface”. If the images in the collection have been annotated with these keywords, retrieval is a trivial matter.

---

[3] http://bulmeurt.uib.no:8500/caim/Maritim/

However, if the collection lacks annotation, or the actual annotation was created for a very different purpose than general image retrieval, retrieval might be problematic. If there are no images annotated with "jumping" or "dolphin", none will be retrieved. If the image collection is large, it is unlikely that each image is annotated with enough keywords or descriptions to satisfy all possible image requests. For example, if the main motivation behind the image collection is the description of different dolphin species, the activities or photo-specific details of the image might not be given much attention in the image descriptions. In this case, the images might be annotated with the Latin names of the dolphins or similar scientific data.

The final image request may present the user with some additional challenges. While it is easy to indicate that the images should contain "dolphins", actually expressing the particular layout of the image, the angle of the shot or the pose of the dolphins using simple keywords may be difficult. The obvious choice for the teacher would be to express this in general terms, such as "jumping dolphin" and browse through the retrieved images. Now, consider the images in Figure 3, which are some of the results of a Google Image search[4], using "Jumping Dolphin" as search terms.[5]

Figure 3 – Google search results using "Jumping Dolphin"

Only the rightmost image appears to be relevant to the teacher's information need. The first image is completely irrelevant, the second image is a drawing, and the third image is obviously manipulated and does not resemble the teachers' request. The fourth image may be relevant, but contains a lot of additional details which the teacher might not be interested in.

If the teacher had used Google Images, he most certainly would have found one or more relevant images. But these might not be the *best* images available. Even if an "ideal" image is among the retrieved images, the teacher would have to manually browse through a set of roughly 2,640,000 images[6].

---

[4] http://images.google.com, August 2009

[5] While Google Images might not be considered a "Maritime image collection", it is not unlikely the results presented above might exist in such a collection.

[6] The number of images returned from the *Google Images* query using the terms "jumping dolphin". The query was performed in August 2009.

As an alternative, the teacher may use *visual queries*. However, this presents him with different challenges. The most used approach for visual query specification is to provide the system with an example image visually similar to the requested images. However, this assumes that he already has access to images which are similar to the images he wishes to retrieve. And if this was the case, his information need might already be satisfied. Alternatively, he might try to express the image request by drawing an example image, either using an appropriate paint program or directly in the visual query interface. Successfully using this approach depends on a number of different factors. First of all, creating a good drawing depends on the teacher's drawing competency. If he is not used to working visually, composing a drawing representing his image request may be a daunting task: he might not even be able to create a drawing resembling a dolphin. It is possible that the interface may assist the teacher in some way, but this is highly dependent on the usability of the interface and the tools available for composing the image. Finally, expressing the query visually may be considerably more time consuming than using keywords, and the teacher might not be willing to spend a long time creating the query.

1.1.2 The Skilled User

Next, consider the case of a designer creating a publication for an environmental organisation. She needs to find some images that can be used to illustrate the magazine's main feature article. She has a very clear notion of the layout of the pages, and has specific needs in terms of both the content of the images and their actual structure, composition and colours.

She needs an image of either a dolphin or a killer whale jumping out of the ocean in front of a whale-safari tour. She also needs a close up of a wild dolphin or a killer whale playing or entertaining or interacting with people in the animal's natural habitat. She also wishes to have an image of a whaling vessel in the process of butchering a minke whale as well as an image of a tame dolphin in an aqua park, entertaining a crowd while playing with a ball. Furthermore, the editor has requested that she includes a specific image in the magazine. Both the editor and the illustrator have seen the image before, but they are unable to recall the name of the image, who the photographer was or where it was taken.

Unlike the teacher, the designer has very developed artistic skills, formal training in image composition and is generally comfortable visually. Despite this, she faces some challenging issues when searching for images.

First of all, the designer might use a text based approach when retrieving these images. However, the visual nature of these requests suggests that a visual approach might be better: expressing the compositional structure of the desired images may be difficult using simple keywords, and while the human perceptual system is capable of quick interpretation of visual impressions, browsing through potentially thousands of images might not be an optimal approach. For example, how should she explain, using linguistic terms that she needs images containing a Cetacean jumping in front of a red boat, both located in the lower right part of the image, with clear blue skies over a near black ocean? A text-based retrieval system would require a very thorough description of every image in order to retrieve images based on these criteria. And, in the case of the particular image, she has no idea of what query terms she should use.

Next, even though the designer is capable of creating good, realistic looking images representing the images she is interested in, these images might not share any similarities with any of the images in the collection. If the retrieval system is based on a direct comparison between the query and the images in the collection, even the best made queries may fail to return any meaningful results.

Furthermore, it is quite possible that the retrieval system might retrieve images that are *structurally similar* to the query, but differ *semantically* from the expected results. Consider the two images in Figure 4. Our designer wishes to find images of a jumping dolphin, in a particular pose. She draws an image similar to the image on the left and uses this as a query.

Figure 4 - Structural similarities between a black-and-white drawing of a dolphin and an image of a banana[7].

While the drawing might be a very good representation of a jumping dolphin, the structural characteristics of the image also make it a very good representation of a banana. The overall shape, salient features, colours and overall composition between the two images are very similar. As a result, the search process might retrieve images, which are similar in structure, but semantically dissimilar from the designer's information needs.

1.1.3 Challenges of Image Retrieval Systems

The above scenarios illustrated four major challenges facing users expressing image requests to current image retrieval systems:

1. The Query Formulation Challenge

[7] Both images were retrieved from the VISI system, available at http://bulmeurt.uib.no:8500/caim/Maritim/

2. The Query Interpretation Challenge
3. The Query Mismatch Challenge
4. The Media Mismatch Challenge

Figure 5 shows a simplified overview of a query process, along with 4 challenging problems related to queries involving visual structures. Each of these challenges is discussed and related to this figure.

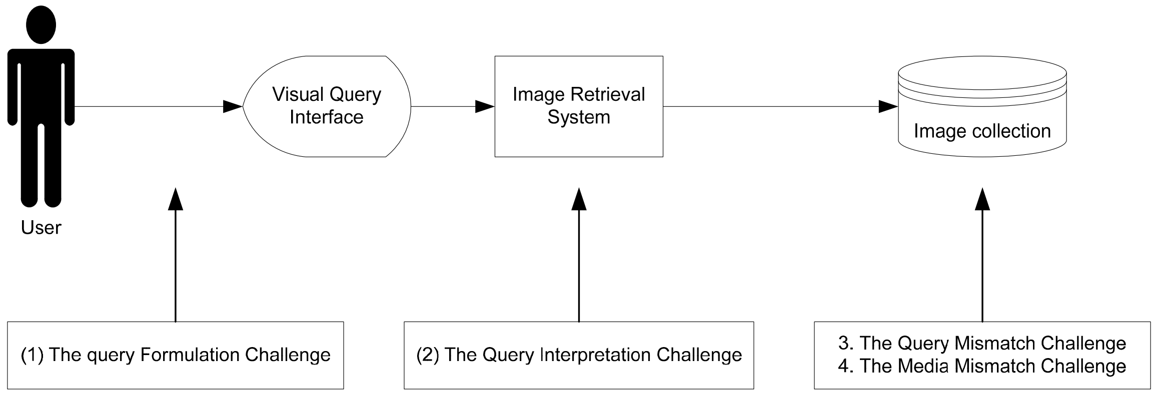


Figure 5 - Challenges of Content Based Image Retrieval

The first set of challenges is related to the user's query specification process. This is illustrated by the problems facing the teacher: Lack of skill, lack of good and reliable tools, and the time required creating these queries. The **user's lack of skills** potentially represents a significant barrier towards widespread use of visual queries: If the user lacks the skills to express visual queries, or at least feels that he or she lacks the skills, they might not be willing to use visual queries at all. This is related to the problem of a potential **lack of suitable tools for expressing visual queries.** The users will require query expression tools that will allow them to express the queries *at their own level of competence*. Finally, visual query specification may be a **time-consuming process**, particularly compared to text-based techniques, and might present a further obstacle towards widespread use of visual queries. These are all issues that may complicate the process of translating an information need into an actual query. This is called **the Query Formulation Problem (1)**.

The second challenge is related to the retrieval system's ability to interpret and process the user's queries. In the above scenarios, this is illustrated by the system's inability to find relevant images from a collection, even though the designer expressed a query which might be considered as a "good" representation of the retrieval task from a human perspective. If the system is incapable of proper segmentation and identification of the visual objects in a query image, it will fail to return relevant images even if there are relevant images in the collection. Furthermore, the actual process of creating the query may contain additional information regarding the importance of the elements in the query. The sequence the objects were drawn or the compositional structure of the query might contain indications towards the importance of the objects, or the relationships between these objects. If the visual query is simply processed and compared directly to the images in a collection based on the colours, shapes and similar structures, this additional and potentially very useful information might be lost. This is called t**he Query Interpretation Problem (2).**

The third challenge is related to the fact that current CBIR systems are primarily based on structural similarity, not semantic similarity. In the scenarios, this is illustrated by the dolphin-banana problem, and describes the fact that a visual query might not share a structural similarity to images in the collection that are semantically similar. This is called **the Query Mismatch Problem (3)**.

The final challenge refers to the problem that, when documents and queries are expressed in different media, matching is difficult, as there is an inherent inter-media mapping process that needs to reformulate the concepts expressed in the medium used for queries (e.g. text) in terms of the other medium (e.g. images). In the scenarios this is related to the designer's difficulties with expressing a very visual query in linguistic terms, e.g. finding images with a certain composition or structure. This is called **the Media Mismatch Problem (4)** (Egenhofer 1997).

Finally, while these scenarios and challenges represent real problems for users, a major problem is that there currently are relatively few (QBD) CBIR systems available to end users. Most of the systems that have been developed have been research prototypes, not fully developed end-user systems. In a real-world situation, neither the teacher nor the designer would have access to large-scale image collections supporting the (QBD) CBIR approach.

1.2 Research Project: Understanding the Query Formulation Challenge

According to Venters et al (2001) there is little evidence to support the usability of visual query formulation tools, and QBD CBIR interfaces remain one of the least researched and developed element of CBIR retrieval systems. The literature generally acknowledges that the main drawback with this approach is that it depends on the user's ability to create good example images (See for example Jaimes and Chang (2002)).

Though CBIR and QBD represent research fields that have been active for almost two decades, but there are still several unsolved challenges, particularly related to these systems' ability to provide the users with results that are semantically relevant to the visual queries. As a result, there are currently only a few CBIR systems that are available to end users.

Consequently, a main focus of this work was to study the needs, expectations, experiences and challenges of users expressing image needs to a CBIR system by drawing visual queries, with a particular focus on the *query formulation challenge*. This was done by gathering empirical data about these issues, and identifying how this material can be used to improve current systems' ability to *process* visual queries expressed by drawing. The challenges of *query interpretation*, *query mismatch* and *media mismatch* have not been directly evaluated, but the results are important factors in understanding and solving some of these other challenges.

Based on this, five major research goals were defined for this project:

1. Understand how users *behave* when expressing image requests by drawing visual image queries
2. Determine the *type of drawing* users draw when expressing image requests by drawing visual image queries
3. Understand how users *experience* expressing image requests by drawing visual image queries
4. Determine if QBD CBIR can be useful tools for end users, despite the current challenges facing these systems
5. Identify potential improvements that can be made to QBD CBIR systems

An important aspect guiding this work is the notion of *expressive convenience*. Users will usually approach an image retrieval system with one or more image information needs, and have to translate this information need into a query in the language provided by the system. While the process of drawing *visual queries* as used in this work might not qualify as a formal language, it might nevertheless be relevant to discuss this process in terms normally used for such languages. One important aspect of formal languages is that a language has a certain *expressive power*, i.e. the potential for what might be expressed using the language, regardless of how easy or hard it is to use the language.

> *The expressive power of an image query interface is defined as the type of image information requests that can be expressed using the interface (Definition* 5*).*

The expressive power represents capabilities of a given language or interface: what *can* be expressed. A complementary notion to this is *expressive convenience:* How a language or interface can be used to express a query (Trovåg 2004; Moe 2006).

> *The expressive convenience of a visual query interface is defined as the ease a user experiences when expressing a given image information request using the interface (Definition* 6*).*

While the *expressive power* and *expressive convenience* of visual queries have not been formally used as evaluation criteria in this work, they represent a fundament for the work and have guided the direction of the research.

The research goals are expressed in the following set of research questions:

- **RQ1:** How do users utilize the visual query interface when they draw visual queries?
- **RQ2:** How realistic are the query images drawn by QBD CBIR users?
- **RQ3:** What are the major challenges encountered when users draw visual queries?
- **RQ4:** How do users feel about expressing image requests by drawing visual queries?
- **RQ5:** What improvements can be made to CBIR systems in order to better support users when drawing visual query images?

The first research question focuses on understanding how the users make use of the tools available for expressing drawing visual queries. Understanding the users' use of, and actions in, the user interface may provide important insights into both how these interfaces can be improved, as well as providing clues on how these interactions might be used to assist the system in interpreting the queries. This research question is operationalized and evaluated in chapter 5.

The second research question focuses on the degree of realism in the query images the users create. Current CBIR systems are primarily based on low-level similarity functions. Successful retrieval is dependent on similarities between the query image and the relevant images in a collection. This is particularly important for the challenges of *query interpretation* and *query mismatch* challenges. Accordingly, query images created by users need to be analyzed. This research question is operationalized and evaluated in chapter 6.

The third question focused on gaining an understanding of the *query formulation problem* and identifying what the users found to be the most challenging aspects of the visual query formulation process. This concerns issues such as what the users find challenging, why it is challenging and what can be done to improve this process. Understanding these challenges is a fundamental step in order to create systems that best can support users when expressing these queries, and increase the likelihood that users will find visual queries a viable alternative to text based queries. This research question is operationalized and evaluated in chapter 7.

The fourth research question covered one of the least evaluated fields within CBIR: how users feel towards expressing image requests through visual queries. Reading through existing literature, one might get the impression that using visual queries might not be a preferred tool for the users as visual queries, as illustrated by the following quote from a peer-review process:

> *I am not surprised at all when the study indicates that users tend to draw simple iconic pictures for simple retrieval tasks. My argument is that users may not want to draw at all for simple retrieval tasks!*

Based on this, it was felt that a thorough evaluation of the opinions and feelings of a set of users using visual queries might be both relevant and interesting for researchers of image retrieval. This research question is operationalized and evaluated in chapter 8.

The final research question this project was focused at identifying which, if any, improvements actual users of visual query interfaces suggest. Having users try different interfaces might identify shortcomings in these interfaces, making it possible to identify improvements based on feedback from these users. This research question is evaluated based on the overall results and data made during the project. Chapter 9 presents an overview and discussion of the suggestions made by the *respondents* in the project, while chapter 0 presents four steps that must be followed in order to promote the current position of CBIR systems as experimental prototypes to powerful tools that may be useful for users expressing specific image requests to an image retrieval system.

An overview of the research questions and their corresponding research hypotheses can be found in Appendix 3 - Research Questions and Hypotheses. While the operationalization and evaluation of these research questions are presented in chapters 5 through 10, the questions are actually answered in section 10.1.

1.3 Methodological Approach and Overview

A user centred research approach was chosen for the project, and three separate studies were performed. Two groups of people with different backgrounds were asked to perform a set of image retrieval tasks using two different image retrieval systems. Several methods have been used to collect data in the three studies. An overview of the methods is presented in Figure 6 and fully detailed in chapter 3.

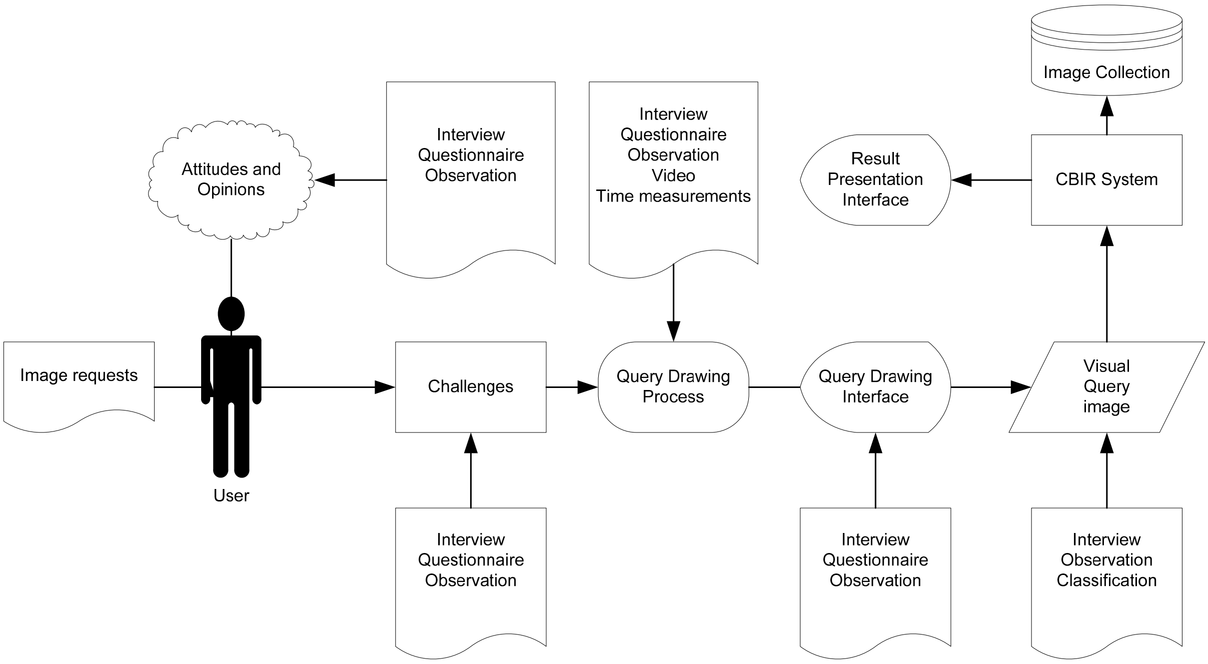


Figure 6 - Overview of the methodological approach used in this work

In the studies respondents expressed a set of image requests in two different CBIR systems supporting Query by Drawing. 30 respondents were selected from two different sources: *students of information science* and *students at the Bergen Academy of the Arts or professionals working with design or fine arts*. Each experiment session was performed in a laboratory setting, and observation, interview sessions and questionnaires were used as the primary tools for data collection. A *grounded-theory* approach was used to evaluate the data sources, and a framework for classifying the visual query images was developed specifically for this project.

1.4 Overview of the Thesis

In addition to this introduction, the thesis consists of 5 major parts.

Chapter 2 presents a theoretical discussion of images, image contents, user image requests and an overview of current techniques for visual query specification. Central elements discussed include what types of *content* can be found in digital images, what types of information needs users have when approaching an image retrieval systems, how these users can *express* these information needs as queries to the retrieval system, and a presentation of how the capabilities current CBIR systems have for processing these queries.

Chapter 3 presents an overview of the methodological approach used in the thesis.

Chapter 4 presents a framework for evaluating and categorizing visual query images. The framework was developed specifically for this work, and is based on the concept of *visual modality* (Kress and van Leeuwen 2006).

Chapters 5 through 9 present the data analysis and evaluation of the major research questions. These chapters present the major empirical data collected in this work.

Chapter 10 presents a discussion of the major results discussed in the previous chapters. This includes answering and discussing the major research questions, a discussion on the quality of the data, and a presentation of how these results can be used to improve current CBIR systems.

In addition to these chapters there are four appendices:

- Appendix 1 presents an overview of the central definitions used in this work
- Appendix 2 presents a summary of different CBIR system reviewed in chapter 2.5
- Appendix 3 presents an overview of all the research questions and research hypotheses used throughout the thesis, as well as an overview of the answers to the research hypotheses
- Appendix 4 presents the data collection tools used during the experiment sessions. These are only available in their original language (Norwegian)

2 Image Retrieval

The main objective in this work has been to gain an understanding of the experiences and challenges faced by users using query by drawing interfaces to retrieve images from general image collections.

Image retrieval has its origins in the field of *Information Retrieval*, which originally focused on text based information items. This research and development area has grown to accommodate “new” digitalized information items, such as video, sound and images. There are two main approaches to image retrieval, text-based and content-based. In text-based image retrieval (TBIR) a user query consists of semantic keywords describing aspects of the desired image(s). In content-based image retrieval (CBIR) the user submits an image example for a search for similar images. The technology behind CBIR systems has its foundations in the fields of *Image Retrieval* and *Image Processing*.

Research in image retrieval started in the 70s, when it became possible to store and process image material. Since then, both the fields of Information Retrieval and Computer Vision have driven research in the field. Today it is an active and important research area, spanning a broad range of research disciplines, such as Information Retrieval, Computer Vision and Image- and Signal Processing.

Image Processing refers to a computer discipline wherein digital images are the main data object. It covers the analysis, manipulation, storage, retrieval, and display of images from sources such as photographs, drawings and video.

The needs and experiences of the users are central to this thesis. Consequently, human interpretation and use of images have been given precedence over more computer centric approaches to images and image content, making theory concerning the nature of images and human interpretation of visual structures central to this work. This theory is primarily based in communication studies, visual culture and the humanities.

Figure 1 Figure 7 presents repeats the CBIR process from Figure 1, but presents how the different sections in this chapter relate to the elements in this process:

- Sections 2.1 and 2.2 discuss various aspects of images and image collections: What are digital images and what types of content can we expect to find in an image.
- Section 2.3 presents a high level overview of Content Based Image Retrieval systems, along with a discussion of some of the major challenges facing these systems.

- Section 2.4 presents an overview of different types of *image requests*. Why do users approach an image collection, and which types of requests can we expect these users to have?
- Section 2.5 presents an overview of the query specification techniques and interfaces offered by past and current CBIR systems, with a discussion of the strengths and limitations of the different query techniques.

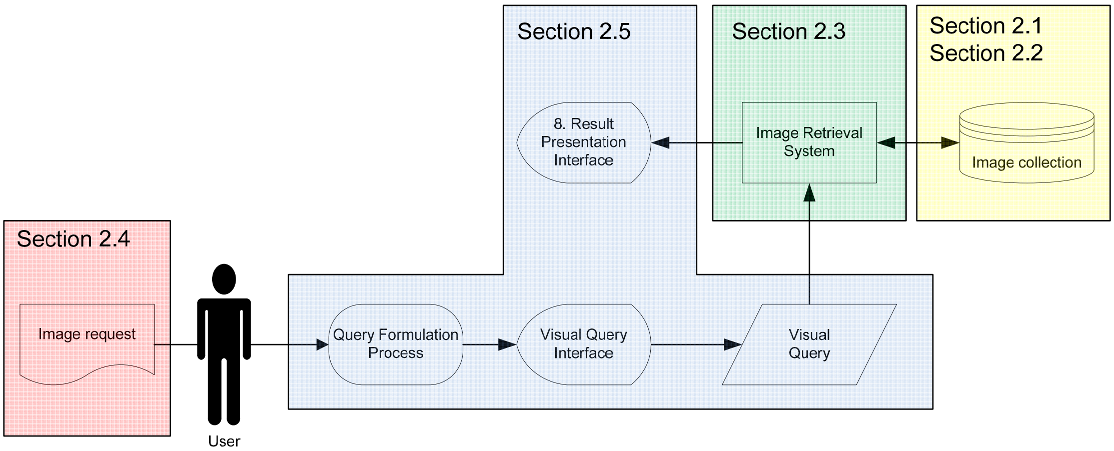


Figure 7 – Chapter 2 sections and the CBIR query process

2.1 Digital Images

The word "Image" stems from the Latin word *imago* (imitation, copy, likeness or bust). In common usage, it is an artefact that reproduces the likeness of some object, at several different levels. At the most basic level, an image represents a response to light perceived by our visual senses. At the most complex level an image represents abstract ideas dependent on the observer's knowledge, experience and mood. In everyday life, terms like *pictures, images* and *digital images* are used interchangeably to describe this concept. The general term "image" is related to several different concepts, particularly when talking about "digital images". Consider the case of an observer viewing an image on a computer screen, as illustrated in Figure 8:

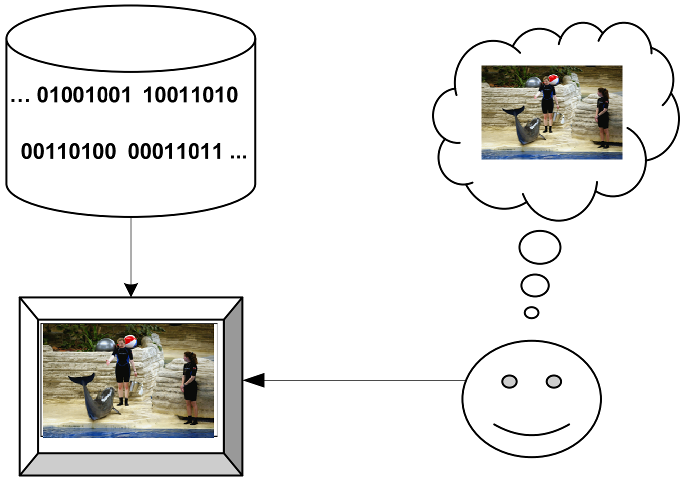


Figure 8 - Different forms of images

Figure 8 presents three different forms of images. First there is the actual visual representation of something. In this example, the observer is looking at a visual representation of a dolphin and two caretakers produced on a computer screen. This is a common understanding of the term “image”, and is called the *visual image*. This visual image is a representation of an object, scene, person or abstraction, produced on a medium.

In Figure 8, the visual image has been produced digitally on a computer monitor. This representation is not synonymous with the actual digital image stored in a computer system - it is merely a representation of it. The actual digital image is a binary file consisting of a two-dimensional array composed of pixels or pixel arrays whose locations hold data about digital colour and/or brightness which, when represented on a suitable digital medium form a visual image. A human cannot directly observe a digital image other than through a representation of the binary file. However, a computer system is capable of processing these structures in a variety of manners, e.g. projecting it on a monitor or printing it to paper.

Finally, the observer has a *mental image* of the ideas, events and objects represented in a visual image. In Figure 8, the mental image is likely to be very similar to the visual image. Mental images may also appear in a person’s brain in the form of an imagination: the act or power of forming an iconic mental representation of something not present to the senses or never before perceived in reality. Such images are only available as images to the person having the imagination, unless the person expresses this in some way or manner (Hartvedt 2008). In the scenarios presented in section 1.1, the teacher and the designer had clearly defined information needs. They probably had some notion of the image they were looking for; a mental image or an internal visualization of the type of image they were requesting. This is called the *mental image*, representing an internal visualization of an object, concept, event, scene or visual image in the mind of an individual. When observing a visual image, their *mental image* may be almost identical to the *visual image.* However, when imagining an ideal image that may represent the goal of an image request, the mental image may represent a particular image they have seen before, a pure imagination of an ideal image, or anything in between.

Throughout this work, the term *image* is used as a common denominator of these three visual concepts when discussing concepts that may be valid for all three forms. The term *image* is then defined as *all representations of objects, concepts, scenes, persons or abstraction, produced or stored on some medium (Definition 7).* The different qualifiers have been used where it has been necessary to distinguish between the different forms of an image.

2.2 Image Contents

Discussing image retrieval requires an understanding of what can be retrieved, e.g. why are images interesting and what types of content do images contain? In ordinary, everyday use, an observer is often interested in the objects or people present in an image, or in the meaning that this content represents. In fields such as cultural studies or art history, the observer might be interested in the stylistic and formal means used to create the image, the connotations that can be derived from the content, or the broader context of the image. In some technical disciplines, images are regarded as a specific form of signal[8], where the important content is defined in the *structure* of the image. An image will in most cases be *of* something. A photographic image normally depicts objects, people, landscapes or activities, while other types of images (e.g. works of art) may consist of some type of abstract content. Figure 9 shows an image of two people, a dolphin and two balls, situated in what appears to be an aqua park. Most human observers are immediately capable of interpreting the image, identifying the objects and the scenes, perhaps identifying the identity of the people or animals present, have an opinion of the activities being performed in the image, or identify the deeper meanings represented by the image.

[8] A signal is an abstract element of information, or more specific usually a flow of information, in either one or several dimensions.

Figure 9 - Image of a dolphin, a ball and two caretakers.

The image in Figure 9 contains a number of different elements at different levels of abstraction. At the lowest level, it consists of colours, shapes and textures. At higher levels of abstraction, it is possible to identify different semantic units such as the dolphin, the beach balls or the caretakers. Some observers may be able to identify and name some of these objects, e.g. the name of the dolphins or the caretakers, or the location of the image. Most observers will most likely also be able to identify and interpret abstractions such as *narratives* in the image (e.g. which actions are illustrated), understand that it is situated in an *aqua park* or even extract *meanings* from the image (e.g. that the image may symbolize how humans exploit animals for economic purposes).

Several classification schemes focusing on different aspects of image contents have been created for categorizing different types of image contents. In order to analyze the research questions presented in this work, a taxonomy of image contents was created based primarily on the works of Jaimes and Chang (2002), Eakins, Burford and Briggs (2003) and Kress and van Leeuwen (2006). The taxonomy is presented in Table 1, and was used as a theoretical basis for the research questions and hypotheses in this thesis. Five different levels of image contents are identified: Perceptual structures, generic content, specific content, narrative content and abstract content. The table presents a brief description of each level, including what knowledge or skills are required by a human observer in order to interpret the content. Each of these levels is discussed in detail in the following sections.

Table 1 - A taxonomy of different Levels of Image Content.

Content Level	Description	Interpretation	Examples
Perceptual structures	The overall syntactical structures present in an image. These are the basic perceptual elements identifiable by the human sensory system. This includes perceptual primitives, geometric primitives and simple two- and three dimensional non-representational forms.	Based on low-level perceptual systems.	Lines, colours, shapes, contours, arcs, circles and textures, as well as the local and global spatial distribution of these
Generic content	The basic semantic units. Generic objects which share a set of attributes which are common to all, or most of, the members of a particular category. This includes images without participants ("Scenery"), images with participants but without background and images with one or more objects or entities placed in a context, constituting a "scene".	Based on everyday knowledge, and is presumed to be universal	Images of single objects, i.e. "Ball", "Dolphin" and "Human" A skyline or cityscape (Generic) images of a forest The image of the dolphin and the caretakers represented in Figure 9.
Specific content	Specific content which can be uniquely identified and named. This can be both single image participants and several participants together in a scene.	Based on personal knowledge and recognition.	The dolphin "Skippy" The Empire State Building
Abstract content	Meanings that can be derived from specialized or interpretative knowledge about what objects depicted in an image *represents,* and what the image is *about*.	Based on contextual, cultural or technical knowledge of objects, motives and symbols, filtered through individual experience.	Interpretation of X-rays or medical imagery Emotions evoked in an observer of an image (Happy, Sad) Themes coved in an image (Sport, Leisure, War)
Narrative content	Actions performed by image participants, interactions between participants or conditions and states of a participant.	Based on knowledge of context and the ability to "read" and interpret the narrative structures present in the image.	A jumping dolphin A person feeding a whale A wounded shark

2.2.1 Perceptual Structures

At the most basic level, an image consists of a set of perceptual structures created by the way patterns of light are reflected on different materials, producing the perception of different elements such as texture, colours and shapes. This is what Jaimes and Chang (2002) call the *syntactical structures* of an image representing the colours, textures, shapes and the local and global distribution of these in the image.

Eakins, Burford and Briggs (2003) distinguish between three categories of perceptual structures. *Perceptual primitives* represent content extracted by low-level perceptual system, e.g. the colours and some types of textural descriptions present in an image. *Geometric primitives* represent simple two- and three-dimensional non-representational forms, e.g. lines, arcs, squares and circles. *Visual* extensions are visual features that do not contain meaning beyond the simple perceptual pattern, e.g. the detection of depth through occlusion or perspective.

2.2.2 Generic Content

The *generic content* refers to content that is not derived purely from the perceptual structures, but represents basic semantic concepts which are defined and named. Semantic concepts are generic objects which share a set of attributes which are common to members of a category. Instances within categories are defined by a set of prototypes[9], each presenting a subjective indicator of membership in the category. Some examples of this are the dolphin, the two humans and the two balls shown in Figure 9. The two first examples might require a large set of “prototypes” in order to cover the large variance of the shape, while the last example might require a smaller set of prototypes, as there are fever variances between instances of the ‘ball’ category. Identification of image content at this level is generally based on everyday knowledge.

2.2.3 Specific Semantic Content

The *specific semantic* content refers to particular instances of a concept that can be identified and named. Specific knowledge of the objects in the image is required, and interpretation relies on the factual knowledge of the observer. Examples include individual persons, such as identifying and naming the two humans or the dolphin in Figure 9 (e.g. “Anna”, “Louisa” and “Skippy”) or identifying the specific aqua park (e.g. “Dolphin World”).

2.2.4 Abstract Content

The *abstract content* refers to image meanings that can be derived from specialized or interpretative knowledge about what the depicted objects represent. The generic and specific semantic content primarily concerns what Jaimes and Chang (2002)refer to as the visual content of the image: what is directly perceived when an image is observed. The *abstract semantic content* primarily concerns information that is closely related to the image, but not present. Identification and interpretation of content is based on the observer’s knowledge of motives and symbols, and filtered through their individual cultural, technical or emotional experiences.

[9] A prototype is this context refers to an original type, form, or instance of a concept or category, serving as a typical example, basis, or standard for other members of the same category. Two examples of such prototypes are “Dolphin” and “Ball”

Eakins, Burford et al. (2003) distinguish between four different categories of abstractions. *Contextual abstractions* refer to information which is presumed to be universal, in that it is derived from knowledge of the environment. An example of this is deciding whether the scene depicted in Figure 9 is indoors or outdoors. *Cultural abstractions* are presumed to be fairly generalized within the general culture of the viewer. Examples of such abstractions may be activities performed in the image or political, cultural, historical and sporting events, or determining that the persons and dolphins depicted in Figure 9 are located in an aqua park, and are performing a show. *Technical abstractions* refer to information that requires specific technical expertise to interpret. An example of this is as x-ray images. Finally, *emotional abstractions* refer to affective or emotional associations or responses people may have to an image. Example of this is retrieval of images with a particular theme ("Love" or "War") or images which may provoke specific feelings in the observer (A "funny" image) .

2.2.5 Narrative Content

The *narrative content* of an image represents the actions performed by image participants, interactions between participants or conditions and states of a participant. Technically, this may be considered as a subcategory of the *semantic* content of an image. However, while the previous sections defined some important types of image contents and the skills and knowledge required when interpreting these concepts, the actual visual structures that assist an observer in understanding and interpreting higher level semantic warrants some additional discussion. Kress and van Leeuwen (2006) present a systematic and comprehensive account of a grammar of visual design, offering a descriptive toolkit for understanding and interpreting images. Hove (2007) presents an analysis of how this work can be used for image retrieval. Four of the most central concepts are presented here: *Participants, actors, goals* and *interaction vectors.*

Participants represent the subject matter of an image: the people, animals, objects or other elements representing the interesting elements in an image, where the emphasis is on *interesting* elements. In theory, every perceptual structure in the image might be considered a participant, even though they might not be directly relevant for interpretation of the image. Consider the scene in Figure 9. The most obvious objects in this image are the dolphin, the caretakers and the beach balls. However, the image also consists of a number of other objects, such as the plants, the two walls or the individual stones in the walls. Depending on the observer's interest, some of these objects may be more interesting than the others. Possibly the dolphin, the two caretakers and the beach balls represent the most interesting elements of this image, while the other objects provide context to the important objects. Consequently*, a participant is defined as an important visual element in an image (Definition 8)* while *a contextual element is defined as a visual element providing situational description to the narrative structures (Definition 9)*.

Participants in an image are often involved in some kind of *narrative process* such as performing an action or interacting with another participant. Kress and van Leeuwen (2006) present a variety of such processes describing various forms of narrative structures, such as *action processes, reactional processes, speech processes* and *mental and conversational processes*. However, as all these processes involve some kind of relationship between participants, they have been grouped together in this overview. Consequently, *a narrative process is defined as an interaction between two participants (Definition 10)*. Narrative processes commonly consist of an *actor, representing the active part in a narrative process (Definition 11)*, and *a goal, representing the receiving part in a narrative process (Definition 12).* Visually, a narrative process between an actor and a goal is represented through an implicit or explicit visual structure called an *interaction vector*. An interaction vector is a visual structure representing or indicating the presence of a narrative process, e.g. the outstretched arm of the leftmost caretaker in Figure 9. The interaction vectors represent the interactions, i.e. the actual action or actions performed by the two participants participating in a narrative process. This represents all potential interactions which might be performed by two or more image participants, and a single image may contain any number of such interaction vectors. Figure 10 presents an example of *participants, actors, goals* and *interaction vectors* present in the image shown in Figure 9.

Figure 10 - Aqua Park Image illustrated with narrative structures.

The important participants have already been defined as the dolphin and the two caretakers, while the stone walls and the beach balls have been identified as contextual elements. The yellow arrows represent four potential *interaction vectors*, which might indicate interesting narrative processes. First of all, a substantial part of the dolphin's body is pointing towards the leftmost caretaker. The dolphin is identified as an *actor*, the caretaker as the *goal.* This can signify that the dolphin is focused on the caretaker, and is presumably involved in some sort of interaction, or transaction, with her. Similarly, the same caretaker's outstretched hand is signalling something to the dolphin, and her eyes are looking directly at the dolphin, possibly representing another transaction - the caretaker is expecting something from the dolphin, maybe passing some sort of instructions. Finally, the rightmost caretaker is watching the pair, possibly focusing on the actions and behaviour of the other caretaker.

2.3 Content Based Image Retrieval Systems

Content-based Image Retrieval/CBIR is based on an analysis of the perceptual structures of images. When a digital image is submitted to a CBIR system, these structures are extracted and indexed, resulting in a set of statistical descriptors of the image. These descriptors are normally represented as feature vectors: a set of descriptors describing one, or more, syntactical image features, represented as numeric quantities. Several different categories of descriptors exist, from very specialized vectors created for a very narrow application domain, e.g. management of x-ray images (Engan and Fretheim 2004) to general descriptors which may be used to describe any type of image, such as *colour, shape, texture* and *spatial composition*:

- **Colour** is an important dimension of human visual perception that allows discrimination and recognition of visual information. Correspondingly, colour features have been found to be effective for indexing and searching colour images in image collections. Generally, colour descriptors are relatively easy to extract and match and well suited for content based queries. For an introduction to colour feature extraction, see for example Smith and Chang (1995).
- **The shape** of a physical object is the external form or contour, the geometry of its external surfaces or contours, the boundary between the object's interior and the exterior, representing the outline or characteristic surface configuration of the object. The shape of an object can also be said to be invariant to variances in location, scale and rotation of the object; it represents the characteristic surface configuration of the object. An introduction to the use of shapes as feature vectors can be found in Li and Kuo (2002).
- **Texture** refers to visual patterns with properties of homogeneity that do not result from the presence of only a single colour or intensity. Pictures of water, grass, a bed of flowers and so on contain good examples of image texture. Many natural and man-made objects are distinguished by their texture. Examples of texture are tree barks, clouds, water, skin and fabrics. A thorough introduction to texture feature extraction is available in Manjunath and Ma (2002).
- **Spatial composition** refers to the structural relationships between the perceptual structures in a digital image. There are two classes of these relationships. The first class, containing topological relationships, captures the relations between element boundaries. The second class, containing orientation or directional relationships, captures the relative position of elements with respect to each other. Examples of topological relationships are "near to",

"within" or "adjacent to". Examples of directional relationships are "in front of", "on the left of" and "on top of". The spatial structure of an image is dependent on other features, i.e. in order to identify *where* a component is related to another component, the *component itself* needs to be identified, through colour, texture, shape or potentially other features. An introduction to the use of spatial composition is available in Li and Kuo (2002).

The process of query and retrieval of images in CBIR systems is usually based on a notion of similarity between two or more such feature vectors, e.g. between the vectors describing a visual query and the description of the images in a collection. Similarity is determined by a *similarity function*. A large number of widely differing similarity functions for computing all kinds of similarities exist. Most these are based on mapping pairs of feature vectors to a number representing the similarity between two images. A similarity function is defined as a mapping between pairs of feature vectors and a positive, real-valued number, which is chosen to be representative of the visual similarity between two images (Li and Kuo 2002). Usually, the number represents the Euclidean distance[10] between two feature vectors. If, for a given feature, two images are identical, the similarity function should be equal to 0. In other words, the less distance there are between images, the more similar they are. For more details, see for example Datta, Joshi et al (2008).

The major challenge facing the CBIR approach is that of *the semantic gap*. Current CBIR techniques are primarily based on similarity functions comparing feature vectors extracted from the perceptual structures of an image. However, when requesting images, users are normally interested in the *semantic* contents of the image. This gap between what current CBIR systems are *capable of processing* and what users *normally request* from an image collection is called the *semantic gap.* As an illustration, consider the three images in Figure 11. A comparison between the three images based on colour features, would likely report a high degree of similarity between the first and the second image, and a low degree of similarity between the second and third image. A comparison based on shape would likely report a high degree of similarity between the second and third, and a low degree of similarity between the first and the second image. It is unlikely that a retrieval system based on feature descriptors will return all three images, even though they all depict a single dolphin.

[10] The Euclidean distance is the straight line distance between two points.

Figure 11 - Different depictions of a dolphin.

Next, consider the image presented in Figure 12. Most people would correctly identify them image as a depiction of a banana. However, looking solely at the perceptual features of the image, it has a high degree of similarity to the third image in Figure 11, i.e. both are gray-scale images, dominated by a single shape, with similar salient characteristics.

Figure 12 - A depiction of a banana.

Some CBIR algorithms operate on a *global scale*, i.e. the feature vectors are extracted from, and compared to, whole images. For whole image matches, a single feature vector is extracted from each image and used for indexing and retrieval purposes. This framework was adopted in early CBIR systems, such as IBM's QBIC (Flickner, Sawhney et al. 1995). While this might be useful for comparing global image features, such as colour distribution, it is insufficient for identification and comparison of objects within an image. An example is the image of the seagull represented in Figure 13. The seagull is easily identified by a human observer. However, the seagull is situated on a background: Grass, water and rocks. While this may be a natural setting for a seagull, the background may be irrelevant in a query for "seagulls", and the background can potentially cause retrieval problems for content based image retrieval systems, e.g. by *not* including images of seagulls on *dissimilar* backgrounds, or by including *irrelevant* objects on a *similar* background.

Figure 13 - An image of a seagull on a "noisy" background[11].

In order to separate the interesting objects from the noise, there needs to be a mechanism for *segmenting* the image into different segments. *Segmentation* is the process by which an image is divided into spatial sub regions. Segmentation can be either *data-dependent* or *data-independent*. Data-independent segmentation commonly consists of dividing an image into overlapping or non-overlapping fixed-size sliding rectangular regions of equal size and extracting and indexing a syntactical feature vector from each such region. This type of segmentation is easy and quick to perform, but generates a large amount of data. In addition, there is no guarantee that the segmentation is semantically meaningful. For some application areas, such as satellite imagery, this does not pose a problem, as one might expect large areas with similar texture. However, for images where there are few, important objects, data-independent segmentation is likely to divide the image in non-optimal locations, i.e. splitting a visual object over several regions.

Data-dependent segmentation is based on dividing the image based on its content, for example trying to identify objects, such as persons, from the background in photographic images. This type of segmentation produces fewer sub regions than data-independent extraction, and the ensuing segmentation can be used for automatic semantic labelling of image components. However, it requires more specialized tools and algorithms in order to produce semantically sound results. One example of this type of segmentation is *Blobworld* (Carson 1997), in which images are segmented using colour and texture features. This method is well-tailored toward identifying objects in photographic images, providing they stand out from the background. A similar example of data-dependent segmentation is the neural network based algorithm presented by Rowley and Baluja et al (1998)This algorithm is trained to identify and segment faces in photographic images. While these examples show that successful segmentation of some types of images is feasible, achieving reliably

[11] The image was retrieved from http://www.flickr.com/photos/turtlemom_nancy/3437417971/. Some rights reserved.

good segmentation quality has proven difficult, primarily related to the computational complexity of this (Datta, Joshi et al. 2008). Without proper segmentation, retrieval based on similarity between elements in an image is hard to achieve.

Next, a digital image is a two-dimensional representation of a three-dimensional space, and the possible variances in scale, rotation and orientation of visual objects are nearly unlimited. Factors such as light conditions, framing and perspective may complicate the matter further. Consider the case of a tree. At a distance, it can be described as a blobby top attached to an elongated bottom. However, as one approaches the tree, large branches become visible, then smaller branches play a role, followed by leaves and so on. Even very small changes in pitch, rotation or lightening conditions between two images of the same object might lead to major changes in the syntactical image features. In order for similarity comparison based on data patterns alone to be effective, the depicted objects must have a high degree of visual invariance, i.e. the quality of an object to be resistant to variations in visual appearance.

Consider the two images in Figure 14, below. Although the two images depict the same visual object and are very similar in semantic content (Both depict “a jumping dolphin”), they share few syntactical similarities. Images, even two images depicting the same object, are often heterogeneous in nature, and retrieval techniques based on syntactic features are by default not capable of overcoming this problem for a general application area, as they lack understanding of semantic concepts.

Figure 14 - Two different images of a “Dolphin Jumping”[12]

Summarized, CBIR systems primarily operate by extracting descriptors of the perceptual structures in an image, and use statistical similarity functions to compare different descriptors, identifying images that share similar perceptual structures. However, as these functions primarily operate on the perceptual structures, achieving successful retrieval based of semantic concepts has proven difficult.

[12] Both images were retrieved from the VISI system.

2.4 User Requests and Image Information Needs

Section 2.2 indicated the *possible* content a user may request from image retrieval system. However, this is not necessarily identical to the requests that users may be interested in. Identifying what the most used request types are may provide another view of the types of queries an image retrieval system should support. Several attempts have been made to classify the attributes of image retrieval tasks.

Shatford (1986) in (Westman and Oittinen 2006) refined Panofsky's three level image semantics scheme, categorizing the subject of image requests as *generic of* and *specific of* (the factual content of an image) and *about* (the expressional content of an image), and adding four facets (who / what / where / when) to each level.

In a study of requests received by the Hulton Deutch Collection, Enser (1993) identified two dimensions of request classification: unique/non-unique and refined/non-refined. The *unique* dimension refers to queries for specific and generic semantic concepts, while the *refined* dimension refers to whether the query is qualified with contextual specifications, such as *time, place* and *location*. Over 2.700 requests for images were analyzed.

Keister (1994) evaluated image requests received by the National Library of Medicine. He identified two major categories: Requests in which the user defined elements that should be present in an image, and topical requests, based on non-specific visual requirements. 239 queries were analyzed.

Rodden (1999) performed a qualitative study on how twelve persons organized their photo collections, how they felt about various approaches for indexing the collection and what their reactions were towards different techniques for searching for images. The participants described different approaches based on their actual retrieval tasks. When searching through their collection rather than browsing, they were usually looking for a particular photograph they have remembered. These searches were often based on remembering when the image was taken, using this to guide their searches. Searches for more generic content were uncommon, and in most cases the requests were for images of particular persons or images of a particular quality.

Markkula and Sormunen (2000) analyzed the topics expressed in image requests and illustration tasks presented to a digital newspaper photo archive. They classified the queries according to the major topics represented by the request: named objects and places, news events, object types, actions and types of events, type of place, film/TV and queries for a known photograph. 108 photo requests were analyzed.

Choi and Rasmussen (2003) presented a study of a number of requests for visual information by a group of students and faculty members at the history departments at three American universities.

The study categorized the queries into 16 types of search requests, each belonging to one of three categories: Specific and individually named content, generic content and abstract content. One of the specific semantic categories was labelled “Linear Time: Date or Period”. While the remaining 15 categories were directly related to the content of the image this category represents the non-visual elements of the image. For the purposes of this survey, this has been re-classified as non-visual content. A total of 38 natural language statement queries were analyzed.

Cunningham, Bainbridge and Masoodian (2004) performed a grounded theory analysis of 404 visual arts queries sent to Google Answers™. The requests were classified into 8 categories: Bibliographic (Metadata), Contents (The *participants* in the image), Genre (Style or genre of the image), Where seen (where the work was seen by the enquirer), Colour (Mention of colours used in the work), Example (copy or representation of the desired image), Abstract (abstract concepts or symbols represented in the work) and Affect (mood or emotional state induced by the image).

Jørgensen and Jørgensen (2005) evaluated professional image users’ queries to a commercial web image database. The queries were classified by the function of terms in the search strings. Queries for specific objects (“nouns”) accounted for the majority of the searches, but queries for descriptive terms (“adjectives”) and thematic queries (“concepts”) were also frequent.

Cunningham and Masoodian (2006) presented a study of 64 image-related searches evaluated through qualitative analysis and interviews. The information requests were categorized in 4 categories: *Specific needs* (referring to a specific person, event or activity, g*eneral nameable needs* (referring to general semantic content expressible in key words such as “a typical New Zealand landscape with sheep), *general abstract needs* (involving abstract concepts such as “an image symbolizing photography”) and *subjective needs* (referring to request satisfying emotional responses as interpreted by the user, such as “a funny photo”).

The different surveys use a range of different labels and categories for classifying the requests. However, with some adaption, these categories may be mapped directly to the levels of image content defined in chapter 2.2, with the addition of requests based on non-visual content (e.g. contextual descriptions and metadata) and requests for specific images. Table 2 presents an overview of these requests levels with some examples:

Table 2 - Categories of image queries

Request level		Description	Examples
1	Metadata / non-visual / contextual	Information requests based on non-visual content of an image.	Photographer, Creation date, title, location, position
2	Specific image	Information requests based on finding a specific target image	Find a particular image I have seen earlier.

3	Perceptual structures	Information requests based on the basic syntactical structures in an image.	Find images consisting of a certain colour or colour distribution, images with particular texture samples or images similar to an example image
4	Generic requests	Information requests based on generic objects, generic scenes or landscapes.	"A typical New Zealand Landscape", "a dolphin"
5	Specific requests	Information requests based on specific objects which can be named and identified.	Images containing "Barack Obama", Images of the London Skyline
6	Requests for narrative content	Information requests based on the narrative content of an image: The actions performed by the participants in the image, conditions of an image's participant, activities depicted in an image.	Find images containing pack of lions hunting a flock of zebras, images containing a "bleeding shark" or images depicting a football match.
7	Requests for abstract content	Information requests based on the abstract and thematic contents of an image	Images symbolizing "happiness", images perceived as "funny" and images about "love", "sport" or "war".

In order to determine if any of these levels are more important for users than the others, an attempt has been made to rank the levels based on the surveys presented above. Comparing and summarizing the results was difficult because of differences in the nature of the collections studied and the different typologies used to classify the requests. In addition, some of the studies were primarily directed towards professional users (Enser 1993; Keister 1994; Sormunen and Markkula 2000; Choi and Rasmussen 2003; Westman and Oittinen 2006), while others were more directed towards casual users or cover a wide range of interests (Rodden 1999; Cunningham and Masoodian 2006). Table 3 presents a ranking of the relative importance of the different request levels based on these surveys. For each survey the request types were given a score from 1 (highest) to 5 (lowest) based on how much that request type was used. An average, weighted score for each request category was created based on the number of evaluations mentioning the request category. *Rank* describes the ranking of the request level based on the surveys. *Request category* names the request category according to Table 2. *Average score* is the average, weighted score of the request category based on the surveys, and *evaluations* presents the number of evaluations mentioning the request category.

Table 3 - Ranking of image query types

Rank	Request Category	Average score	Evaluations
1	Metadata and non-visual content	1,20	5
2	Specific content requests	1,33	9
3	Generic content requests	1,78	9
4	Specific image requests	2,33	3

5	Narrative content requests	3,67	3
6	Abstract and thematic requests	3,86	7
7	Perceptual structures	4,50	2

The overview presented in Table 3 should be read with some reservations. The methodological approach used may have influenced the score and ranking. Some of the categories were only present in a low number of studies (*specific images*, *narrative content* and *perceptual structures)*. Next, with the exception of Rodden (1999), these surveys were primarily based on text based image retrieval systems. This may have biased the results towards requests that the users were familiar with or felt confident that they would be able to express. Consequently, requests that may be difficult to express using text may be underrated in these surveys. Additionally, these studies were based empirical data on which queries the users have made, not necessarily the types of queries they would *prefer* to use if they could choose. With these reservations in mind, some general observations can be made.

First, requests based on metadata and non-visual content were the most used category. It was rated as the most used and / or most useful query type in all the surveys that included it. It seems particularly important for some application domains, particular visual art requests (Choi and Rasmussen 2003; Cunningham, Bainbridge et al. 2004).

Second, requests for specific contents rated almost as highly as requests based on non-visual content. This category also obtained a high score in all surveys including it, and appeared to be particularly important for users browsing personal image collections (Rodden 1999) and for professional illustration purposes (Sormunen and Markkula 2000). In Enser (1993) 69% of the queries were after specific semantic content. In Sormunen and Markkula (2000) 54% of the requests were after named objects. In Choi and Rasmussen (2003), 66,3% of the queries were after specific content. In (Westman and Oittinen 2006), 40% of the requests and queries were for a specific person. In Cunningham, Bainbridge et al (2004), 42% of the requests were after named objects, people and events depicted in an image.

Third, requests based on generic image contents seemed to be useful. While this category was not as widely used as the two first categories, it represents a very large number of requests submitted to generic image collections (Jørgensen and Jørgensen 2005; Cunningham and Masoodian 2006; Westman and Oittinen 2006). In most of the surveys, queries for the generic content represented between 20 and 40 percent of the queries.

Fourth, specific image requests were not included in more than three of the surveys. These queries seems to be primarily important for people searching through their personal image collection looking for a particular image (Rodden 1999), or for illustration tasks requiring a specific image (Sormunen and Markkula 2000).

Fifth, the surveys reported that requests for both narrative and abstract content were generally not much used, generally representing between 5 and 15 percent of the requests (Shatford 1986; Choi and Rasmussen 2003; Jørgensen and Jørgensen 2005; Cunningham and Masoodian 2006).

Finally, requests for images based on *perceptual structures* were given the lowest possible score in both evaluations which included these. In Rodden (1999) the users could not describe any situations where they found this would be useful. Furthermore, of the requests analyzed in Cunningham et al (2004), only 10% were based on perceptual structures, and all of these were requests for images containing a named colour. It should be noted that it has been suggested that this form for request may be of high importance for users working with visual structures, e.g. artists and designers (Enser and Sandom 2003).

Note that these evaluations primarily reflect *what queries have been used*, which is not necessarily the same as *what queries are important* for users. It is possible that users, or at least some users, may be *interested* in performing other types of queries, but are either unable to do this, or find expressing these queries difficult.

2.5 Querying with CBIR systems

The previous sections gave an overview of what images *are*, what types of *contents* they may contain, what types of *requests* users submit to image retrieval systems, and how basic CBIR systems *index* and *compare* images. The final question is: How do current systems support users when *expressing their image requests* as queries to these systems?

A number of reviews of image retrieval systems have been published, such as Huang and Rui (1999), Eakins and Graham (1999), Veltkamp and Tanase (2000), Venters and Cooper (2000), and Jaimes and Chang (2002) Kherfi , Ziou and Bernardi (Jaimes and Chang 2002; 2004). While these studies present a broad overview of CBIR systems, they do not have a detailed study of past and current techniques for specifying visual queries.

Based on a combination of the mentioned studies and an evaluation of currently available CBIR systems, a survey of techniques and interfaces for general visual query specification has been made. The systems included in the survey are all suited for visual image retrieval. Systems focusing mainly on narrow domains (i.e. fingerprint and face identification) have not been included in the survey. Where possible, evaluation of the systems was done by first hand examination. However, several of these systems are no longer publically available. Evaluation of these systems has been based on original publications describing the systems where available, or through the abovementioned surveys where no other material has been available. The main objective of this survey was to identify and classify the different approaches available to a user when creating a visual query in current and past image retrieval systems. 59 systems have been evaluated. Table 62 (Available in appendix 2) presents an overview of these systems, along with the query formulation techniques supported by these systems. Six techniques for image query formulation were identified based on the survey:

1. **Query by Text (QBT).** Queries are either expressed as keywords or through selection of collection categories.
2. **Query by Features (QBF)**. Queries are based on user specification of low level features such as colour, texture and shape.
3. **Query by Internal Example (QBIE)**. Queries are based on selecting images already present in the image database.
4. **Query by External Example (QBEE).** Queries are based on having the user submit images that do not exist in the database.
5. **Query by Image Area (QBA)**. Queries are based on the specification of a region of interest in an example image.
6. **Query by Drawing (QBD).** Queries are created by the user by composing a query image representing their information needs using one or more drawing tools.

Several of the surveys mention *browsing* as a method of navigating through the image collection, and there are indications that many users find this a useful and convenient way of searching for images (Rodden 1999). However, since browsing is based on sequentially navigating through a collection it has not been included as a query method in this survey. Table 4 presents an overview of the query specification techniques used in the systems reviewed. Most of the systems support one (23) or two (25) query specification methods, but 9 systems support 3 query specification methods.

Table 4 - Summary of query methods

Type	Number	Percentage
QBIE	38	64,41 %
QBT	18	30,61 %
QBEE	17	28,81 %
QBD	13	22,03 %
QBF	12	20,34 %
QBA	8	13,56 %

Table 5 presents an overview of the combinations of query specification techniques used in the systems reviewed. Four combinations represent a majority of the systems. QBIE alone represents the most used query specification technique, used in 13 of the systems. Next, QBIE used in combination with text alone is used in 7 systems, while the combination of QBIE, QBT and other modalities are totally used in 12 systems. QBEE alone is used in 4 systems. The remaining categories are only used in 3 or fewer systems. QBA alone is not used in any of the reviewed systems.

Table 5 - Combinations of Query Specification Techniques

Query Technique	Number	Percentage
QBIE	13	22,03 %
QBIE+QBT	7	11,86 %
QBEE	4	6,78 %
QBIE+QBA	3	5,08 %
QBIE+QBEE	3	5,08 %
QBD	3	5,08 %
QBF	3	5,08 %
QBIE+QBF	2	3,39 %
QBEE+QBD	2	3,39 %
QBD+QBF	2	3,39 %
QBIE+QBD	1	1,69 %
QBT	1	1,69 %
QBEE+QBF	1	1,69 %
QBEE+QBT	1	1,69 %
QBD+QBA	1	1,69 %
QBD+QBT	1	1,69 %
QBIE+QBEE+QBF+QBT+QBA	1	1,69 %
QBF+QBT	1	1,69 %
QBIE+QBEE+QBT	2	3,39 %
QBIE+QBF+QBT	2	3,39 %
QBIE+QBEE+QBF	2	3,39 %
QBIE+QBD+QBA	1	1,69 %
QBIE+QBEE+QBD	1	1,69 %
QBIE+QBT+QBA	1	1,69 %
QBA	0	0,00 %

In most of the systems combining QBIE with other techniques, these other techniques are often used as a point-of-entry to the system. The initial query to the system is often expressed the other techniques, while QBIE is used to either refine the query using a relevance-feedback loop, or to initiate a new query based on one or more of the retrieved images. For example, in the AMORE system (Mukherjea, Hirata et al. 1997), the user may select a category of images through a textual label (e.g. “arts” or “travel”), or by choosing a random set of images.

Most of the systems surveyed shared a common structure for retrieval:

1. **Query Formulation.** This is normally the first stage in the query process. The user initiates the search process by creating or submitting a query using one of the available query specification tools.

2. **Specification of query parameters.** In this step, the user may have the option to qualify, limit or otherwise modify the query using one or more parameters. The parameters available are normally dependent on the methods used for indexing and querying.
3. **Image Retrieval**. This is the actual image retrieval phase, in which the query is processed and similar images are retrieved.
4. **Presentation of results**. In this step, the query results are presented to the user.
5. **Refinement of the query.** In this stage the user may choose to either refine or rephrase the query. Not all systems support this, and several different refinement and methods exist. Some examples are initiating a new search based on one or more of the retrieved images, refining the query results through a relevance-feedback loop and repeating the query using new query parameters.

Not all systems follow this sequence strictly; some systems do not allow the user to specify query parameters. Other systems have a more dynamic process where the user might not experience query formulation, image retrieval and result presentation as separate steps.

Each of the six query types are discussed in the following sections with regards to their strengths and limitations related to the different levels of user requests described in chapter 2.3 (Table 2, page 30)

2.5.1 Query by Text

First of all, why include text based queries in a survey of CBIR systems? By definition, queries expressed in textual terms are not visual queries. However, a large number of the systems either include QBT, or use QBT as the primary method for initiating the query process. As a result, including the QBT approach in this survey is required.

Text based image indexing and retrieval techniques refer to the use of textual descriptions of digital images. Retrieval is based on similarities between a textual search string and textual indexing, such as keywords or free text annotation of the images in the collection. These queries allow all retrieval techniques from traditional, text based information retrieval to be used for image retrieval. As long as the image collection has been properly annotated, it is possible to achieve very high levels of recall and precision for most image retrieval tasks. Tools can be used in order to expand and improve these queries, e.g. a thesaurus (Foskett 1997). A comprehensive overview of some text based retrieval techniques is available in (Baeza-Yates and Ribeiro-Neto 1999).

The main strength of text based queries is that they have both high expressive power and expressive convenience. Almost anything that can be expressed using language can be expressed using text.

The major challenge using text based queries for image retrieval is the *query mismatch* problem, i.e. the queries are expressed in one format (text), while the requested objects are in a different format (images). In order to retrieve images based on textual queries, the images *must* be indexed using textual descriptors. As no reliable methods for automatically creating such textual descriptors have been created, indexing images using text is predominately a manual process. As a consequence, image retrieval based on text faces three major challenges: The problems of *volume*, *subjectivity* and *explicability*; the images in the collection must be correctly and completely annotated. Missing or incorrect image annotation will result in poor search results. No matter how well the query is formulated, it will fail if the annotation is inadequate. Furthermore, the user may not share the vocabulary of the annotator, which might lead to a mismatch between the search criteria and the annotation.

The problem of volume refers to the fact that manual annotation of an image is a time consuming task. Indexing times quoted in the literature range from about 7 minutes per image from stock photographs at Getty Images, to more than 40 minutes pr image for a slide collection at Rensselaer Polytechnic (Eakins and Graham 1999). While it is relatively easy to create annotations for a small number of images, even a small personal computer now has the possibility to store millions of images, making manual annotation a daunting task, at best.

Furthermore, the combination of rich image content and differences in human perception makes it possible for two individuals to have very diverging interpretations of the same image. As a result, the description is prone to be both subjective and incomplete. Consider the image in Figure 15, below.

Figure 15 - A man feeding the killer whale 'Keiko'[13]

One possible textual annotation of this image could be "the killer whale 'Keiko' being fed by a man". However, this does not include the name of the man, when and where the image was taken, or the context of the depicted situation. Furthermore, while some might see this as an image representing how humans and animals interact, others might regard this as an example on how animals are exploited by humans. This is called the problem of *subjectivity*.

[13] The image was retrieved from the VISI-Maritime collection

Finally, there are some limitations when using text to describe visual structures. Some syntactical image features are difficult to describe with words. As an example, consider the case of the Norwegian Postal Services (Posten Norge AS). In September 2008 they launched a new logo (Figure 16) for their services. This resulted in international press coverage as the logo was very similar to a *Pokéball*[14] (Bryne 2008; Joystiq.com 2008).

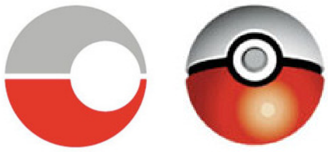


Figure 16 - The logo for the Norwegian Postal Services (left) and a *Pokéball* (right).

While it is easy for a human observer to visually see the similarities between the two objects in Figure 16, it is at best very difficult to give a precise, objective and non-biased textual description of the shapes and the similarities between them. While this is primarily a curiosity, similarities between trademarks may result in legal difficulties. When designing new trademarks and logos, it is important for the designers to identify potential infringements and similarities as early as possible. Searching for potential similarities between a new logo and existing trademarks using text may be very difficult. Similarly, queries based on colours, textures and other syntactic image features may be equally difficult. Even though one might have a clear understanding of the colour "red", there are so many different nuances and shades that correspond to the word, and there might be different names depending on who you ask. This is called *the problem explicability.*

As a final remark on the challenges of text based queries, it should be noted that the successful use of textual queries depends on the user's ability to express themselves verbally. This excludes young children, illiterate people and some people with learning difficulties. These groups are more or less excluded from the vast amount of information available as long as text-based queries are the only means of access.

In the case of level 1 requests (Metadata / non-visual / contextual), this type of information is almost exclusively described using language and available in textual format. This may be manually annotated (i.e. title and photographer), automatically annotated based on image context (Karlsen and Nordbotten 2008) or embedded in the EXIF[15] data of a digital image (i.e. time, place and resolution).

[14] An important artifact in the *Pokémon* animated TV series and computer games.
[15] Exchangeable image file format – a specification for the image format used by digital cameras. The specification uses the existing JPEG, TIFF Rev. 6.0, and RIFF WAV file formats, with the addition of specific metadata tags, such as date, time, image characteristics and user defined tags.

As the QBT approach is based on text, it is ideally suited for this type of requests, and the user may conveniently express his or her request in textual terms.

For level 2 requests (requests for a specific image), the usefulness of the QBT approach depends on the users' knowledge of relevant non-visual information such as photographer or location and that the image is properly annotated with this data. If the user has access to this kind of knowledge and the images are properly annotated, the use of these queries is identical to queries based on requests for non-visual contents. Without this information expressing these requests as text based queries is very difficult.

In the case of level 3 requests (requests based on perceptual structures), the QBT approach may prove to be very difficult, depending on the actual perceptual structures involved and the nature of the request, as described by the *problem of explicability*.

Requests of levels 4 through 7 (Generic, specific, narrative and abstract content) are, in most cases, based on some sort of semantic concept: Generic or specific objects, individuals, activities, themes or emotions. Common for these is that they are all described and classified using some sort of linguistic labels which can easily be expressed verbally through keywords or simple textual sentences or statements. Requests of level 4, 5 and 7 might be the easiest: This might only require simple keywords (e.g. "Dolphins", "Barack Obama", "New York Skyline" and "War"). Queries of level 6 (Narrative) may present some more problems. Simple queries may be easily expressed ("Jumping Dolphin", "People dancing Reel dance"). More complex request may present some difficulties, particularly if the compositional structure is of importance. However, the biggest challenge is that the images must be properly indexed with textual descriptors for text based queries to be useful for these requests.

The above discussion points in one direction: With the exception of queries based on perceptual structures and possibly queries for specific images, the text based approach seems to be a very convenient and efficient tool for expressing image queries. Almost all types of queries may be easily expressed using this approach. If the collection is properly annotated and supports text based queries, retrieval is a trivial task, only dependent on the capabilities of the text-based similarity retrieval algorithms. The usefulness of QBT is currently limited by the *media mismatch challenge* and the problems of *volume*, *explicability* and *subjectivity*.

2.5.2 Query by Features

The *Query-By-Features* (QBF) approach is based on specification of perceptual structures. Queries are created by creating and manipulating these perceptual structures through methods such as defining colour histograms[16], creating or selecting texture samples or similar feature specification techniques. One example of this approach is the MARS system (Ortega, Rui et al. 1997).

The expressive convenience of this query type depends on the query interface provided to the user. Working with colour histograms or manipulating textural patterns might not be the preferred approach for the average user, somewhat depending on the actual query interface. It is possible that this approach is primarily of use and interest to users with very specific needs or persons used to work with perceptual structures.

It is very difficult to perceive how this approach can be used in requests for non-visual content. Such tasks are primarily based on textual labels, and QBF are strictly visual in nature. However, QBF may be more useful when requesting a specific image, particularly if the requested image has very characteristic syntactical structures, such as a particular colour distribution or the presence of particular textural patterns.

The QBF approach may be useful for level 3 requests. If the user is interested in digital images containing specific colour distributions or textural patterns, the ability to precisely define and express this in the query may be very useful. If the query interface is easy to use, allows the user to precisely express the desired features, and the retrieval system has good support for similarity retrieval based on these structures, it is possible to achieve very high levels of recall and precision using this approach. Two of the participants in Rodden (1999), mention this. The participants described themselves as “very visual” (an architecture student and an art student). They claimed that the possibility of querying images based on perceptual structures would be very useful, particularly when querying after abstract photographs.

Finally, using QBF for requests for semantic content is not likely to be very useful. Most types of semantic content may be difficult to express using low level perceptual structures. Some potential exceptions are requests for objects that have very characteristic structures (i.e. the stripes of a Zebra or the fur of a leopard), or very abstract requests, such as using the colour “red” to express emotions such as “love” or “anger”.

Summarized, the QBF approach is most useful for requests based on perceptual structures, and it may be possible to use it in some specific semantic retrieval tasks.

---

[16] A color histogram is a numerical representation of the distribution of colors in an image, describing the percentage of pixels of each of given set of color ranges in a typically two-dimensional (2D) or three-dimensional (3D) color space.

2.5.3 Query by Internal Example

Query by Internal Example (QBIE) allows queries based on images already present in the image database. Queries are generally made by selecting one or more images from the collection and using these as the basis for the similarity search. Different approaches for selecting the initial image(s) were available in the surveyed systems, primarily by browsing through the collection or by choosing from a set of sample of images. Selection of sample images can be static and predefined, randomly drawn from the collection or based on a subset of the images based on an initial textual query. Most of the systems supported additional QBIE searches based on the result set obtained in a query. One example of the QBIE approach is the FIRE system (Deselaers, Keysers et al. 2008).

QBIE represent a simple and fast manner of expressing a query, as the user simply has to select a suitable query image. However, this simplicity comes at a cost, as the user is limited to the images in the initial sample or resulting from a prior query. Images buried in the collection are normally unavailable to the user.

In theory, the QBIE approach allows retrieval based on a broad concept of similarity: retrieve images that are similar to the query image. The expressive power of this query is dependent on how freely the user can define his or her notion of "similarity", e.g. the breadth of the query parameter selection. In theory, this might allow the user to express a wide range of queries. For example, a user submitting the image represented in Figure 10 (page 23) could refer to a number of different "similarity based" retrieval tasks:

- Retrieve images taken by the same photographer
- Retrieve images with similar visual content
- Retrieve images containing the same people
- Retrieve images in a similar location
- Retrieve images with a similar theme
- Retrieve images provoking similar feelings

If the user is free to define his notion of similarity, QBIE may have a high expressive power, limited only to the images available to the user and the way the users can define the query parameters.

However, in practice this does not seem to be the case. None of the surveyed systems support a wide range of query parameters. In most cases, the parameters are limited to specifying the weight of the low-level features (e.g. shape and colour) or comparable system-specific similarity criteria. While a lot of research has been done, and is currently being done, current CBIR systems are very limited in their ability to identify and compare other types of content than perceptual structures, as described by the problem of the *semantic gap*. This efficiently limits QBIE's usefulness for most types of semantic, narrative, thematic and abstract queries.

The usefulness of QBIE for requests for non-visual content is limited. If the digital image used in the query contains additional metadata, such as information about the photographer, GPS coordinates or creation date, this metadata can be used to identify images in the collection with similar metadata. However, this is not used to any extent in the surveyed systems. Expressing these terms through text based queries is likely much easier than using images.

For level 2 request, QBIE is not an ideal approach. First of all, if the user has information about the non-visual elements of the image (e.g. title or photographer), expressing this type of query through QBIE immediately appears inconvenient. If this information is not available, QBIE may be used if the user can access structurally similar images.

For requests based on perceptual structures (level 3) QBIE may be a powerful way to express the queries, depending on the actual retrieval task. This is particularly true for queries based on colours and textural patterns, as current CBIR systems can achieve relatively high recall based on these structures. If the retrieval task is of a general nature, and system can provide the user with relevant sample images, the system may be able to retrieve a number of relevant images. However, if the user has very specific retrieval tasks (such as finding a very specific textural pattern, shapes or specific colours), he or she still faces the problem of obtaining suitable example images.

The usefulness of QBIE for requests based on semantic content (levels 4 through 7) depends on the actual retrieval task and the capabilities of the retrieval algorithms. As noted above, the theoretical expressive power of QBIE is limited to how freely the user can express his notion of similarity. However, since current CBIR systems are more capable of comparing structural similarities, the usefulness is limited.

First of all, it may be relatively easy to express queries for basic semantic concepts (e.g. "dolphin" or "human") if the user has access to images containing these elements. However, it is difficult to image why a user would choose to express this using example images if text based queries are available. QBIE may be more useful if the user requests images with specific compositional structures, e.g. the request for a dolphin jumping out of the water in a particular way (described in section 1.1.1) and the request for images fitting a magazine layout (described in section 1.1.2).

However, as the complexity of the retrieval requests grows, successful use of QBIE becomes difficult. First of all, requests for objects with a high visual variance, such as living objects, is complicated by the large number of possible shapes of these objects. Next, if the query image contains additional participants or contextual elements, the user is faced with the problem of defining which parts of the image contains the relevant objects, and which parts of the image can be considered noise. Even if the query image contains very good depictions of relevant content, the retrieval system is faced with the challenge of successfully identifying and segmenting these objects, both in the query image and the digital images in the collection.

Finally, for more complex queries (narrative or abstract), it may be impossible to actually define which elements represent the interesting content. However, unlike the lower query levels, it may be more difficult for the user to express the exact similarity criteria which should be used for the image. Consider for example queries for “jumping dolphins” or “people dancing folk dance”. Images represent a snapshot of time. Hove (2007) discusses the possible of identifying movement and activities using the tools described in chapter2.2.5., but none of the systems surveyed provide the user with a possibility to define such structures, and none of the surveyed systems support retrieval based on these or similar structures.

Based on the above the following observations can be made. First of all, the poor quality of existing tools for segmenting, identifying and comparing semantic content represent a major obstacle for successful retrieval based on QBIE. Next, for actual query specification, QBIE may represent a fast and convenient manner of expressing queries. But users are limited to using the images presented as query candidates by the retrieval system. Furthermore, the user is limited to using the entire image as a query, and it may be very difficult for the user to identify which parts of the image contains the most relevant parts. Finally, QBIE does not represent a convenient or efficient format for expressing queries for non-visual requests or requests for specific images.

2.5.4 Query by External Example

QBEE allows the user to use images not already present in the system when using query-by-example. Queries are made by submitting one or more images to the system. The query image is analyzed similarly to the digital images in the collection and used for a basis for the similarity search. Two recent examples of this approach are the Retrievr system (Langreiter 2006) and TinEye (Idée 2009)

For all practical and technical purposes, the approach is identical to the QBIE approach. The main difference is that the user has more freedom when expressing queries. As long as the query image is in a format supported by the retrieval system, they may use any digital image in their possession, including images they have created or manipulated prior to the query process.

In the case of level 1 queries QBEE is identical to QBIE in all respects. And in most respects, they also similar for level 2 queries. However there are two scenarios in which QBEE may be useful for queries for specific images. First of all, a user may have in their possession a digital image they have no additional information about. For example, consider the case of someone working on a thesis concerning digital images who wishes to use a particular image as an illustration. The image was obtained from a web query, but no information about who the photographer was, where it was taken and if there are any limitations to the use of the image was included with the image. In order to use the image in the work, it is necessary to obtain this information. In this case, the user is interested in finding the same image, but with additional non-visual information. Another example is the case of a photographer suspecting that some of her images are used illegally on the internet. In this case, the user has all the necessary information about her image, but wishes to know if the image, or parts of it, has been used without permission. Changing the non-visual content of the image is a trivial task, and text-based queries do not represent a reliable approach. However, the syntactical characteristics of the illegally used image are likely to be similar to the original image, unless the image has gone through major changes. As current CBIR generally work by measuring the structural similarities between digital images, it is very suited for this form of retrieval. The *TinEYE* application (Idée 2009) is an example of a system providing support for these types of image requests.

In the case of level 3 queries, the QBEE approach provides the user with more flexibility than the QBIE approach. As long as the user has access to digital images containing the relevant perceptual structures, or he or she is capable of creating her own images containing these structures, successful retrieval is based on the retrieval capabilities of the image management system.

Similarly, the use of QBEE for the remaining query types is very similar to the QBIE approach. The main difference is that the user has more freedom as he or she has can use any digital images with a compatible format as queries. It is even possible for the user to use images internally in the collection, even if the system does not provide support for this, either by downloading the digital image or creating a new digital image using screen-capture. The user also has the possibility of manipulating the image, for example by cropping away noise. This may provide the system with less ambiguous visual queries. The other limitations of QBIE are also valid for QBEE.

Summarized, the QBEE approach is very similar to the QBIE approach, but with additional freedom for the user. Furthermore, it may be particularly useful for some scenarios in which the user is interested in finding a particular image. However, the main challenge of this approach is represented by the semantic gap.

2.5.5 Query by Area

Query by Area (QBA) is based on selecting a region of interest in an example image and using this as a basis for a similarity search. It is in most cases a special class of QBIE and QBE, in that the queries are expressed by selecting a sub-section of another image. However, it is included as a separate class as it offers the user some more freedom in expressing his request than the other approaches. One example of this approach is the visual search available at *Like* (Like.com 2009)

For most practical and technical purposes, the approach is identical to QBIE and QBEE. However, the main difference is that the user has the ability to be more precise in his or her queries. By selecting a particular region of the image, the user can help the system by removing irrelevant content. While the system still has to face most of the challenges presented by the semantic cap, this approach reduced the probability that the retrieval system will focus on irrelevant parts of the query image.

2.5.6 Query by Drawing

Query by Drawing is based on having the users compose a visual query representing their image request directly within the query interface using one or more drawing tools. This approach has also been described as *query-by-sketch*. One example of this approach is the Retrievr system (Langreiter 2006). QBD was first used in IBM's QBIC System (Niblack, Barber et al. 1993; Faloutsos, Barber et al. 1994).

11 of the systems reviewed support this query type. Five different tools for creating the queries were identified:

1. **Freehand drawing (F)**. This refers to the use of a mouse (or similar tactile input devices) to create a drawing in a similar manner to drawing with pen and paper. This allows the user a high degree of freedom to express any types of content, only limited to the user's competence with freehand drawing. Freehand drawing was available in 7 systems.
2. **Colour specification (C).** This refers to creating visual queries through the use of colours in combination with other tools. This allows the user to specify which colours should be present in the query image, as well as the spatial distribution of the colour. Colour specification was available in 9 systems.
3. **Geometric Primitives (GP)**. This refers to creating visual queries by using geometric primitives such as circles, squares and lines. These primitives can be used to build the spatial composition of the query image. Geometric primitives were available in 4 systems.
4. **Prototypes (SP)**. This refers to creating visual queries through the use of example shapes or shape prototypes, representing real-world objects. This allows the user to use these shape prototypes to spatially arrange the query participants within the query image. Shape prototypes were available in 3 systems. One example of this is ImageScape (Lew 2000).
5. **Texture (T).** This refers to creating visual queries through the use of texture samples or texture specification tools. This allows the user to express which textures should be present

in an image, as well as specify the spatial arrangement of these textures. Texture tools were available in 4 systems. One example of this is NETRA (Manjunath and Ma 1999).

Table 6 shows an overview of the tools available in the systems supporting QBD:

Table 6 - QBD Tool support

System	Tool types				
	F	C	GP	SP	T
CHROMA (McDonald, Tait et al. 2001)	x	x			
DrawSearch (Kherfi, Ziou et al. 2004)	x	x			
Hermitage Museum (Hermitage 2003)	x		x		
ImageScape (Lew 2000)	x			x	
IME (Petraglia, Sebillo et al. 2001)				x	
NETRA (Manjunath and Ma 1999)		x			x
Picasso (Del Bimbo, Mugnaini et al. 1997)	x	x			x
QBIC (Niblack, Barber et al. 1993)		x	x		x
Query by Visual Keywords (Lim 2000)				x	
Retrievr (Langreiter 2006)	x	x			
VisualSeek (Smith and Chang 1997)		x			x
VP Image Retrieval system (Veltkamp and Tanase 2000)	x		x		
WISE (Wang, Wiederhold et al. 1997)	x	x	x		
Total	7	9	4	3	4

According to Venters et al (2001), there is little evidence to support the usability of query tools based on *query-by-sketch*, and QBD remains one of the least researched and developed element of CBIR retrieval systems. However, literature generally acknowledges that the main drawback with this approach is that it is highly dependent on the user's ability to create good example images. Evaluation of the expressive power and expressive convenience of the tools in particular and the approach in general should be performed.

QBD gives the user a very high degree of freedom for visual query specification. Unlike QBIE and QBEE, the user is not limited to existing images; the user has total freedom when expressing visual queries. Theoretically, this freedom can be used in a number of different ways. First of all, it might be possible to use the spatial nature of QBD to identify the spatial characteristics of the requested images. One example of this is *ImageScape*. The user places representative icons, or shape prototypes, on a canvas where the major image features should be used. The system then returns images that have pre-classified database image features at similar locations to the query (Lew 2000). A similar approach was used in the EPIC system (Jose, Furner et al. 1998)[17]. The user could specify a spatial query by drawing and labelling rectangles on a sketchpad, in order to represent the relative positions and names of the objects desired within any retrieved image. According to Jose et al, their respondents rated the spatial approach higher than a text-based approach, claiming that they were better able to express a mental image representing their image requests.

It could be possible to create a basic visual query language allowing for queries based on narrative content, by using an approach similar to the language used to query spatial-temporal described in (Bonhomme, Trépied et al. 1999) or by building a query language based on the *narrative structures* presented by Kress and Van Leeuwen (Kress and van Leeuwen 2006; Hove 2007). However, with the exception of *ImageScape* (Lew 2000), *Query by Visual Keywords* (Lim 2000)and *IME* (Petraglia, Sebillo et al. 2001), the systems surveyed process the visual query images created through QBD in the same way as queries submitted using QBIE or QBEE. The visual query image is analyzed, and the resulting feature vector is compared to the vectors of the images in the collection. While the spatial nature of the query is processed at a syntactical level, no use is made of the semantics offered by the spatial structure of the query.

Depending on the nature of the image collection, retrieval based on similarity comparisons of feature vectors may have a major impact on the retrieval results. Consider an image collection primarily consisting of photographs. Depending on factors such as skill level and the time spent to create the query, the resulting query image may not have a structure resembling such images, as illustrated in Figure 17. Similarity functions based on perceptual structures will not find a large degree of similarity between these images. Consequently, QBD may be even more sensitive to the problems of the semantic gap than QBIE and QBEE. Finally, the while the user does not have to spend time looking for a suitable image, the actual time and effort needed to create a visual query may be a considerable obstacle.

[17] The EPIC system was not included in the survey; it is not based on CBIR technology.

Figure 17 - Two depictions of a "happy girl". A visual query and an image titled "Happy girl"[18]

As a final note, the QBD approach may represent a way in which users without a written language may gain access to image collections.

It is difficult to imagine how QBD can be successfully used for requests for non-visual content. The images or feature sets are created by the user, and it is not likely that these images will include any relevant metadata, as this would have to be entered by the user or generated during the creation of the image. This would be a very inconvenient way of expressing this type of image request.

QBD may be more useful for requests for particular images. If the user has a clear mental image of the requested image and the query creation interface has adequate tool support, the user may be able to create a query which is similar to the target image in terms of perceptual structure. This is supported by the results presented in Jose, Furner and Harper (1998): In cases where the users had a clear mental image of the images they would like to retrieve, the spatial approach was found very useful and convenient.

Depending on the tools available in the visual query interface, QBD may also be a good alternative for requests based on perceptual structures. Depending on the drawing tools available in the query interface, the user may express any type of perceptual structures. If the interface has a large number of colours and it is easy to compose an image using these, the hands-on approach provided by QBD may be more convenient than using QBF. Various shapes can be expressed using either freehand or by using predefined shapes, while textures can either be created using the drawing tools or, if supported, chosen directly from a selection of shapes.

Some types of requests for generic content may also be easy to express using QBD. Simple requests, e.g. requests for "humans" or "dolphins" may be expressed using simple drawings. These concepts could be expressed as icons, or through more realistic depictions, depending on the drawing competency of the user. However, with the possible exception of users without a written language, it

[18] The image of the happy girl was retrieved from Flickr, photographed by D Sharon Pruitt. Some rights reserved.

is difficult to imagine why users would prefer to use QBD rather than text to express such requests. But if the request is based on spatial characteristics, such as the particular shape of the interesting object, or the spatial arrangement of the important objects in the requested images, QBD may provide the user with the ability to express these more easily and more completely than a textual approach. And compared to QBIE and QBEE, the user has considerably more freedom in terms of the actual spatial characteristics of the query. Unfortunately, given that in most cases the CBIR approach is based on directly comparing the signature of the query image to the signature of the digital images in the collection, the query would have to be very detailed and realistic in order for a simple similarity function to correctly identify similarities between the images.

Concerning requests for specific content the usefulness of QBD may be very dependent on the nature of the task. If the user knows the name of the objects of interest it is difficult to imagine why the user would choose to draw the objects rather than using keywords. However, if the requested objects have very characteristic visual features (e.g. the Eiffel tower or the Golden Gate Bridge), current CBIR technology may be capable of retrieving images, particularly if the user is capable of creating a query visually similar to the real object.

In terms of requests for narrative content, QBD provides the user with more expressive power than the QBE based approaches. For QBE, the user is limited to using existing images, which may not be very explicit in terms of the narrative content. Using QBD, the user may create queries which are very explicit in their narrative content. For example, in the case of requests for images depicting “persons feeding a dolphin”, it might be possible for a skilled user to create drawings representing this content in a realistic manner. The user may also use some of the techniques described by Kress and van Leeuwen (2006) to create very explicit narrative structures in the query image, as suggested in (Hove 2007). However, with the possible exception of the ImageScape system (Lew 2000), no attempts have been made to evaluate or examine this approach.

Finally, in the case of requests for abstract content QBD shares the same problems as QBE. The user may create queries which may represent some aspect of their retrieval task, i.e. using specific colours, but specifying criteria for similarity might present a major challenge. Furthermore, the retrieval system still faces the challenges of identifying the content created by the user and retrieving images that are semantically similar.

In conclusion, QBD provides more freedom than QBIE and QBEE, but expressing the queries may be more inconvenient for the users. QBD may be more time consuming, and might possibly require the user to be competent in visual composition. QBD may also provide the user with the ability to be very precise when expressing some requests, particularly requests where the spatial distribution is important. Finally, QBD may provide the user with the option to precisely express certain types of narrative requests. However, as most current CBIR systems rely on a comparison between the visual query and the image collection, this might not be very beneficial to the actual retrieval process. A further evaluation of the possibilities provided by QBD should be done.

2.5.7 Visual Query Techniques: A Summary

Table 7 presents a summary of the main strengths and limitations of the different image query modalities.

Table 7 - Summary of techniques for visual queries

Query Type	Strengths	Limitations
QBT	Convenient Fast High expressive power	Is dependent on textual descriptors, which introduces the problems of Volume, Subjectivity and Explicability
QBF	Allows precise definition of perceptual structures	Difficult to express other query types May be difficult to use
QBIE	Convenient to use	Dependent on availability of query images Sensitive to visual noise Difficult to specify criteria for semantic similarity CBIR algorithms limited to syntactical similarities
QBEE	Convenient to use More freedom than QBIE	Dependent on availability of query images Sensitive to visual noise Difficult to specify criteria for semantic similarity CBIR algorithms limited to syntactical similarities
QBA	Convenient to use More freedom than QBIE and QBEE Less sensitive to visual noise than QBIE and QBEE	Dependent on availability of query images Difficult to specify criteria for semantic similarity CBIR algorithms limited to syntactical similarities
QBD	Large degree of freedom Can support precise definition of spatial arrangements and visual variance	May be dependent on the drawing competency of the user May be time consuming CBIR algorithms limited to syntactical similarities

Summarized, five main observations can be made by the above discussion:

1. There is a lack of studies on the usability and convenience of the different approaches for visual query specification. Most discussions concerning visual query specification are primarily based on argumentation, not empirical data. The problems of the semantic gap is still a limitation of the CBIR approach, but empirical studies on the users' behaviour in and conceptions about the visual query approaches may possibly uncover new application areas and other potential advances for CBIR technology.

2. The text based approach appears to be the fastest and most convenient way of expressing most types of image requests. The main exceptions are requests based on perceptual structures, requests for images with specific compositional structures, and requests for images where the participants are depicted from a particular angle or in a particular pose. However, in these cases, the text-based approach can be used to identify a relevant subset of the image collection, reducing the impact of the semantic cap for visual queries. Consider for the example illustrated in Figure 4 (structural similarities between a dolphin and a banana, page 7). Even if textual annotations do not describe the actual syntactical structure of the images, narrowing the image collection could remove the irrelevant content (fruits) and allow the CBIR algorithms focus on the relevant content (maritime mammals). Furthermore, the importance of queries based on non-visual content shows that the inclusion of text based queries is important for all image retrieval systems. However, the text-based approach is dependent on good annotation and textual indexing, a process that requires manual effort and is prone to the problems of volume, subjectivity and explicability.
3. The QBF approach has a very low expressive convenience, and is probably most useful for requests based on perceptual structures. It also does not represent the most convenient way of expressing image requests, but it is possible that QBF may be useful for people working professionally with visual structures. The actual usefulness and expressive convenience of this approach needs to be determined by additional empirical studies.
4. The different Query-by-Example approaches offer a fast and convenient way of expressing queries, but are limited to the actual syntactical features present in the query image. These approaches are also sensitive to the presence of visual noise, but that can be reduced by using QBA. The approach might also be beneficial for people unable to express queries using written language. The main problem with the QBE approaches is the semantic gap.
5. The QBD approach currently suffers from the problems of query interpretation and the problem of the semantic gap. The expressive convenience of the approach depends on the user's level of competency with visual structures. Next, it may be more time consuming than the other approaches. Also, if the user does not have a clear mental image of the desired image, deciding how to begin composing the query may represent a potential obstacle (Lai, McDonald et al. 1999; Lee, Jeong et al. 2004). However, it is possible that the improved freedom provided by QBD may present interesting opportunities for visual queries, particularly for requests based on narrative structures and requests for images where the spatial, compositional or representational structures are important. In addition, the approach may allow users without a written language to express queries. Finally, the users with a visual background mentioned in (Rodden 1999) claimed that the possibility of

querying using spatial techniques would be very useful for them, particularly when searching for more abstract photographs.

Based on this, a further and detailed evaluation of the possibilities of QBD might provide important insights into how CBIR systems can be improved, particularly for queries that are difficult to express using a text based approach. It would also be very interesting to examine how people working professionally with images and visual structures could use visual queries.

3 User Centred Evaluation: Methodology and Data Collection

A substantial amount of research in CBIR has been focused on improving the quality and efficiency of CBIR retrieval capabilities. In most cases, measurements such as *recall and precision* have been used to evaluate the quality and efficiency of CBIR systems. However, as this project has focused on the *user* a *user-centred* approach was chosen. The methodological approach used in this project was based around *users* expressing *image requests* as *visual queries* through *query by drawing*. The methodological framework, the data collection tools and the analysis tools were chosen based on their usefulness for evaluating the *needs, experiences and challenges* of the *users,* as expressed by the five main research questions.

As noted in the previous chapters, Content Based Image Retrieval may be used by different user groups for very different image retrieval tasks. There are a large number of different approaches for interface design, and different tools for expressing visual queries through drawing have been used. In order to obtain as generalizable results as possible, two CBIR systems with different tools and different mechanics for the query specification process were used in the study. Furthermore, as most CBIR systems are based on a direct comparison between the visual query image and images in the image collection, it was decided to include users with a background in drawing, design and related visual competencies in order to present an additional level of detail to the empirical data, assuming that these users are capable of drawing images that are more realistic than other users.

3.1 **Methodological Framework**

Figure 18 presents an overview of the methodological framework used in the project. The main engine for data collection used in the project was the *visual query drawing process*, in which 2 groups of respondents were observed when performing a set of image retrieval tasks in two different CBIR systems. The respondents were interviewed after performing the retrieval tasks, and the query images created in the process were analyzed.

The work consisted of three studies performed in laboratory settings. Each experiment session took place in a room with only the respondent and the researcher present. Different elements were included in the three experiments, and the data collection tools and methods were improved between the experiment sessions.

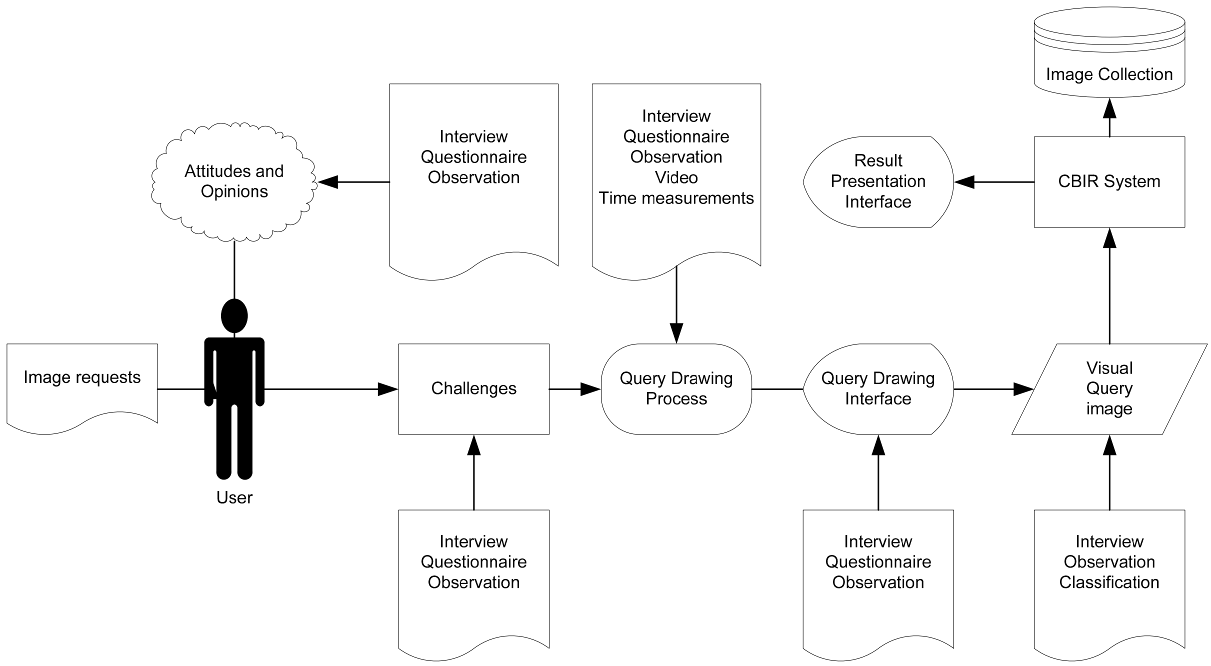


Figure 18 - Overview of data collection methods.

3.1.1 Respondents

Two groups of respondents were included in the experiments: 17[19] Information science students from the Department of Information and Media Science (IFIM) and 13 students from the Bergen Academy of Fine Arts. The first group was included primarily based on their availability: they were easily recruited. Even though these students represent a certain demographic group, particularly with regards to their knowledge of software systems in general and retrieval systems in particular, it was believed that the students are relatively heterogeneous and representative of a wider population with regards to drawing competency and experiences with visual image queries. There were 14 male and 3 female students in this group, which will be referred to as the "IFIM group"[20].

The second group represents respondents with a "visual background". An arrangement was made with the Bergen Academy of the Arts, providing an opportunity to present the project for bachelor level students of Visual Communication and Design. Seven students (3 female, 4 male) were recruited from Visual Communication and two students (both male) were recruited from Design. In addition, four respondents (3 male and 1 female) were recruited through other channels. Two respondents were pursuing a MA in fine arts, one respondent had finished a similar MA, and one

---

[19] 18 students had initially volunteered, but respondent 13 did not attend the experiment. As all material had been prepared in advance, it was decided to use the original numbering scheme. Consequently, respondent 13 is excluded rather than renumbering the following respondents.

[20] Named after the Norwegian name of the Department of information science and media studies: "Informasjons- og Medievitenskap".

respondent was working as an industrial designer, with a MA in industrial design. The respondents in this group all worked with images and visual structures on a regular basis, and were considered as "professional" users. This group is referred to as the "KHIB group"[21].

In all, there were 23 male respondents and 7 female respondents. The average age was 27 years. Most respondents in the IFIM group had prior experience with various image management and retrieval systems such as Google Images. Most of these respondents also used text-based image retrieval on a regular basis. Respondent 15 stated that he had knowledge of how CBIR systems work. Respondent 12 had tried different CBIR and QBD systems, but claimed very little knowledge of the underlying software system. Respondents 20, 21 and 22 had some prior experience using the VISI retrieval system[22], and one had prior experience working with the CBIR system in Oracle. The remaining had little or no experience with CBIR.

In the KHIB group, all respondents were fairly proficient with various image retrieval systems such as Google Images. 5 had never used image management software such as Flickr or Picasa. Most of the respondents used text-based image retrieval on a regular basis. None of the respondents had any prior experience using visual image queries or QBD. 9 of the respondents considered they were skilled or very skilled in drawing and were drawing frequently. Table 8 presents an overview of the respondents. The final row shows the mean value for the numeric columns.

[21] Named after the Norwegian name of the Bergen Academy of the Arts ("Kunsthøyskolen i Bergen").
[22] One of the CBIR systems used in the experiments.

Table 8 - Overview of Respondents

	Age	Gender	Group	exp	search	mgt exp	Exp	freq	Work or private	skills	frequency	skills	drawing	skills	Dig pen freq
1	26	Male	1	5	4	2	2	Weekly	Mostly private	2	N/A	2	N/A	1	N/A
2	31	Male	1	5	5	4	2	Monthly	Equally	1	N/A	2	N/A	1	N/A
3	22	Male	1	5	5	1	1	Weekly	Mostly private	2	N/A	5	N/A	1	N/A
4	34	Female	1	5	3	4	2	Monthly	Mostly private	4	N/A	2	N/A	1	N/A
5	27	Male	1	4	4	1	1	Monthly	Mostly private	2	N/A	2	N/A	2	N/A
6	26	Male	1	4	4	2	2	Weekly	Mostly private	3	N/A	3	N/A	1	N/A
7	30	Male	1	5	5	4	1	Weekly	Mostly private	1	N/A	2	N/A	1	N/A
8	29	Female	1	4	2	2	2	Monthly	Mostly work	3	N/A	4	N/A	1	N/A
9	26	Male	1	5	4	2	2	Monthly	Equally	2	N/A	2	N/A	1	N/A
10	24	Male	1	4	3	2	2	Monthly	Mostly private	1	N/A	2	N/A	2	N/A
11	23	Male	1	4	3	3	1	Monthly	Mostly work	1	N/A	3	N/A	3	N/A
12	23	Male	1	5	4	5	3	Weekly	Equally	2	N/A	3	N/A	2	N/A
14	30	Female	1	5	4	4	1	Monthly	Only private	1	N/A	1	N/A	1	N/A
15	31	Male	1	4	2	2	2	Weekly	Mostly work	2	N/A	1	N/A	1	N/A
16	28	Male	2	4	4	3	1	Weekly	Equally	4	Daily	3	Monthly	2	Seldomly
17	32	Male	2	4	4	2	1	Weekly	Mostly work	5	Daily	1	Seldomly	1	Never
18	30	Female	2	4	4	1	1	Weekly	Mostly work	5	Daily	2	Seldomly	1	Never
19	33	Male	2	5	5	5	1	Weekly	Mostly work	3	Weekly	2	Monthly	2	Seldomly
20	21	Male	1	5	4	2	4	Weekly	Mostly private	1	Monthly	3	Seldomly	1	Never
21	24	Male	1	5	4	1	1	Weekly	Equally	1	Seldomly	2	Seldomly	1	Never
22	22	Male	1	5	3	2	1	Monthly	Mostly private	1	Seldomly	2	Seldomly	1	Never
23	43	Male	2	4	2	1	1	Monthly	Equally	2	Weekly	2	Seldomly	1	Never
24	20	Male	2	5	5	5	1	Weekly	Equally	4	Weekly	4	Weekly	4	Weekly
25	24	Female	2	4	4	2	1	Weekly	Equally	5	Daily	3	Monthly	3	Monthly
26	26	Male	2	5	5	2	1	Monthly	Mostly work	2	Weekly	5	Monthly	2	Seldomly
27	22	Male	2	3	3	1	1	Monthly	Mostly private	4	Weekly	2	Seldomly	1	Never
28	24	Female	2	4	4	3	1	Weekly	Mostly work	3	Weekly	2	Seldomly	3	Seldomly
29	27	Female	2	5	5	1	1	Daily	Only work	5	Daily	4	Monthly	4	Weekly
30	26	Male	2	4	4	1	1	Weekly	Mostly work	4	Daily	2	Monthly	1	Never
31	25	Male	2	5	3	2	1	Daily	Mostly work	4	Daily	2	Never	5	Weekly
Avg.	27	-	-	4.50	3.83	2.40	1.43	-	-	2.67	-	2.50	-	1.73	-

- *Group* describes the group of the respondent: IFIM (1) or KHIB (2).
- *Search exp* describes the respondent's experience with ordinary search engines, e.g. *Google*.
- *Image search exp* describes the respondent's experience with text based image retrieval engines, e.g. *Google Images*.
- *Image mgt exp* describes the respondent's experience using image management systems such as *Picasa*[23] or *Flickr*[24].
- *VQ Experience* describes the respondent's experience with visual image queries.
- *Search freq* describes to how often the respondent usually performs an image search.
- *Work or private* describe whether the respondent's image searches are primarily work (or study) related or personal
- *Drawing skills* describe the respondents own description of their drawing abilities
- *Mouse skills* and *dig pen skills* describe the respondents' own description of their skills drawing with a mouse and a digital pen.

[23] http://picasa.google.com/
[24] http://flickr.com/about/

- *Drawing Frequency*, *Mouse Drawing Frequency* and *Dig Pen frequency* describe how often the respondent draw, draw with a mouse and draw with a digital pen. This question was not presented to respondents 1 through 15.

Each characteristic is provided subjectively by the respondent on a 5 point scale where 1 indicates *no experience* or *no skill* and 5 indicates *very experienced* or *very skilled*. S*earch frequency* is divided into *Daily, weekly, monthly, more seldom* and *never*. Similarly, *work pr private* is divided into *only private, mostly private, equal, mostly work* and *only work*.

Two of the respondents in the IFIM group reported that they suffered from a slight handicap making it difficult to draw precisely.

All respondents were presented with a written introductory letter, and were asked to sign a form of consent. Both are included in Appendix 4 - Data Collection Tools (In Norwegian).

3.1.2 Visual Image Query Interfaces

Two different CBIR systems were used in the experiment, the VISI prototype[25] and the Retrievr system[26]. These systems presented different approaches for the query process, and offered the respondents with very different drawing tools and interface options. The use of two different interfaces would make it possible to identify results that might be common for both interfaces, as well as identify results that can be related to certain aspects of a particular interface. It would also allow the respondents to compare the two interfaces and describe what they liked or disliked about the two approaches. The two systems are detailed below, and the differences between them are summarized in Table 9.

VISI (Vortex Image Search Interface) is a prototype web based CBIR system, developed as a test bed for content-based image retrieval using a Cold Fusion front-end to Oracle 9i InterMedia CBIR software. The system consists of three main elements: *Visual query specification*, *query parameter specification* and *result presentation and browsing.*

Query by Drawing (QBD) is provided through a sketch interface, illustrated in Figure 19. The tool is based on a Java Applet, *J-Painter*[27], and provides a set of basic drawing tools: Freehand drawing; basic geometric shapes such lines, rectangles, circles; colour selection, a limited palette of pens and two simple texture tools (special pen tools).

---

[25] Available online at http://link.uib.no/?6grcZ
[26] Available online at http://labs.systemone.at/retrievr/
[27] http://www.izhuk.com/painter/

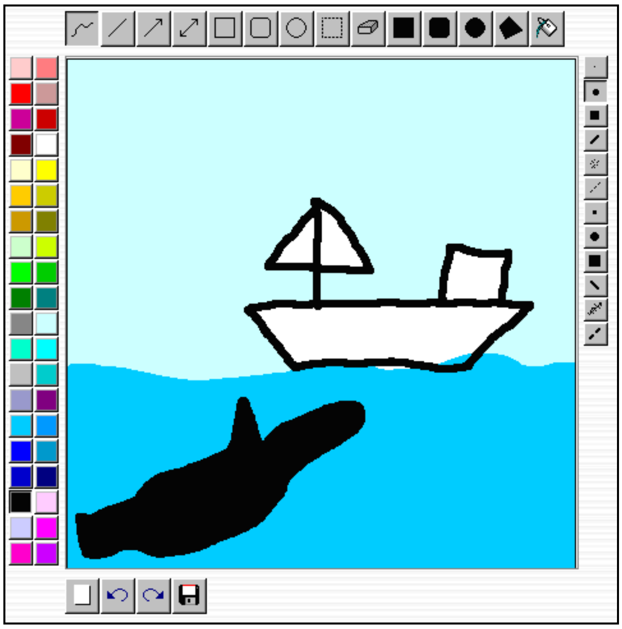


Figure 19 - VISI Sketch tool with a visual query image (Query 5)

Specification of the query parameters is provided through manipulating sliders. The user can specify how much weight the retrieval system should give to *shape, colour, texture* and *spatial arrangement* in the visual query, and specify a similarity threshold for the query results. This is shown in Figure 20. These parameters are directly related to the input parameters of the CBIR algorithms of Oracle 9i (Ward 2001).

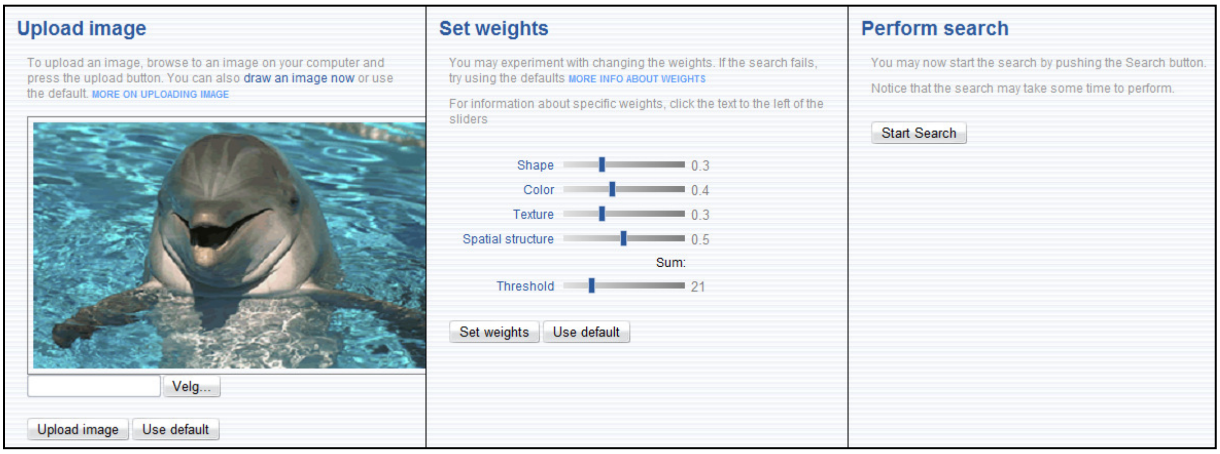


Figure 20 - Query parameter specification in the VISI prototype.

Query results are presented as shown in Figure 21. The query image and query parameters are presented on the left part of the screen, while thumbnail images of the results are shown in order of system determined relevance on the right part of the screen. The user can click on the thumbnails in order to show a larger image or use the image as a QBIE search.

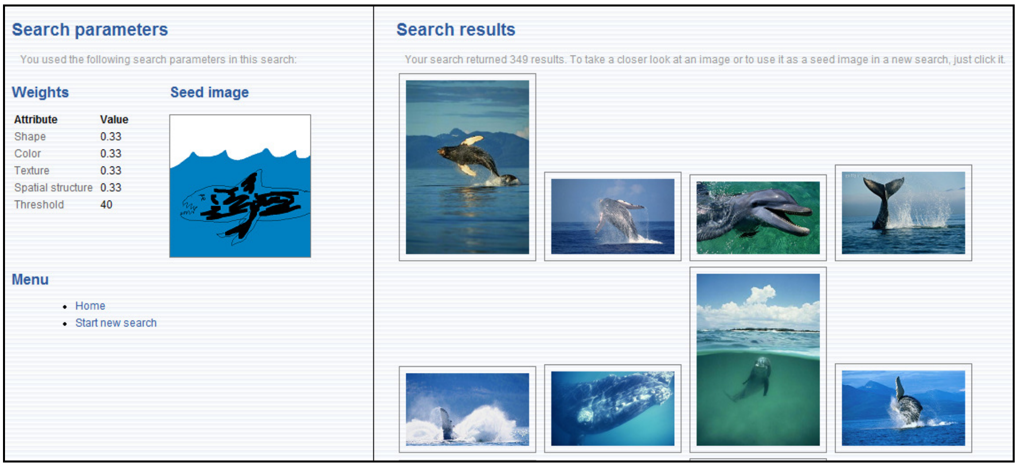


Figure 21 - VISI query result presentation.

The drawing tools are representative of the state-of-the art of QBD interfaces, as described in chapter 2.5. The performance of the CBIR similarity search is limited by the capabilities of the Oracle InterMedia package (Hove 2003). A full description of the prototype is available in (Næss 2007). A report on the usability of the prototype is available in (Egeland 2007).

The retrievr system is an experimental service which lets users search and explore in a selection of Flickr images by drawing a rough sketch (Langreiter 2006). The *retrievr* system has a different approach to the visual query process than the VISI prototype. It has a more basic QBD interface than the VISI system. Figure 22 shows the visual query specification interface for the Retrievr system next to the VISI interface.

The Retrievr interface only supports freehand drawing using one of four pen sizes (10, 20, 30 or 50 pixels). In addition, it supports a wide range of colours: 72 different shades of 12 primary colours, totalling 864 colours and colour nuances.

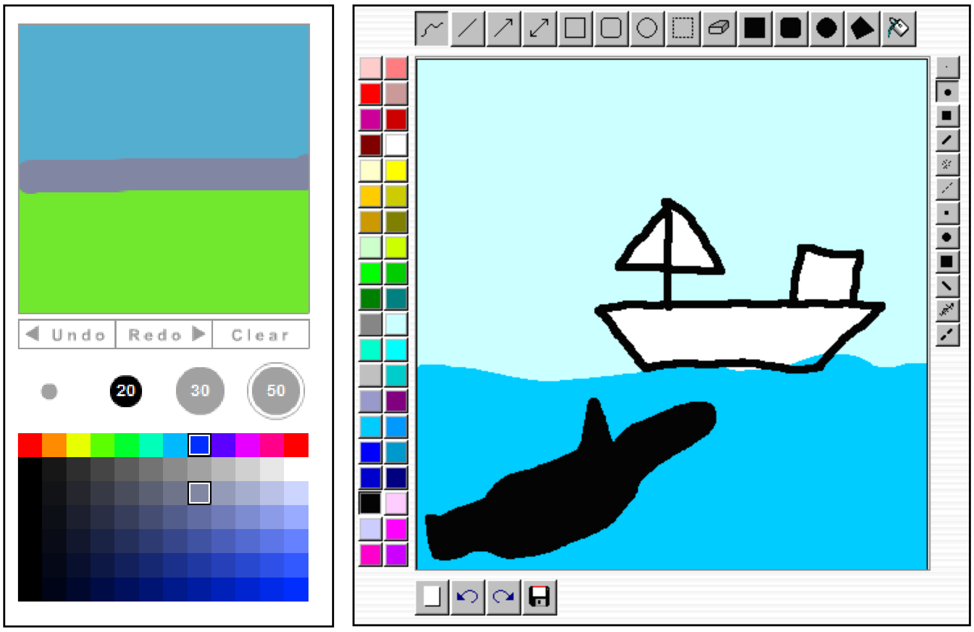


Figure 22 - The Retrievr interface compared to the VISI interface

Retrievr does not include specification of query parameters, and the user has no control over how the system processes their visual query image. Query results are updated whenever the user stops drawing, and are continuously displayed on the right part of the same screen as the query specification interface, as shown in Figure 23.

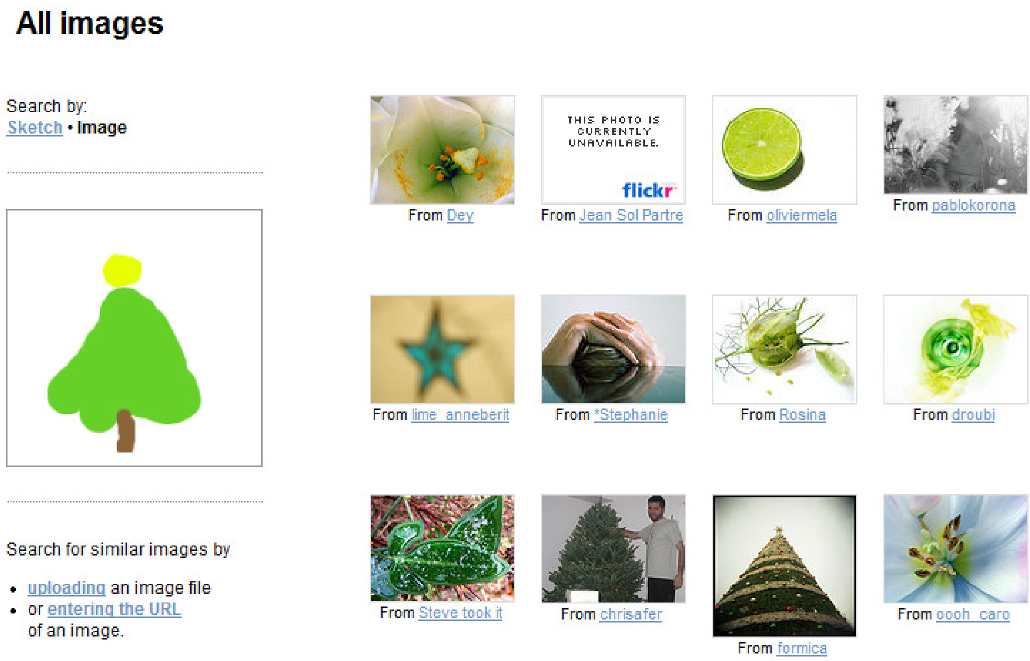


Figure 23 - Result presentation in the Retrievr interface, showing query #175.

This presentation of the results reflects a different search dynamic than the VISI interface. The VISI search process is based on having the user create a query and submit it to the retrieval system, which then retrieves the results, and it is not possible to modify the initial query. Retrievr allows for a higher degree of interactivity and refinement, and it is possible to modify and adapt the query based on the query results. Finally, the drawing canvas of the VISI interface is about 4 times as large as the canvas in Retrievr.

The Retrievr system is based on the work presented in (Jacobs, Finkelstein et al. 1995). The interface is built using Python with a frontend made in Macromedia Flash (Langreiter 2006).

Table 9 - Overview of differences between VISI and Retrievr

Interface option	**VISI**	**Retrievr**
Pen types	**12** Different sizes and shapes	**4** Different sizes
Freehand	**Yes**	**Yes**
Tools	**Yes** Geometric shapes, duplicator, eraser, fill tool and polygon tool	**No**
Colours	**40**	**864 (12 colours with 72 shades each)**
Query parameters	**Yes** User can specify the relative weight of colours, textures, shapes and their overall composition	**No**
Canvas size	400 x 400 pixels	180 x 180 pixels
Query process	**Static** The user must first draw the query, then specify the parameters, then browse the results. The query image cannot be modified.	**Dynamic** Results are shown and refreshed while the query is created. The query image can be updated several times.

Figure 24 and Figure 25 shows the relative sizes of the query images created in VISI (Figure 24) and Retrievr (Figure 25).

Figure 24 - Actual size of VISI query images.

Figure 25 - Actual size of Retrievr query images.

3.1.3 Input device

A Wacom Intous 3 graphical tablet[28] with a digital pen was used as the primary input device for both interfaces. While a mouse might be the most commonplace input device for computers, it might not be the preferred tool for drawing. Recent studies (Barthelmess, Kaiser and Lunsford 2006) have

[28] http://www.wacom-europe.com

shown that interfaces that depart from existing practices require a higher cognitive load from the users, taking focus away from the task at hand. In addition, research done by Venters et al (2001) indicates that a graphic tablet might be a more natural and direct input for visual queries. The respondents should have optimal conditions for sketch input, and consequently a state-of-the-art drawing tablet was provided as an input device.

3.1.4 The Image Collections

Two image collections were used in the project. The image collection used in the VISI system contained almost 400 images depicting maritime animals, people involved in maritime activities and different objects and activities related to this. The collection was originally custom built for the *Virtual Exhibits on Demand* project (Nordbotten 2005).

The Retrievr system operates on a large number of the *interesting*[29] images of the Flickr photo management application. The number of images available in the Retrievr engine is not known, but experimentation has shown that appears to be a large number of different images. The collection covers a wide range of images, image content and image categories, as the images are uploaded to Flickr from a very large user base.

3.1.5 Image Retrieval Tasks

In order to perform the queries, the respondents were given a set of image retrieval tasks. These tasks were developed based on the retrieval tasks described in chapter 2.3. Three types of image requests were used:

- **Requests for generic semantic content:** These are image retrieval tasks where generic objects represent the major information requests, e.g. retrieve images of a dolphin or retrieve images of a flower. In the rest of this thesis, these image requests and the queries created based on these are referred to as either TYPE 1 REQUESTS and TYPE 1 QUERIES.
- **Requests for narrative content:** These represent image retrieval tasks where the narrative content of the image is the major information request, e.g. retrieve images of a jumping dolphin or retrieve images of people practicing sports. In the rest of this thesis, these image requests and the queries created based on these are referred to as TYPE 2 REQUESTS and TYPE 2 QUERIES.
- **Scenes**: These represent image retrieval tasks where the scene depicted in the image represents the major information request, e.g. *retrieve images of a pair of whales in an arctic*

[29] See http://www.flickr.com/explore/interesting/ for more information.

landscape or *retrieve images of a skyline*. In the rest of this thesis, these image requests and the queries created based on these are referred to as TYPE 3 REQUESTS and TYPE 3 QUERIES.

As described in chapter 2.3, requests for generic content represented one of the most used image requests sent to generic image retrieval systems. It was included in this project in order to evaluate how the respondents used QBD as a tool for this type of query.

Requests for narrative content were included in order to see how the users expressed this type of queries using QBD. As indicated in chapter 2.5, the QBD approach may potentially be a powerful tool for expressing this type of request.

Requests for scenes were included in order to see how the users behaved when expressing request which may have an element of spatial structure.

While it definitely would be interesting to evaluate QBD for the other requests described in chapter 2.5, particularly for requests based on levels 3 (requests for perceptual structures), 5 (requests for specific content) and 7 (requests for abstract content), it was necessary to limit the number of queries and tasks evaluated in order to keep the scope of the project and the length of each experiment session at a manageable level.

Two different approaches for creating the actual requests were adopted: Predefined tasks and scenario based tasks. *Predefined tasks* were short sentences describing a generic image retrieval task, such as “Find images depicting one or more sharks”. The purpose of these tasks was to provide the respondents with a common set of retrieval tasks that would allow for a certain degree of comparison of the visual query process. These were all classified as one of the three query levels described above.

The use of such predefined tasks may not be representative of realistic image retrieval tasks. As noted by McDonald, Tait et al (2001), the use of scenarios might simulate how respondents might typically interact with image retrieval systems in a non-experimental setting. Consequently, the predetermined tasks were supplemented with *scenario based tasks*, in which the respondents were presented with a text and asked to define their own image retrieval tasks based on this text. The level of these queries was determined post study, based on the respondents’ own description of the retrieval task.

Table 10 presents an overview of the image retrieval tasks used in the experiments. A total of 31 different retrieval tasks were used. The tasks were originally expressed in Norwegian, and have been translated into English. The original tasks are available in Appendix 4 - Data Collection Tools (In Norwegian). Tasks 1 through 16 represent the predefined image retrieval tasks presented to the respondents in the VISI system. Tasks 17 through 21 and tasks 23 and 31 represent tasks defined by the respondents themselves through the maritime scenario text. Tasks 24 through 29 represent the predefined image retrieval tasks presented to the respondents in the Retrievr system, while task 30 represents the scenario based requests defined by the respondents for the Retrievr system. Task 22 represents queries which were “unique” - i.e. tasks defined by the respondents either through the scenario texts or made up, and are only represented once in the material. The final column “Queries made” represent how many visual queries were made based on a given task.

Table 10 - Overview of image retrieval tasks

#	Query Text	Queries made
1	Find images of a seagull	16
2	Find images of a scuba diver	18
3	Find images containing one or more sharks	16
4	Find images containing one or more whales	6
5	Find images containing one or more birds	11
6	Find images of one or more seagulls eating	7
7	Find images of one or more sharks attacking	22
8	Find images of a seagull eating a fish	7
9	Find images of an injured bird	3
10	Find images of a happy dolphin	4
11	Find images depicting a maritime animal	18
12	Find images of a ship	18
13	Find images of a predator attacking another animal	13
14	Find images of humans nursing a beached and injured whale	13
15	Find images of two dolphins entertaining humans in an aqua park	16
16	Find images of one or more animals swimming in an arctic environment	14
17	Find images of a dolphin playing with a ball	11
18	Find images of humans interacting with a dolphin	11
19	Find images of a dolphin swimming with a boat	11
20	Find images of a jumping dolphin	4
21	Find images of a dolphin	3
22	Unique queries*	13
23	Find images of an injured dolphin	2
24	Find images depicting a flower, a tree or another type of plant	20
25	Find images depicting furniture or an interior object.	34
26	Find images of humans practicing sports	34
27	Find images of a happy girl	18
28	Find images of a skyline	21
29	Find images of several people and / or animals gathered in a rural setting	16
30	Find images of a forest	4
31	Find images of a dolphin entertaining people in a boat	9

3.2 Three Experiments

Data were collected over the course of three experiment sessions, each focusing on one specific topic. The first experiment was the initial study, and the goal was to gain a basic understanding and overview of the research questions. This experiment included respondents 1 through 15, all from the IFIM group. The respondents were self-selected based on an e-mail sent to all students at the Department of Information Science and Media Studies, University of Bergen. All queries were made using the VISI interface. The predefined tasks were based on the maritime nature of the image collection. For the scenario based tasks, the respondents were given excerpts from a newspaper article describing a dolphin visiting the Norwegian fjords. They were asked to describe an ideal image which they would like to use to illustrate the article, and then try to express this request using a visual query. The tasks used and the newspaper article are included in Appendix 4 - Data Collection Tools (In Norwegian). An initial evaluation of this data was performed immediately following the data collection, and is reported in (Hove 2007).

The goal of the second experiment was to include respondents with a background in visual arts. The experiment was conducted using a similar approach to the first experiment, but the tools used to collect data were improved based on the experiences from the first experiment. The intention was to recruit 10 people with a background in fine arts, visual communication or similar fields. Unfortunately only 4 people could be recruited at the time of this experiment (Respondents 16, 17, 18 and 19). The respondents were recruited based on an advertisement posted on a web-forum frequented by people with this background. All respondents were self-selected. All queries were made using the VISI interface. The retrieval tasks were based on the tasks used in the first experiment, with some new tasks added based on the experiences from the first experiment. The tasks are available in Appendix 4 - Data Collection Tools (In Norwegian).The scenario used was identical to the text used in experiment 1.

Two goals were identified for the third experiment: Include more respondents with a visual background, and have the respondents use both query interfaces. The retrieval tasks used for the VISI system were based on the two first experiments, with some new tasks added. This was done in order to have an equal number of queries for each task type. The same text was used for the scenario based tasks. A new set of tasks were developed for the Retrievr system. These were similar to the tasks used for VISI, but adapted to the images available in Retrievr. For the scenario based task, the respondents were presented with a poem (“I skogen” / “In the Forest”), and asked to find at least two images which could be used to illustrate the poem. The retrieval tasks and the poem are included in Appendix 4 - Data Collection Tools (In Norwegian). 9 new respondents were recruited from the Bergen Academy of the Arts (respondents 23 through 31). In addition, three additional information science students were recruited (respondents 20, 21 and 22). It was intended to have at least 6 information science students in this group in order to support comparison between the two groups' use of the different interfaces, but only three could be recruited in time for the experiment.

3.3 Data Collection: Methods and Materials

As noted by (McGrath 2000), different methods have their own strengths and weaknesses, and an approach based on multiple strategies is advocated: "credible empirical knowledge requires consistency or convergence of evidence across studies based on different methods". Five main methodological approaches have been used to collect data: questionnaires, observation and think-aloud protocol, semi-structured interviews, video log and video analysis, and a classification scheme for the visual query images. While some of these methods are often used in *usability testing*, the use of these methods in this project has not been to evaluate the usability of VISI and Retrievr, but to analyse the *approaches* represented by the two systems.

Questionnaires represent a fast and easy way of gathering comparable measurements of the respondents' background and prior experiences as well as their subjective level of satisfaction and their evaluation of different aspects of visual query specification. While questionnaires are less flexible and open-ended than an interview, they can be analyzed more rigorously, and might provide a basis for comparison between the different respondents in a study (Dix, Finlay et al. 2004).

Observation, think-aloud sessions, video recording and protocol analysis represent a way of gathering information about actual use of a system which is not available through questionnaires and interviews. *Think-aloud* represents a form of observation where a respondent is asked to talk through what she is doing while she is observed, describing what she believes is happening, why she is performing in a certain way or what she is trying to do. *Protocol analysis* represents analysis of the observation notes and the think aloud sessions following the experiment session. *Video recording* has the advantage that we can see *what* a respondent is doing.

Interviews provide a direct and flexible way of gathering information about a subject, and are particularly useful for accessing information which might be difficult to obtain using other methods, such as questionnaires and observation. The flexibility provided by interviews allow for high-level evaluations, such as eliciting information about user preferences, impressions and attitudes. They may also reveal problems, issues or whole areas that could not have been anticipated by the researcher, or not have occurred during observation. Interviews can also be a very useful for providing confirmations or modifications of results obtained using other methods (Dix, Finlay et al. 2004).

3.3.1 Questionnaires

Two questionnaires were used in the project. A general questionnaire was used to collect demographic data about the respondents. This included questions about gender, age, familiarity with retrieval engines, image management and drawing. This was presented to the respondents prior to the experiment session. A second questionnaire was presented to the respondents following the retrieval sessions. This questionnaire was used to determine the respondents' opinions about visual queries, the visual query process and the interfaces and tools available when creating the queries.

The questionnaires were primarily built around scalar questions, in which the respondents were asked to answer a specific question on a numerical scale ranging from 1 to 5. A granularity of 5 was chosen, giving the respondents adequate room to differentiate, while still retaining clarity in meaning. The data from the questionnaires were coded in SPSS[30].

In addition to age and gender, 12 questions were given in the first questionnaire. Some new questions were added to this questionnaire in the second and third experiment. The actual questions used are shown in Table 11. The results of these questions are given in Table 8, page 57.

Table 11 - Questions used in the first questionnaire.

#	Text	Exp 1	Exp 2	Exp 3
1	How much experience do you have with search engines such as Google, MSN or "Kvasir[31]"?	X	X	X
2	How much experience do you have with image search engines, such as Google Images or "Kvasir bildesøk"[32]	X	X	X
3	How much experience do you have with image management systems such as Flickr?	X	X	X
4	How much experience do you have with visual queries - queries where you draw your queries?	X	X	X
5	How often would you say you search for images using different image retrieval engines?	X	X	X
6	When searching for images, is your search private or work related (studies)?	X	X	X
7	How would you rate your own drawing skills?	X	X	X
8	How much experience do you have with drawing on a computer using a mouse?	X	X	X
9	How much experience do you have with drawing on a computer using a digital pen and tablet?	X	X	X
10	How often would you say that you draw?		X	X
11	How often would you say that you draw on a computer using a mouse?		X	X
12	How often would you say that you draw on a computer using a digital pen and tablet?		X	X
13	If you have any prior experience with visual query systems, please list them below.			X
14	If you have any education in visual communication, fine arts or similar studies beyond elementary school, please describe below (College, Bachelor studies or master studies)			X

[30] SPSS is a computer program used for statistical analysis. For more information see http://www.spss.com/.
[31] A Norwegian web search engine, similar to Google
[32] A Norwegian image search image, similar to Google Images

The second questionnaire was designed similarly to the first questionnaire. The questions were related to the visual query process and the search tasks the respondents performed. 24 questions were given. 23 were given on a 5-point scale. One question (question 12) presented the respondent with a set of adjectives. The respondents were asked to mark those words they felt described the visual query process.

New questions were added to the questionnaire in experiments 2 and 3. These questions were added based on the experience from the first experiment.

Table 12 presents the questions used (translated from Norwegian). The three last columns indicate in which experiments the questions were used.

Table 12 - Questions from the second questionnaire

#	Text	Exp 1	Exp 2	Exp 3
1	How well did you enjoy searching for images using Visual Queries?	X	X	X
2	How easy was it to express a search using Visual Queries?	X	X	X
3	How easy did you find it to create visual queries in this [VISI] search tool?	X	X	X
4	How easy did you find it to use a digital pen and tablet for drawing?	X	X	X
5	How satisfied were you with the choice of drawing tools? [VISI]	X	X	X
6	How easy did you find it to draw using freehand? [VISI]	X	X	X
7	How easy did you find it to draw using the predefined shapes? [VISI]	X	X	X
8	How easy was it to understand the concept and use of the weights? [VISI]	X	X	
9	How easy was it to understand the concept and use of the threshold value? [VISI]	X	X	
10	If a system such as this was publically available, how likely is it that you would use it over a text based search?	X	X	X
11	If a system such as this was publically available, how likely is it that you would use it in *addition* a text based search?	X	X	X
12	In the table below, please mark those words you feel best describe image retrieval using visual queries.	X	X	X
13	How time-consuming do you experience this form of image search?		X	X
14	How problematic did you find the time required by this form of image search?		X	X
15	To what degree did you feel that your own drawing skills influenced your ability to create good queries?		X	X
16	To what degree did you feel that the tools available in the interface influenced your ability to create good queries?		X	X
17	How satisfied were you with the choice of colours [In VISI]?			X
18	How easy was it to draw query images in this [Retrievr] search tool?			X
19	How easy was it to draw using the free-hand tool [Retrievr]?			X
20	How satisfied were you with the choice of drawing tools [Retrievr]?			X
21	How satisfied were you with the choice of colours [Retrievr]?			X
22	How well did you like that the results were shown continuously?			X

Table 13 shows the terms used in question 12. The choice of terms was based on the assumptions which underlie the set of hypotheses. This was not meant as an exhaustive list of possible descriptions, and was included as an additional source of data on the respondents' attitude towards visual queries. The respondents were not given the option to add new terms. This was done in order to have a common ground for comparisons of the respondents.

Table 13 - Terms used in questionnaire 2, Q12.

Term	English Term	Norwegian Term
1	Time-consuming	Tidkrevende
2	Quick	Hurtig
3	Enjoyable	Morsomt
4	Difficult	Tungvindt
5	Useful	Nyttig
6	Toy	Leketøy
7	Useless	Unyttig
8	Complicated	Komplisert
9	Usable	Brukbar
10	Easy	Enkelt
11	Effective	Effektivt
12	Efficient	Arbeidssparende
13	Creative	Kreativt
14	Demanding	Arbeidskrevende
15	Boring	Kjedelig
16	Insufficient	Mangelfull

The original questionnaires used are included in Appendix 4 - Data Collection Tools (In Norwegian).

3.3.2 Observation and Think-Aloud Protocol

The think-aloud protocol was used extensively during the experiments. The respondents were encouraged to think-aloud during the query process and were asked questions and asked to elaborate on particular choices made, or actions taken. This was done in order to encourage the respondent to explain particular incidents or special situations, and to provide input for the interview sessions. Audio from the sessions was captured using a digital audio recorder and transcribed.

Observation was used to take paper based notes concerning the query specification and retrieval process. These were used both as a basis for the following interview, and as aids during the later video and audio analysis.

3.3.3 Semi-structured Interviews

The main purpose with interviews was to encourage the respondents to talk about their experiences during the visual query process. Letting the respondents describe these experiences in their own words may provide a "thick description" of the concept of visual query and the process of creating them. A *thick description* represents a description of a concept which not only describes concept itself, but a includes contextualization and interpretation of the concept (Geertz 1972) in (Gentikow 2005).

A semi-structured approach based on a predefined interview guide was used for the interview session. The interview guide was constantly evolved and improved over the course of the different interview sessions and experiments, following the principle of *theoretic sampling*:

> [Theoretic sampling] is a method of data collection based on concepts / themes derived from data. The purpose of theoretical sampling is to collect data from places, people and events that will maximize opportunities to develop concepts in terms of their properties and dimensions, uncover variations and indentify relationships between concepts. (Strauss and Corbin 2008:143)

The interview guide was also supplemented with notes and observations made during the query specification process. The guide was not followed strictly during the interview sessions, but used as support for the researcher to ensure that all relevant topics were covered. Not all questions in the interview guide were used, and follow-up questions were given during the interview. The topics and some sample questions from the interview guide are presented in Table 14. The actual interview guide used is included in Appendix 4 - Data Collection Tools (In Norwegian).

Table 14 - Overview of the interview guide

Theme	Description	Sample Questions
Use of Visual Queries and Query by Drawing	The respondent's attitudes towards using visual queries and Query by Drawing	Is this something you would consider using on a regular basis? How did you enjoy searching for images using this approach? Do you understand why this approach may be useful?
Query specification tools	The respondent's attitudes towards the visual query tools	Were you satisfied with the choice of tools in the interface? How did you enjoy specifying queries using the freehand tool? How easy was it to understand and specify the query parameters (Weights, threshold)?
Major challenges	The major challenges the respondent experienced when using visual queries	How easy or difficult was it to express queries visually rather than using text? What were your major challenges? How did you express complex content, such as actions and interactions?
Suggested changes	Any suggestions the respondent might have for improving or changing the interface or the visual query formulation process	Is there anything that might have made this process easier? Were there any tools you were missing? Do you see any features that could be added to the interface to better enable you to use visual queries?
Expressing Visual Queries	Discussion about the actions and choices the respondent made during the query formulation process.	Can you tell me something about how you proceeded when you created the visual queries? Can you tell me something about the placement of the objects in your queries? Did you consider the "white space" on the canvas as part of your image, or as blank space? Did you have a "mental image" of the how the visual query should appear?

3.3.4 Video Log

Video capturing software[33] was used to capture a movie of all actions performed by each respondent while using the visual query interfaces. The video was synchronized with the audio recording made during the sessions.

3.3.5 Visual Query Images

One of the important goals in the project was to determine the actual structure and composition of the visual query images. Accordingly, it was necessary to analyze these images. For the VISI system, the queries were stored on the server and retrieved after each session. A total of 255 visual queries were created in VISI.

[33] Snapz Pro - http://www.ambrosiasw.com/utilities/snapzprox/

The queries created using the Retrievr system could not be accessed in a similar manner, as these query images were not directly accessible. Images were captured and extracted from the interface videos using screenshots. A total of 157 queries were created in Retrievr.

A total of 414 visual query images were created based on the scenario texts and retrieval tasks. Table 15 presents an overview of the distribution of the queries, based on respondent group, retrieval system and query type.

Table 15 - Overview of query images created.

Group	Overall	VISI	Retrievr	Generic	Narrative	Scenes
IFIM	181	148	33	79	53	49
KHIB	233	108	125	85	91	57
Total	414	256	158	164	144	106

3.4 Data Analysis: Tools and Approach

Three main approaches were used to analyse the data captured during the experiments: Analysis of the visual query process, analysis of the query images and an evaluation of the interview material and the video of the query process. Data from the questionnaires were used as a supplement to these main approaches.

While the data was collected in three separate experiments, it was analysed as a whole. Where possible, the different respondent groups, the two retrieval systems and the three query categories have been compared. Where applicable, SPSS has been used to perform statistical tests on the data material. When different categories of data have been compared, a significance level of 0.01 has been used to determine if the observed results were significant. Unless otherwise stated or defined by the hypotheses, two-tailed hypothesis tests have been used.

3.4.1 Analysis of the Visual Query Process

Analysis of the process of creating the visual image queries was done in order to evaluate several of the research questions. This analysis was primarily based on the interface videos, observation and the transcripts of the think aloud protocol, supported by data from the 2nd questionnaire. This was used to:

- Determine the tools used when creating the visual image queries
- Determine the time spent on creating the visual queries
- Identify break-down situations or other interesting elements of the query process

- Identify interesting observations or remarks made by the respondents during the query process

In order to do this, the interface videos were edited into separate video clips for each of the visual queries. For the queries made in the VISI systems, each video was edited into three shorter clips showing the actions performed in the interface for each query:

1. The actions performed in the visual query specification interface
2. The actions performed in the query parameters interface
3. The actions performed in the result browsing interface

The *Query specification* segment contains all the actions performed in the drawing interface when creating the visual query image. The length of this clip was used as a measurement of the time spent expressing the visual query. This segment was also used to identify and classify the different tools used to create the visual queries. This was performed by analyzing each video noting which interface tools were used to create the query.

The *Query parameters* segment contains all the actions performed when the respondents defined the query parameters in the query specification interface, i.e. setting the weights and the threshold value for the query.

The *Result Browsing* segment contains all the actions performed by the respondents when browsing through the results of the search. This segment was primarily used to record the respondent’s comments on the CBIR process.

Editing the videos from the Retrievr retrieval process proved somewhat more difficult than the VISI videos. As a result of the interactive nature of the system, it was sometimes difficult to determine when the actual search ended, as some of the respondents refined their image several times.

The respondents were asked to press the “clear” button when they had finished each task. This separated the individual tasks, and the video of each query was edited from the frame when activity began on the query (e.g. selection of a colour, choosing a pen size or starting to draw), and cut on the frame when activity in the interface ended. The seconds spent on this represents the time spent on each query.

The results from this analysis (Measurements of time and classification of tool use) was quantified and plotted into SPSS. The video and audio from the query process were also used as an additional data source for the qualitative analysis.

Note that because of corruption in some of the videos, only 379 (92%) of the queries were timed. Table 16 shows the number of queries timed for the two interfaces, the two respondent groups and the three query categories.

Table 16 - Queries timed

Category	IFIM	KHIB	VISI	Retrievr	Generic	Narrative	Scenes
Queries Timed	155 (86%)	233 (96%)	221 (86%)	158 (100%)	144 (87%)	136 (94%)	99 (95%)

3.4.2 Analysis of the Query Images

In order to analyze the query images, a framework for evaluation of visual queries expressed through QBD was developed. The framework and the use of the framework are described in chapter 4.

All query images were classified by the primary researcher. In order to ensure the quality of the classification work, two external evaluators were given a random sample of the query images and asked to classify these images based on the framework. The evaluators were presented with both a written presentation of the framework (included in Norwegian in Appendix 4 - Data Collection Tools) and a discussion concerning the framework prior to the classification.

The first evaluator (Female, 35) had no formal education or background in visual arts or related fields. She evaluated a random sample of 100 query images, selected from all three experiments. The second evaluator (male, 34) holds a degree in visual communication from the Bergen Academy of the Arts. He was not involved in any of the experiments. He evaluated a random sample of 100 query images, selected from all three experiments.

The classification performed by the external evaluators was compared to the classification done by the researcher. In most cases, there were small differences between the three evaluations. There were some cases there were significant differences between the evaluators. These differences were discussed with the evaluators. Some images were reclassified based on these discussions. In cases where there appeared to be structural differences in the evaluations, these are mentioned and discussed in the analysis.

The results from this analysis were quantified and plotted into SPSS for subsequent analysis and comparison between the two respondent groups, the two interfaces and the different query categories.

3.4.3 Questionnaires

Finally, the two questionnaires were used to support the other data. Data from both questionnaires were coded in SPSS and used for reference and in analysis of the different research questions. Only the questions that describe interesting elements of the analysis have been included in the evaluation.

3.4.4 Qualitative Analysis and Grounded Theory

The research questions related to the respondents' behaviour and experiences when using QBD were evaluated using a qualitative approach based on grounded theory (Strauss and Corbin 2008). The primary data source used in this approach was the semi-structured interview sessions. However, the questionnaires, the observation notes, the video files and the actual query images were used to support the interview sessions. This approach was used to:

- Support and interpret the findings and results obtained in the analysis of the query process and the query image classification (RQ1 and RQ2).
- Determine the challenges facing the users (RQ3)
- Determine how the users feel towards using visual queries (RQ 4)
- Determine any potential improvements the users would like to see (RQ5)

The interviews were recorded on a digital audio recorder and transcribed into text files. Most of the transcription work was performed by a research assistant. The quality of this work was ensured through random sample of comparisons between the audio log and the transcribed text. The interviews with participants 16, 17, 18 and 19 were transcribed by the researcher. During the transcription process, the major questions were numbered and added to the transcript, and each of the statements made by both the interviewer and the interviewee were numbered.

The transcribed interview sessions were processed, structured and imported into QSR Nvivo 8[34], a software package for processing and analysing qualitative data. An adaptation of grounded theory was used during the analysis of the interview data. In its purest form, grounded theory is used to construct theory without any prior categories, hypotheses or described framework (Strauss and Corbin 2008). The approach is based on discovering concepts and categories from the empirical data, and using these to construct theory. However, in this project a set of hypotheses regarding visual queries were explicitly stated and were used to provide an initial set of concepts and categories for the data analysis. As such, this method does not adhere strictly to the standards of grounded theory.

[34] http://www.qsrinternational.com/products_nvivo.aspx

None-the-less, the approach used for the interview sessions and the following analysis work was based in methodologies from grounded theory.

All discussions and interviews were performed in Norwegian, and the transcripts were written in Norwegian. The language transcripts were not normalized, and the language, grammar and sentence structures in the transcripts were kept as close to the oral source as possible.

Excerpts from these transcripts are used extensively in the following chapters. These excerpts have been translated by the author. Because these had to be translated, the excerpts presented as illustrations have been normalized, and should be considered paraphrases and not direct quotes. This rewording and paraphrasing was done in order to improve the readability of the text, but every effort has been taken to preserve the original meaning of the text. Furthermore, some of the respondents used a high degree of profanity in their statements. As translating profanities between different languages is very difficult, these have been removed from the paraphrases.

The excerpts and paraphrases included in the text have been chosen for their strength as illustrations of how the respondents have voiced opinions and expressed themselves about the different topics. Unless otherwise mentioned, these are not the only statements concerning these topics, but were chosen as the most descriptive and readable of the statements. In most cases, the number of respondents who have made similar statements is not directly mentioned, and unless otherwise stated it may be assumed that the included paraphrases reflect opinions that are shared by several respondents or all the respondents in a particular group.

4 A Framework for Visual Query Image Classification

An important goal in this project was to evaluate *how* users compose visual image queries, and to determine if there are any variations in the way users compose queries based on factors such as background, image retrieval task and variations in the query interface. In order to achieve this goal, a framework for evaluating the visual queries was required. This framework should:

1. Provide a set of precisely defined concepts which can be used to evaluate the hypotheses
2. Provide a set of tools capable of evaluating these concepts

The concept of *visual modality* was chosen as the foundation of the framework. The term 'modality' comes from linguistics and refers to the truth value or credibility of statements about the world. Kress and van Leeuwen (2006) discusses the concept of modality applied to images, *visual modality*. For the purposes of this thesis, *visual modality is defined as the degree to which an image represents a naturalistic rendition of the concepts depicted in the image (Definition 13).* In order to quantify this concept, Kress and van Leeuwen present a number of *modality markers*, i.e. visual indicators which can be used to determine the visual modality of a an entire query image or an element of the image.

There are several reasons for why this approach was used. A substantial number of existing approaches for CBIR are based on a direct comparison between a visual query image and a set of images. As these comparisons are based on mathematical measurements of feature similarity, the achieved similarity score is based on the similarity of the compared features. Consider the two images in Figure 26. Both are photographic images depicting a single seagull, and while they depict two different scenes, they are relatively similar with regards to representational and compositional structures.

Figure 26 - Two representations of a "Seagull"

Next, compare these to the three images represented in Figure 27. These are drawings made in a visual query interface, and represent seagulls in various forms. First of all, even though most observers would identify these as birds, or even seagulls, they do not have a very high similarity to either of the images in Figure 26. The obvious difference between the images is that the images in Figure 26 are photographs, while the images in Figure 27 are drawings. Next, while all the images in Figure 27 are drawings, there are a number of differences between them. Figure 27a and Figure 27b are created as black and white drawings, while Figure 27c includes the colours yellow and grey in addition to black and white. Finally, Figure 27b is likely to be considered a more realistic depiction of a seagull than the other drawings. A retrieval system based on direct similarly comparisons between the visual query images of Figure 27 and a collection of visual images including those of Figure 26 is unlikely to achieve a high level of successful retrieval.

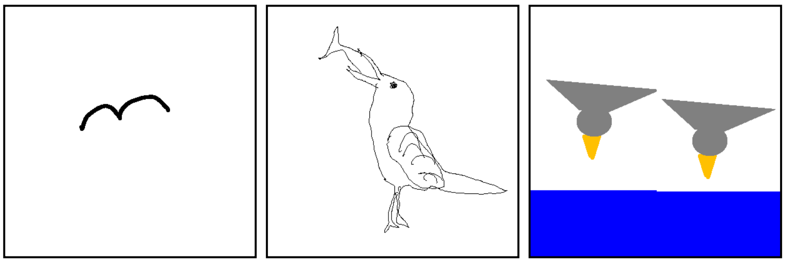


Figure 27a, b and c - Three visual queries for a seagull[35].

It has been important to determine if users draw query images in a manner that a CBIR system based on direct comparison is capable of interpreting and processing. It is assumed that a query image with a *high visual modality*, i.e. drawn in a realistic manner, will have a higher similarity to a photograph than a query image with a *low* visual modality. Consequently, using visual modality was found to be a suitable approach for evaluating the query images drawn by the users.

Understanding how users compose and create these images might be relevant for understanding *how* QBD queries could be made a more useful tool for expressing visual image queries. This understanding could also be used to improve a CBIR system's ability to respond to QBD queries. The framework provides tools for deconstructing the query images into several basic elements that could be analyzed, compared and evaluated.

[35] The visual queries created during this project are frequently displayed in the following text. Unless otherwise noted, the queries have been resized (scaled down) and adapted to the text. A frame has been added to the images in order to illustrate the borders of the image. If the white space in the query images have been cropped in order to fit the text, this have been explicitly stated. In these cases, the frame is not included.

4.1 The Framework

The framework for visual modality presented in Kress and van Leeuwen (2006) is a generic framework that could be applied to *any* type of image. A pilot evaluation using the framework directly on the visual query images created in this project was performed. This evaluation showed that using the framework directly as it is described in Kress and van Leeuwen (2006) would not be a good framework for evaluating drawn visual query images. While it would allow for comparisons between images drawn as visual queries and other types of images, it would not provide enough details to identify and categorize differences *between* different types of drawn queries. Consequently, it was decided to create a new framework for analyzing and categorizing images drawn as visual queries, *based* on the visual modality framework described in Kress and van Leeuwen (2006).

This customized framework consisted of four main components:

1. A classification scheme for drawn query images based on four customized modality markers
2. A method for personal evaluation of query image modality based on these modality markers
3. The number of unique objects present in a visual query image
4. The number of colours used to create a query image

The four components are detailed in the following section. An example of how the framework is used is presented in Figure 43 (page 93).

4.1.1 Modality Markers

In this framework, a modality marker is defined as an indicator used to determine the visual modality of a query image (Definition 14). Four such modality markers were used:

1. **Contextualization:** The use of contextualization elements in an image
2. **Colours:** The use of colours in an image
3. **Representation:** The degree of abstraction used in an image
4. **Composition:** Different compositional effects used in an image

For each modality marker, a set of criteria were determined. A criterion is a condition that may be satisfied in a visual query image. Ideally, these criteria should be distinct and mutually exclusive, i.e. for a given query image, a certain criterion is either satisfied not satisfied. However, the different visual elements in an image might have different levels of modality: One element of an image might satisfy a criterion, while another element in the same image might not satisfy the same criterion. Consequently, with one exception (monochromatic), the criteria are not necessarily mutually exclusive within a given image. The modality markers and the criteria are presented in the following sections, and an overview of the markers and criteria is presented in section 4.1.1.5.

4.1.1.1 Use of Contextualization

Contextualization represents the degree of *completeness* in a query image, e.g. to what degree are background and contextual elements used to give a context to the subject matter of the query. Four contextual criteria were used, presented in Table 17:

Table 17 - Contextualization Modality Criteria

Criteria	Conditions for satisfaction
Participant	A visual element representing the subject matter of the query is included in the query image
Background	Inclusion of background other than a "neutral" background in the query image
Symbolic contextual elements	A visual element with strong contextual signifiers is included in the query image
Minor contextual elements	A visual element without strong contextual signifiers, but which nevertheless may be present in a realistic image, is included in the image

Three elements are likely to be present in a visual query image: *background*, *participants* and different *contextual elements*. A visual query may include one or more participants, e.g. visual elements representing the subject matter of the query. An example of this is the running person in Figure 28a - a query based on task 26 - Find images of humans practicing sports[36]. However, it is possible that a query does not include any objects of interest, as shown in another query based on the same task (Figure 28b). This query consists of a single green area, possibly representing a football field. Note that while "objects of interest" represent the subject matter of a query and not context, it is included in this marker as it is closely related to the criteria for contextualization.

[36] The actual image retrieval tasks used in the project are described and detailed in Table 10, page 68

Figure 28a, b - Illustrations of the use of objects of interest (Queries# 304 and 181).

Next, a query may or may not contain a background. The running person in Figure 28 is an example of a query without background. This query has no other content than the object of interest. Consequently, *background* describes the inclusion of background other than a "neutral" canvas.

Finally, a query may consist of one or more contextual elements: an element that is not the subject matter of the query, but provides contextualization for the object or objects of interest.

An illustration of the use of contextual elements is shown in Figure 29. Both queries are based on image retrieval task #2 - Find images of a scuba diver. The first image (Query # 51) only contains the diver, while query #122 also contains a fish and an underwater plant. These provide contextualization for the scuba diver.

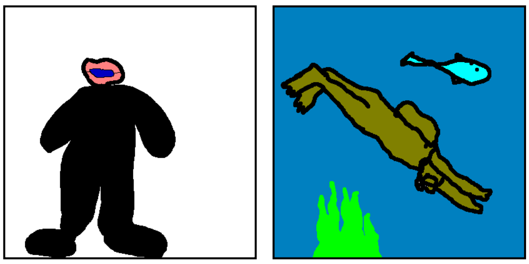


Figure 29a, b - Illustration of contextual elements (Queries# 51 and 122). The images have been resized.

Contextual elements may be either symbolic or minor. A symbolic contextual element is a visual element with strong contextual signifiers, while a minor contextual element is a visual element without strong contextual signifiers, but which nevertheless may be present in a realistic image.

An example of a *symbolic contextual element* can be the inclusion of "the sun" to represent either that the query represents "outside" or "a sunny day", or a straight or curved line to indicate the surface of the sea. An example of a *minor contextual element* may be the inclusion of small rocks and stones on a query image containing a beach.

4.1.1.2 Use of Colours

This modality marker combines the different colour markers presented in Kress and van Leeuwen (2006:p160). It describes the degree to which colour is used to create a realistic image. Five modality criteria were used to describe colours, presented in Table 18:

Table 18 - Colour Modality Criteria

Criteria	Conditions for satisfaction
Monochromatic	The image is created exclusively using a single colour on a neutral canvas. This criterion is mutually exclusive with the other colour-based modality criteria.
Basic colour use	One or more objects in the image are represented by a single colour. Different objects may have different colours.
Varied colour use	One or more objects in the image are represented by more than a single colour.
Colour gradients	One or more objects in the imaged is coloured using colour gradients.
Illumination	The use of implicit or explicit light sources to create variances in brightness, shadows or other effects of light in an image.

First of all, some images may be drawn using one single colour, i.e. they are monochromatic: created exclusively using a single colour on a white canvas. In most definitions, monochrome includes the use of different shades of a single colour. However, for the purposes of this framework, monochromatic represents the use of a single colour, i.e. the use of black lines on a white canvas in query 141 (Figure 30).

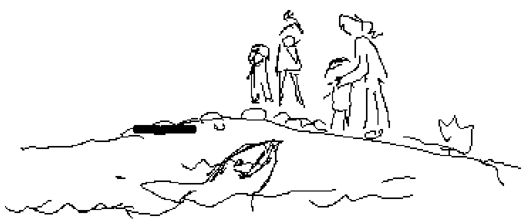


Figure 30 - A monochrome drawing of humans interacting with a dolphin (Query #141). White space has been cropped from the borders of the image.

Images may be created using different combinations of colour. In this framework, three criteria for colour use have been established. First, images can be created with their elements represented by a single colour, such as the dolphins and icebergs drawn in query 201 (Figure 31). While each object is represented using only one colour, the image as a whole uses more colour than the monochromatic image in Figure 31. This is defined as basic colour use: One or more objects in the image are represented by a single colour.

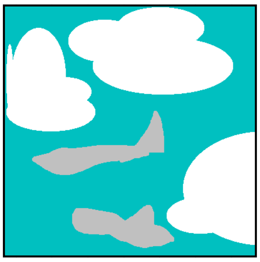


Figure 31 - Dolphins and icebergs represented using single colours (Query #201).

Note that some objects are coloured in a single colour, but also contain an outline (usually black). An example of this is the boat and the dolphin in query 49 (Figure 32). In the evaluation, these objects have been considered as represented with *basic colour use*.

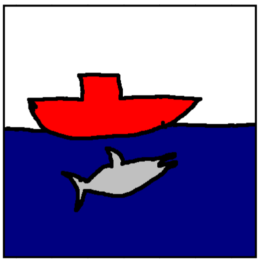


Figure 32 - Illustration of the use of contour lines. (Query #49).

Furthermore, the objects in an image can be created using more than one colour. The flower in query 302 (Figure 33) is composed of three colours. This represents varied colour use: One or more objects in the image are represented by more than a single colour.

Figure 33 - A flower created using different colours (Query #189).

It might also be possible to create images using colour gradients. Colour gradient represents the transition between one colour and another, i.e. the way the sky gradually transforms from dark blue to a lighter blue in Figure 34. In this framework, this is defined as the use of colour gradients: The background or one or more objects in the image is represented using colour transitions.

Figure 34 - Illustration of colour gradients[37]

Finally, it is possible to create images using one or more implicit or explicit light sources which provide highlights, shadows, play of light or other variances in brightness. This is defined as illumination: The use of implicit or explicit light sources to create variances in brightness, shadows or other effects of light in an image.

4.1.1.3 Representation

All images and drawings represent some degree of abstraction from the real world. *Representation* describes the process of simplifying a visual object, from a completely realistic representation to a simpler representation, while still retaining a connection to the original object. An example is the two seagulls depicted in Figure 35.

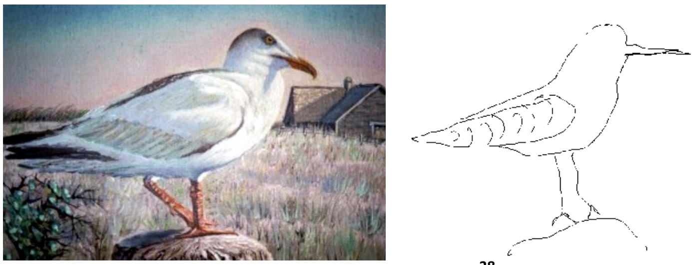


Figure 35 a, b - Two depictions of a seagull. "*Oil painting of a seagull*"[38] and visual query for a seagull (query# 32). Whitespace has been cropped from the query image.

For the purposes of this framework, four representational criteria have been selected, represented in Table 19:

[37] Photograph by Lars-Jacob Hove.

[38] By Giovanni Ramacciato. Image obtained from Flickr (http://flickr.com/photos/elmolise/219974883/). Some rights reserved.

Table 19 - Representational Modality Criteria

Criteria	Conditions for satisfaction
Geometric primitives	simple two- and three-dimensional non-representational forms, such as lines, arcs, squares and circles have been used to create at least one object in the query image
Outlines	a simple representation of the basic shape of an object has been used to create at least one object in the query image
Symbolic representational components	Visual elements that represent visual cues which are of high symbolic value when identifying an object have been included in at least one object in the query image
Detailed representational components	Visual elements that do not have a high symbolic value, but provide a higher degree of realism to an object in the image, have been used in at least one object in the query image
Texture	The use of patterns to illustrate the surface structure of at least one object in the query image

First of all, the most basic way of drawing an object is through the use of geometric primitives, e.g. using simple two- and three-dimensional non-representational forms, such as lines, arcs, squares and circles (Eakins, Burford and Briggs 2003). *Geometric primitives* represent the use of these forms used to describe the use of these primitives to represent one or more image participants. This includes the use of simple shapes to represent objects such as the two dolphins and a boat in Figure 36a, and the use of a combination of lines and circles to represent objects as stick figures, such as the seagull in Figure 36b or the humans in Figure 36c.

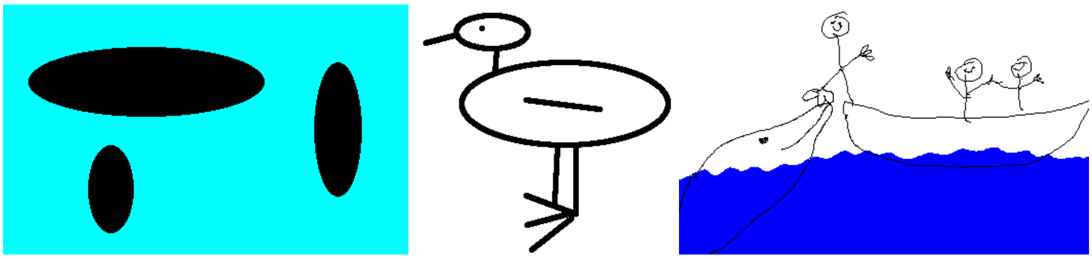


Figure 36 - An example of use of geometric primitives (Queries# 15, 11 and 68). The images have been cropped to the main motives.

Next, objects can be represented through the use of outlines: a simple representation of the basic shape of an object. The whale and the football in Figure 37 are two examples.

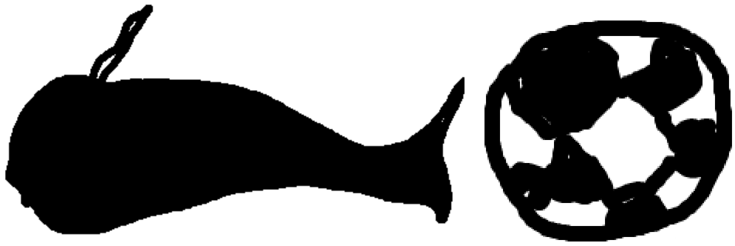


Figure 37 - Illustration of the use of outlines. (Queries# 156 and 323). The images have been cropped to the main motives.

When determining whether a particular outline or a set of geometric primitives represents an amoeba, a human, a ship or a dolphin, one or more visual cues, or *representational components*, might be present to help distinguish and identify the shape. Two examples are the eye of the dolphin in Figure 36c and the black and white areas of the football in Figure 37b. The framework distinguishes between two categories of representational details. Symbolic representational components represent visual cues, which are of high symbolic value when identifying an object. Examples of such components are the inclusion of a pair of human eyes or limbs, a dolphin's fin or beak, or a pair of sails on a boat. Next, detailed representational components represent components that do not have a high symbolic value, but provide a higher degree of realism to a participant in an image. Examples of such components are individual strands of hair or fur, leaves on a tree or the hinges of a door.

Finally, an object might be drawn in plain colour, or texture could be used to signify the object's surface structure. *Texture* refers to the use of patterns to illustrate the surface structure of an object.

4.1.1.4 Composition

Finally, the elements of an image will always be structured in a certain way. This is the *composition* of the query image: the organization or grouping of the different elements in a query so as to achieve a unified whole. Three criteria were defined for this framework, represented in Table 20:

Table 20 - Compositional Modality Criteria

Criteria	Conditions for satisfaction
Realistic scaling	Two objects in the query image are represented in a realistic scale to each other
Overlap	One or more objects in the query image are occluded by the overlapping of another participant
Central perspective	Central perspective has been used to compose the image

First of all, if more than one participant is represented in an image, there may be an element of scaling between them. The participants may be realistically scaled, e.g. a realistic scale has been used to represent the relative size of two or more participants in the query image, or the scaling may *not* be realistic, e.g. the relative size is not represented in a realistic manner.

Next, a common way of illustrating depth in an image is to use *overlapping participants*: Distance, depth or scale is illustrated through the occlusion of one participant by another participant. Overlapping can be used to represent the order of objects, as well as scale and distance, as seen in query #287 (Figure 38). The two dolphins overlap the stand and people observing the dolphins. Another example is query #4 (Figure 39), in which the dolphin and the persons are overlapped by the sea. In combination with the use scaling, this indicates that the two participants are placed equally distant from the observer.

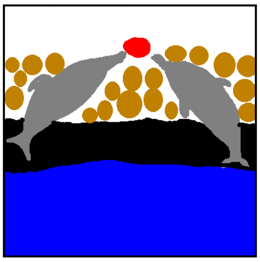


Figure 38 - Illustration of overlapping elements (Query #287).

Figure 39 - Illustration of overlapping elements (Query #4).

In query 5 (Figure 40) overlapping has not been used. It is difficult to determine whether the whale is above or below the surface of the sea, and the boat appears to be floating on top of the water. As a result it is difficult to determine the actual scale and order of the two elements.

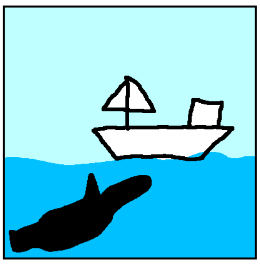


Figure 40 - Illustration of non-overlapping elements (Query #5).

Finally, according to Kress and van Leeuwen (2006), by criteria of standard naturalism the highest possible modality for an image is represented by the central perspective. Central Perspective is a technique used within art to suggest distance by having all parallel lines running through the image perpendicularly to the motive, converge centrally at height with the observer's viewpoint. This criterion represents the use of central perspective in order to represent depth, size and scaling of the depicted participants.

4.1.1.5 Overview of Modality Markers and Modality Criteria

Table 21 presents an overview of the modality criteria of the four modality markers used in the framework, along with the conditions required to classify visual query images by the criteria.

Table 21 - Overview of modality markers and modality criteria

Modality markers	Criteria	Conditions for satisfaction
Contextualization	Participant	A visual element representing the subject matter of the query is included in the query image
	Background	Inclusion of other background other than a "neutral" background in the query image
	Symbolic contextual elements	A visual element with strong contextual signifiers is included in the query image
	Minor contextual elements	A visual element without strong contextual signifiers, but which nevertheless may be present in a realistic image, is included in the image
Colour	Monochromatic	The image is created exclusively using a single colour on a neutral canvas. This criterion is mutually exclusive with the other colour-based modality criteria.
	Basic colour use	One or more objects in the image are represented by a single colour. Different objects may have different colours.
	Varied colour use	One or more objects in the image are represented by more than a single colour.
	Colour gradients	One or more objects in the imaged is coloured using colour gradients.
	Illumination	The use of implicit or explicit light sources to create variances in brightness, shadows or other effects of light in an image.
Representation	Geometric primitives	Simple two- and three-dimensional non-representational forms, such as lines, arcs, squares and circles have been used to create at least one object in the query image
	Outlines	A simple representation of the basic shape of an object has been used to create at least one object in the query image
	Symbolic visual elements	Visual elements that represent visual cues which are of high symbolic value when identifying an object have been included in at least one object in the query image
	Detailed visual elements	Visual elements that do not have a high symbolic value, but provide a higher degree of realism to an object in the image, have been used in at least one object in the query image
	Texture	The use of patterns to illustrate the surface structure of at least one object in the query image
Composition	Realistic scaling	Two objects in the query image are represented in a realistic scale to each other
	Value scaling	One or more objects in the query image are occluded by the overlapping of another participant
	Overlap	Central perspective has been used to compose the image
	Central perspective	Two objects in the query image are represented in a realistic scale to each other

4.1.2 Personal Evaluation

While the modality markers and their criteria provide an opportunity to identify how the query is composed, and the different methods used to create the visual query image, they do not necessarily present a description of the overall modality of query image. In order to obtain a measurement of this, each modality marker was evaluated based on a personal judgement of the evaluator. Each marker will be rated on a score from 1 (very low modality) through 5 (very high modality). This will allow an evaluator to classify the images based on his or her own background. It should be noted that this measure will be *subjective* by nature.

4.1.3 The Number of Individual Objects

In addition to determining whether a given query image contains either objects of interest, contextual elements or both, it may be interesting to know the number of individual elements present in a query image. An individual element is a visual element that represents a unique interesting element or contextual element. For example, a the disembodied head in Figure 41a would be one individual object, while the head, torso, oxygen tank and feet of the scuba diver in Figure 41b person might be considered a single object, even if they were not drawn completely together.

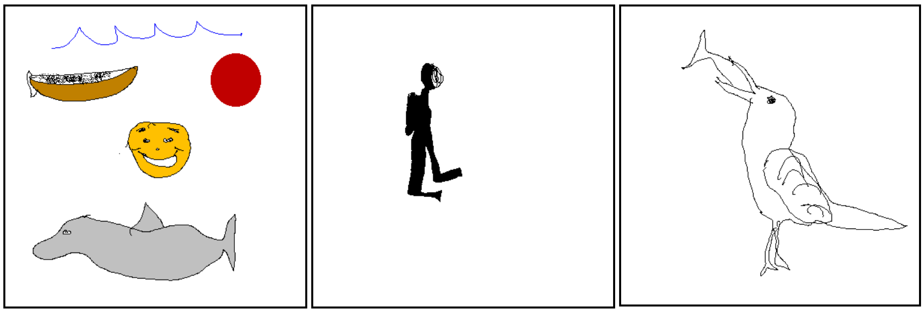


Figure 41a, b, c - Illustrations of the counting of image participants (Queries# 113, 51 and 33)

The image of a seagull eating a fish in Figure 41c (Query 33) contains 2 individual objects (Seagull and fish) even if they are connected in the drawing.

It should be noted that this measurement *should not* be used as the primary measurement of query completeness, but used in support with other measurements. Some images may be complete even if there is only a single object present, illustrated by the image of the Statue of Liberty shown in Figure 42. In this example, the image consists of a single object, without any contextual details.

Figure 42 - An image of the Statue of Liberty[39]

4.1.4 Number of Colours

The colour modality criteria describe whether colours are used in an image. However, they do not provide any data on *how many* colours are used in an image. Counting the number of unique colours present might also provide additional information about the complexity and modality of a given query image.

4.1.5 Evaluation Complexity

This framework presents a method for evaluating a set of query images. However, in addition to determining each of the modality markers, it would be of interest to identify any images that were difficult to evaluate. This will provide both a way to test the strength of the framework, as well as a way to identify and classify images that might represent outliers or special cases.

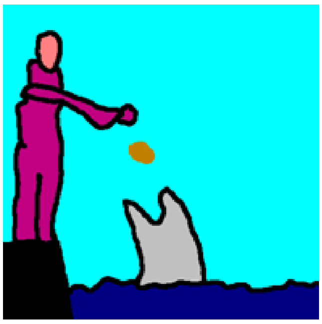
In order to provide for this, the evaluator should rate the complexity of categorizing each image on a scale from 1 (very easy) to 5 (very difficult). Finally, the evaluator should take notes of any problematic cases encountered when performing the evaluation.

4.2 Using the Framework

Each query image was entered on a custom form, shown in Figure 43. The original form was created in Norwegian. The text in Figure 43 has been translated. The original form is available in Appendix 4 - Data Collection Tools (In Norwegian).

[39] Image retrieved from *Featherboa*'s flickr profile (http://www.flickr.com/photos/featherboa/43040507/). Some rights reserved.

Query number 127



Query details
Query number: 127
Participant: 16
Participant query: 6
Seconds: 142
Colours: 8

Query task details
Query level: Level 2
Interface: Visi

Query task:
Retrieve images of a human feeding a dolphin

Tool use

Freehand:	*YES*	*NO*
Shapes:	*YES*	*NO*
Colours:	*YES*	*NO*
Texture:	*YES*	*NO*

Unique objects __3__

Classification of modality markers – Modality criteria

Colour:	Monochrome	Basic	Varied	Modulated	Illumination
Context:	Objects	Background	Symbolic	Minor	
Representation:	Primitives	Outline	Symbolic	Detailed	Texture
Depth:	RelScape	ValScale	Overlap	Central perspective	

Classification of modality – Qualitative evaluation

Colour:	1	2	3	4	5
Context:	1	2	3	4	5
Representation:	1	2	3	4	5
Depth:	1	2	3	4	5

Use of shapes:

Defining major areas :	*YES*	*NO*
Drawing geometric shapes:	*YES*	*NO*
Representing other objects	*YES*	*NO*

Evaluation complexity	1	2	3	4	5

Notes:

Uncertain if the rock the person is standing on should be counted as background or a separate element. I've counted it as a part of the background, while the person, the fish (?) and the dolphin have been counted as separate objects, resulting in 3 objects. The ocean was considered as background. The outline of both the human and the dolphin has been disregarded when evaluating colour use.

Figure 43 - Illustration of the evaluation form used for the framework

5 The Query Formulation Process

The first study objective was the *query formulation process*, e.g. *how* the respondents used the query interfaces when drawing query images, as expressed in **research question 1**:

> *How do users utilize the visual query interface when they draw visual queries?*

Four main research hypotheses have been evaluated to address this question:

- **RH1.1:** Respondents make frequent use of graphical drawing tools rather than drawing by freehand
- **RH1.2**: Respondents prefer the query interface provided by VISI to the query interface by Retrievr
- **RH1.3**: Respondents draw query images more quickly in the interface provided by VISI than in the query interface provided by Retrievr
- **RH1.4:** Respondents with a visual background express queries faster than respondents without this background

Graphical drawing tools represent predefined shapes such as *lines*, *circles*, *squares* and *polygons*.

It should be noted that it is not the two interfaces that have been evaluated, but the user experience *with* these interfaces.

These research hypotheses are detailed and operationalized in the following sections. Section 5.1 explores the respondents' use of and opinions of the two interfaces (RH1.1 and RH 1.2), while section 5.2 discusses elements related to the *time* the respondents spent expressing the queries (RH 1.3 and RH 1.4). A summary of the results is presented in section 5.3.

5.1 Respondent Use of the Query Interfaces

The two retrieval systems evaluated in this project represented different approaches for query by drawing. There were differences in the tools available, the colours available, the size of the canvas, and in the general dynamic of the query process[40] . Each of these aspects of the query interfaces has been evaluated in detail and is discussed in the following sections, with a particular focus on RH1.1 and RH1.2:

---

[40] See Table 9, page 64 for an overview of the differences

- **RH1.1** Respondents make frequent use of graphical drawing tools rather than drawing by freehand
- **RH1.2:** Respondents prefer the query interface provided by VISI to the query interface by Retrievr.

Note that the VISI system was at a prototype stage when the study was conducted, and several issues with usability and functionality were discovered during the evaluations. However, when evaluating the questionnaires, the interview sessions and the observation logs, the main focus was on identifying general observations and issues with the *approaches* represented with the interfaces, not focusing on usability issues directly related to the interfaces.

5.1.1 Tools

The two interfaces offered the respondents different tools for drawing the query images. It was hypothesized that the respondents would prefer to express queries in VISI, primarily because of the additional tools available. Two measurements were used to determine the respondents' tool preference:

1. Questionnaire questions related to tool use
2. Classification of tool usage in the two interfaces

In addition, the respondents' use of tools was a major topic in the interview session, and data from these sessions were used to support and examine these measurements in detail.

The first measurement was constructed from the answers from the second questionnaire related to the tools available in the two interfaces:

- Question 5 ("How satisfied were you with the choice of drawing tools [In VISI]?")
- Question 6 ("How easy did you find it to draw using freehand [In VISI]?")
- Question 7 ("How easy did you find it to draw using the predefined shapes [In VISI]?")
- Question 20 ("How satisfied were you with the choice of drawing tools [Retrievr]?")

Table 22 presents descriptive data of these questions. Questions 5, 6 and 7 were asked in all three experiments (N=30), while question 20 was only asked in the third experiment (N=12). A score of 1 indicated "very dissatisfied" or "very difficult" while 5 indicated "very satisfied" / "very easy".

Table 22 - Respondent satisfaction with tools in VISI and Retrievr

Q	Min	Max	Mean	Std. Dev.
5	2	5	3.40	0.855

6	1	5	2.83	1.053
7	1	5	3.00	1.592
20	1	5	2.25	1,357

There were no significant differences in response between the two respondent groups. Concerning questions 5 and 20, there was a difference in the answers to the two questions, indicating that the respondents were generally more satisfied with the tools available in VISI than in Retrievr. However, the low number of respondents, particularly in question 20, implies that these results should not be given much weight.

Next, concerning questions 6 and 7, these indicate that the respondents found it easier to draw using the drawing tools than using freehand drawing (In VISI), but this was a very small difference.

The second measurement was the actual *use* of tools in the two interfaces. In VISI, the respondents could choose between *drawing by freehand* or use the *drawing tools*. *Freehand drawing* represents use of one of the pen tools to freely draw, or sketch, the query image. *Drawing tools* represents the use of one of the graphical tools provided by the interface, e.g. lines, boxes, squares or circles. The supporting tools (Colour filler, duplicator and eraser) were not included in this category.

Overall, freehand drawing was used in 95.7% of the queries while drawing tools were used in only 29.7%. Freehand was used alone in 81.6% of the queries, drawing tools were used alone in 2.7% of the queries, and the combination of freehand and tools was used in 15.7% of the queries. Table 23 presents an overview of tool usage broken down by respondent group and query category. The upper part shows the overall use of the techniques, while the lower part shows the combinations of techniques:

Table 23 - The respondents' use of drawing techniques in VISI.

		Group		Query type		
Drawing technique	Overall	IFIM (148)	KHIB (108)	1 (108)	2 (87)	3 (61)
Freehand	95.70 %	95.90 %	95.40 %	93.50 %	98.90 %	95.10 %
Drawing tools	29.70 %	31.80 %	26.90 %	26.90 %	21.80 %	45.90 %
Only freehand	81.60 %	68.20 %	73.10 %	71.50 %	78.20 %	54.10 %
Only tools	2.70 %	4.10 %	4.60 %	6.50 %	1.10 %	4.90 %
Both freehand and tools	15.70 %	27.70 %	22.20 %	20.40 %	20.70 %	41.00 %

These results show that contrary to RH1.1, freehand drawing was the most used drawing technique, while tools were only used in less than a third of the queries. There were no significant differences between the two respondent groups[41]. However, a comparison of the respondents who rated themselves with “high drawing skill” or “very high drawing skill” with the other respondents shows that the former group used drawing tools significantly less[42] than the latter group, while there was no difference in their use of freehand drawing, indicating that the respondents who rated themselves as good drawers used less drawing tools than the other respondents.

The use of freehand drawing was high for all three query categories, while the use of drawing tools increased in category 3 queries. There was a significant difference between category 1 and 3[43] and between category 2 and 3[44].

Table 24 presents *which* tools the respondents used in the 76 VISI queries created using drawing tools. The column “Tool #” reflects the sequence of the tool (from left), shown in Figure 44. The upper half of the table shows the drawing tools, while the lower half shows the supporting tools. There were no significant differences in use of the drawing tools between the two respondent groups or the three query categories.

Table 24 - Overview of tool use in VISI

Tool	Tool #	Use (N=76)
Filled square	10	35,50 %
Line	2	26,30 %
Filled circle	12	25,00 %
Circle	7	23,70 %
Polygon	13	17,10 %
Filled rounded square	11	3,90 %
Square	5	1,30 %
Rounded square	6	0,00 %
Arrows	3,4	0,00 %
Fill tool	14	55,90 %
Eraser	13	0,15 %
Copy tool	8	0,01 %

41 Freehand: Mann-Whitney U [256] = -0.224, $p > 0.01$. Shapes: Mann-Whitney U [256] = -0.847, $p > 0.01$
42 Mann-Whitney U [256] = -1.945, $p < 0.01$
43 Mann-Whitney U [256] = -0.412, $p < 0.01$
44 Mann-Whitney U [256] = -1.387, $p < 0.01$

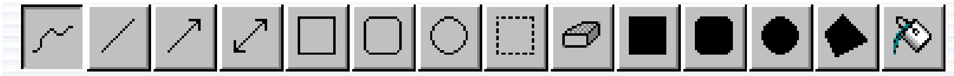


Figure 44 - Illustration of VISI drawing tools. The first tool is "Freehand drawing".

Concerning the supporting tools, only the fill tool ("Fill bucket", tool 14) was used to any degree. There were no significant differences between the respondent groups. There were some differences between the three categories; tools were used significantly more in queries for *complete scenes* (75.5%) than in queries for *generic objects* (40.7%) and queries for *narrative content* (50.5%). This difference was most likely caused by the increased level of *completeness* of these queries. Queries for *complete scenes* were generally more *complete* than the other queries, i.e. they were created using more *contextual details* and more *background* (discussed in detail in chapter 6.1). As queries of category 3 often included background or other larger areas that had to be coloured than queries of categories 1 and 2, the fill tool was used more often.

VISI had 12 different pen sizes (or pen splits) available for use with the different drawing tools, illustrated in Figure 45 (rotated 90 degrees).

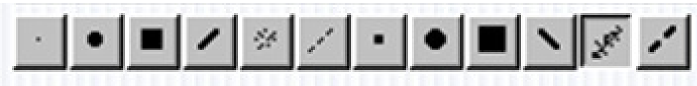


Figure 45 - Different pen splits available in VISI. The toolbar has been rotated 90 degrees.

Table 25 shows the use of these pens in the VISI queries. Only two of the tools saw any considerable use: *The medium circle* (Pen # 2) and *the point* (Pen # 1). Figure 46 shows two queries for "Scuba Diver", the leftmost made entirely using the "point" pen while the rightmost was made using entirely using the "medium circle" pen.

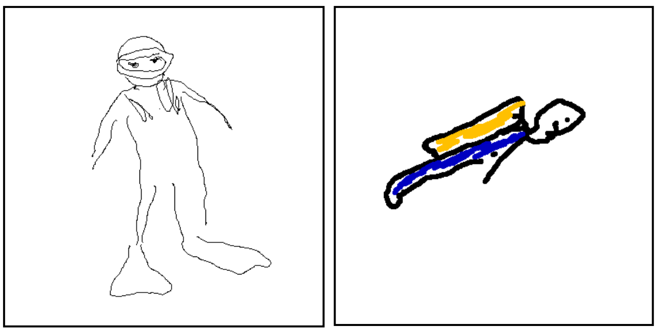


Figure 46 - Two queries for "Scuba diver" (Query 27 respondent 4 and query 73, respondent 9). Query 27 was made using the "point" pen type, while query 73 was made using the "medium circle" pen type.

Table 25 - Overview of pen split use in VISI

Pen	Pen #	Use (N=256)
Medium Circle	2	60,2 %
Point	1	50,8 %
Thick circle	8	4,7 %
Texture A	5	4,3 %
Medium square	3	2,3 %
Right split	4	1,2 %
Thick square	9	0,4 %
Thin square	7	0 %
Dotted line	6	0 %
Left split	10	0 %
Texture B	11	0 %
Thick dotted	12	0 %

There were no significant differences between the two respondent groups or the three query categories.

When using Retrievr the users do not have as large a choice of tools. They can choose from four different pen sizes: 10, 20, 30 and 50 points. The smallest pen size was used in all queries, while the remaining sizes were used in 5%, 3% and 4% of the queries. The largest size was primarily used as a tool for quickly colouring the entire canvas, while the two middle pens were used in some queries where the respondents created large objects, such as the skyscrapers in queries 247 and 411 (Figure 48).

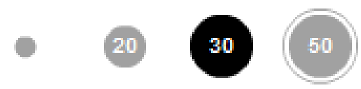


Figure 47 - The pen sizes available in Retrievr.

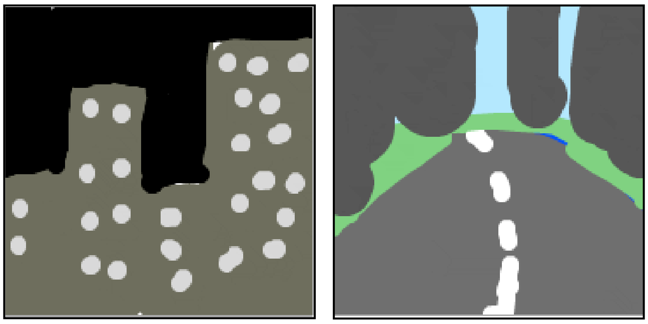


Figure 48 - Retrievr queries made using the larger pen tools. Query 247 (Respondent 24) and 411 (Respondent 31)

The descriptive analysis of tool usage can be summarized as:

- The respondents appeared more satisfied with the tools available in VISI than in Retrievr
- The respondents used *freehand drawing* as the primary method for creating visual queries. A large majority (81.6%) of the queries made in VISI were made using only freehand drawing.
- There were no significant differences in tool usage between the two respondent groups, but there were some indications that respondents who rated themselves as *skilled drawers* used drawing tools *less* than users who rated themselves as *low skilled drawers*.
- When using drawing tools in VISI, the respondents used primarily basic geometric shapes such as squares, circles and lines, or the polygon tool. Arrows were not used in any queries.
- There was an increased use of drawing tools in queries for *complete scenes*
- In VISI, the respondents used either a medium circle or a single point for drawing, and in Retrievr they used almost exclusively the smallest pen size, indicating that the respondents preferred using small pen sizes.

Based on this, it is possible to evaluate research hypothesis 1. 1:

RH1.1: The hypothesis must be *rejected*. The respondents used *freehand drawing* as the primary drawing technique. However, there was an increased use of drawing tools in queries for *complete scenes*.

The above observations raise some questions that should be evaluated further:

1. How do the respondents compare and describe the tools available in the two interfaces?
2. Why did the respondents use freehand drawing as their primary query expression technique?

3. Why did the respondents use more drawing tools (In VISI) when querying for *complete scenes*?

Overall, the respondents generally agreed that they were satisfied with the tools available in the VISI system. The choice of tools and pen sizes felt familiar to most of the respondents, the tools generally worked as they expected that they would, they were fairly easy to use and the respondents were generally satisfied with the basic shapes present. The only tool the respondents missed was an easy way to draw triangles. However, there were several tools they didn't see any use for, and several respondents reported that they would prefer that these were removed, as illustrated by respondent 19:

> *It was very nice and lucid, and in a sense very recognizable. But I think that it doesn't need… For my part, I think I would have removed most of these symbols [points at the arrows]. You don't really need them. And I never used the eraser - just drawing works better.* ***Respondent 19***

The two texture tools were almost unused. Three respondents used one of the texture tools (Pen # 5) in an attempt to add texture to the query image (Figure 49). However, they were generally not satisfied with the results, and would have preferred a proper texture tool, as illustrated by respondent 14 when asked about query 112:

> *I used that spray [texture tool a] in order to get a more... Maybe a more living image. You know, waves and so on are not generally just blue, there is foam, movement and... those things. And a shark is normally not all grey, but different shades of it. So I tried to use the texture spray thing, but the results were pretty awful, weren't they?* ***Respondent 14***

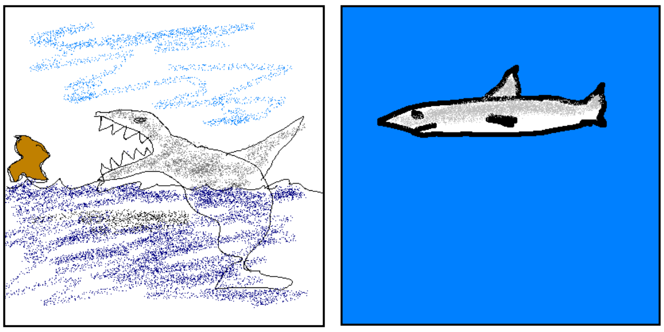


Figure 49 - Use of the "texture pen" to add textures to a query. Query 112 (Respondent14) and 116 (respondent 15)

It should be noted that the use of the *polygon* tool (17.1%) was almost exclusively by a few of the respondents, particularly respondents in the KHIB group. These respondents stated that they used it as a cross between freehand drawing and drawing tools, as it presented them with almost the same freedom as freehand drawing, but giving them more straight edges and a closed shape filled with their selected colour:

> *It's because... It is quick and easy to use. It almost gives me the same freedom as the freehand tool, but it saves me some time as I don't have to go back and fill the shape with colours when I'm done. So... Yeah, I guess it's a kind of trade-off between freedom and speed and ease-of-use.* ***Respondent 31***

An example of the use of the polygon tool is shown in Query 396, Figure 50. Respondent 31 used freehand drawing to create the dolphins, and used the polygon tool twice to create the top and bottom of the iceberg.

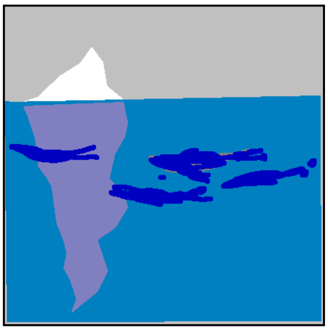


Figure 50 - Example of the polygon tool. Query 396, respondent 31.

Several respondents also stated that the different pen types were unnecessary, and that it would be better if it was replaced with different pen sizes, as illustrated by respondent 29

> *I used [the medium square] one or two times, but it wasn't really necessary. If you could just choose from the smallest to the biggest point size, that would cover most needs.* ***Respondent 29***

Concerning the actual use of the pen in the VISI tools, the two most used were the "point" and the "medium circle". The respondents were often very consistent in their use of these two tools: Some respondents clearly preferred using the *point tool*, and made almost all their queries using this tool. These respondents generally agreed that this pen size gave them the most flexibility and that the other tools were "too coarse" for their drawings style, as illustrated by respondent 4:

> *The default pen type was far too coarse. I don't... It didn't fit my drawing style very well, it felt too much like I was a child drawing. I couldn't add very much... details and so on.* ***Respondent 4***

When these respondents used any of the other pen sizes, it was primary done to emphasize some part of the query image e.g. the edge of the pool in query 137 (Figure 51, respondent 17).

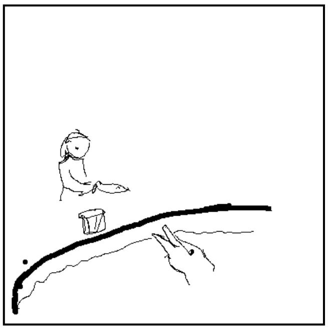


Figure 51 - Example of using the *medium circle* pen to emphasize an important part of the query

Other respondents used mainly the “medium circle” pen. It is likely that the widespread use of this pen can be attributed to the fact that it was the default selected pen size. Several respondents mention this, as illustrated by respondent 1 when asked about why he chose that pen:

> *I didn’t actually consider it very much. I just started drawing, and it worked for me. I think I used a smaller size in that one query where I was trying to find a shark. [Points at query 6] Yes, that one. I wanted to draw [gills], and I felt that I had to draw them smaller than the rest of the dolphin, so I used the tiny one.* ***Respondent 1***

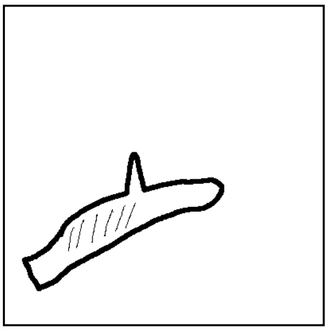


Figure 52 - Query 6, where respondent 1 used the point tool was used to draw gills on the shark.

The respondents appeared less satisfied with tools available in Retrievr than the tools in VISI, as indicated by their responses in the second questionnaire. When asked about this in the interview sessions, the respondents highlighted three aspects of this: Lack of tools, a limited choice of pen sizes, and a lack of supporting tools.

Most of the respondents who used both interfaces reported that they were less satisfied with the choice of tools in Retrievr than in VISI. While most of them did not *use* the drawing tools in VISI to any extent, they reported that they enjoyed having access to them, particularly in order to quickly define large areas of the canvas.

Some of the respondents missed the **supporting tools**, particularly the fill bucket. They had no means of filling the canvas with colour other than using one of the pen sizes and manually colouring the entire canvas. Several respondents stated that they would have enjoyed having access to a fill bucket in order to make the query process more efficient:

> *The biggest problem [with retrievr] was the lack of a colour bucket... Not being able to... I couldn't simply colour big areas in a simple manner.* ***Respondent 24***

Some of the respondents also stated that a copier tool or an eraser would have been nice. The very low use of these in VISI indicates that the need for these may not be very pressing.

All respondents reported that they were dissatisfied with the pen sizes available in Retrievr, particularly the size of the pens compared to the canvas. This made adding details to the queries difficult.

Concerning the considerable use of freehand sketching, three major causes for this were identified:

- The drawing tools available in VISI were not very useful for the domain of the retrieval tasks
- The use of freehand drawing required less effort and time than the use of tools
- Freehand drawing was more convenient, natural and fun than using tools.

First, several respondents stated that they felt that the shape of the drawing tools included in the VISI system were not well-suited for the retrieval tasks they were performing. Some respondents stated that they found the figures "too mathematical" or "better suited for architecture or engineering tasks". Generally, the respondents agreed that the available tools would not be very useful when searching for animals, people or other living things, as illustrated by respondent 12 when asked why he didn't use drawing tools very much:

> *[It was] because the shapes were square, circle, rounded square [..], and there were very few things I wanted which were square. Neither dolphins nor sharks appear very square in my head.* ***Respondent 12***

This is also reflected in the way the tools were *used* in the queries. In 43% of the queries which used tools, tools such as squares or circles were used to define larger areas of the image, particularly background, as illustrated by query 154 (Figure 53), where a blue square was used to create "the ocean".

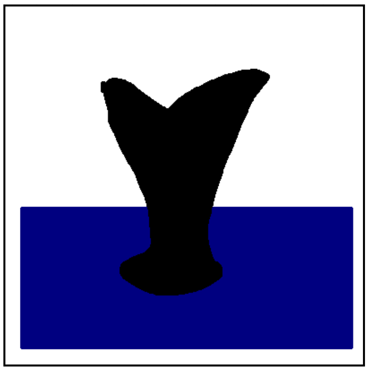


Figure 53 - The use of a rectangle to create an area representing the ocean (Query #154).

In 28.6% of the queries, tools were used to draw objects which have a "real-world" shape similar to the shapes in the interface, e.g. the use of a line and a circle to create the "wheel of fire" in query 200 (Figure 54).

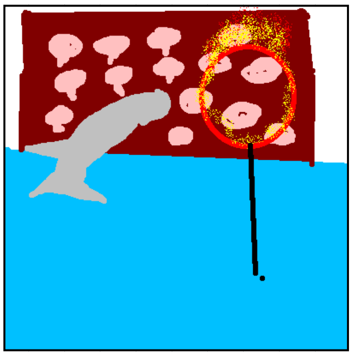


Figure 54 - The use of the *line tool* and the *circle tool* to draw a "wheel of fire". Query 200, respondent 21

Several respondents commented that they found the tools useful when drawing such objects, as illustrated by these statements from respondents 12 (referring to query 94, Figure 55a) and 15 (Referring to queries 115 and 121, Figure 55b and c):

> *I tried to draw the oxygen tank of the scuba diver with this tool in order to compose a circle and a square in a kind of rounded unity, and put it on the back of the diver. This seemed like the best way to use these squares in a simple way.* ***Respondent12***

> *In the drawings I made, using predefined shapes wasn't something I felt a need for, except for a round ball and the one with the diver's mask, for simplicity's sake.* ***Respondent 15***

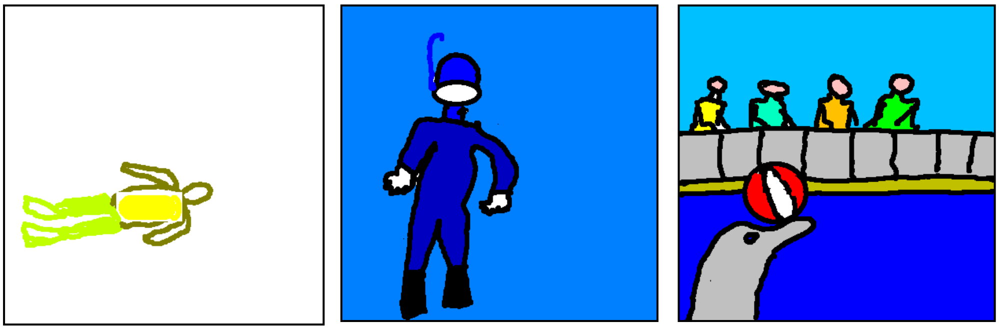


Figure 55 - Queries where shape tools were used to add details (Query 94, respondent 12 and queries 115 and 121, respondent 15).

In 51% of the queries made using drawing tools, tools were used to represent real-world objects that *do not* share a "real-world" similarity to the interface tool, e.g. the use of a circle to represent a "fish" in query 12 (Figure 56). For most of the queries where shapes were used in this manner, they were used to create minor elements in the image such as the fish in Figure 56. While there were some examples of using the drawing tools to create the major participants in the query, e.g. the boat in query 106 (Figure 57a) or the boat and shark in query 16 (Figure 57b), this was not a very common use of the drawing tools.

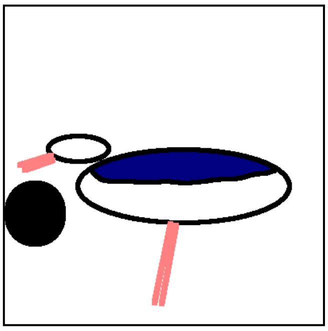


Figure 56 - Use of a circle to represent a fish (Query 12, respondent 2).

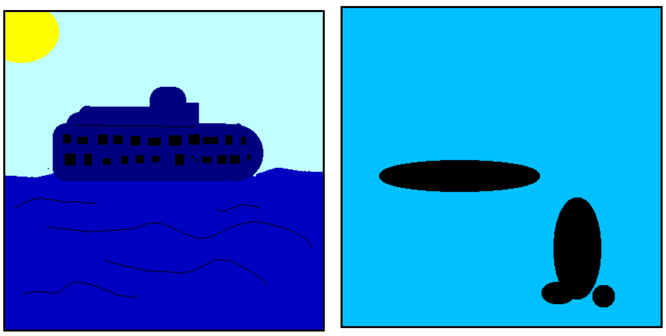


Figure 57 - The use of circles and squares to create a ship (Query 106, respondent 12) or ovals to represent a boat and a shark (Query 16, respondent 2)

Next, several respondents stated that they preferred to use freehand drawing as they felt it was much *faster and efficient* than using the drawing tools, as illustrated by respondent 29:

> *When using only the [freehand] tool, it's much faster to draw. You don't have to swap between the different tools, and if I need a circle, it's easier to just draw it than to click on the circle box, draw the circle, and click back on the circle tool. It's faster, and it keeps focus on the drawing process.* ***Respondent 29***

Similarly, several respondents stated that they *knew what they wanted to draw*, and that they *didn't want to* mess-around *with the drawing tools*, as illustrated by respondent 3:

> *I knew what I wanted to draw. I guess I could have used some of the tools, but... I was ready to start drawing, and I didn't want to mess about with the tools. It would have taken me more time, and I wanted to keep it as simple as possible while still getting [the results] I wanted. I just wanted to try to draw.* ***Respondent 3***

This observation is in part supported by the time spent by the respondents when drawing the query images. Table 26 shows the mean time (in seconds) used to create the queries in VISI using freehand, tools and the combination of these.

Table 26 - Query time (in seconds) for the different tool combinations.

Drawing tool	Minimum	Maximum	Mean
Only freehand (156)	12	416	127,7
Only tools (10)	44	207	86,4
Freehand and tools (55)	29	431	164,87

The least amount of time was spent on the queries created using only drawing tools. These queries were generally very simple, as shown in Figure 58. While the low number of queries created in this way (10) makes any generalizations difficult, it indicates that using only tools may allow users to express simple queries in a very fast manner.

Figure 58 - Queries made using only tools (Queries 11, 15, 166 and 205). The images have been cropped.

However, comparing the queries created using *only* freehand with the queries combining freehand and drawing tools reveal a significant difference of 37 seconds[45]. While the difference is not large, it indicates that using freehand drawing alone is slightly faster than combining it with drawing tools, and may be important for users who find the time required to create the queries to be an issue.

Finally, several of the respondents stated that they found freehand drawing a much more natural way of expressing the queries, using the tools would “remove some of the fun” of the query specification process, as illustrated by respondent 27:

> *I liked the way I could sketch [the queries]. It was really nice, and I didn’t feel like using those squares [drawing tools]. Sketching was simply more fun. Using [the tools] would take away that. Take away the fun.* ***Respondent 27***

Several respondents also stated that choosing the freehand drawing was not a conscious choice, but an act of impulse, as illustrated by respondent 16 when asked why he used freehand:

> *It was pure impulse, actually. An old habit, you might say. Yes... It was really a pure impulse, and not a conscious choice.* ***Respondent 16***

It appeared as if the respondents felt that freehand sketching represented a more direct, or hands-on, approach than using the drawing tools: Freehand drawing appears to be more *convenient* than the use of drawing tools, as illustrated by respondent 19:

> *Freehand sketching was more convenient, in most cases. I think... Actually, the paint bucket [Fill tool] was the only tool I actually wanted to use much. I don’t really need help drawing a square, and I don’t think the computer notices any difference between my squares and the squares made by the drawing box. But filling the boat with colour using the bucket was nice. But yes, freehand sketching was definitely most convenient for me. And way faster!* ***Respondent 19***

This observation seems to be particularly true for the KHIB group and the respondents who rated themselves as good drawers, although some of the IFIM respondents also mentioned this.

It should also be noted that the two respondents who stated that they had slight handicaps causing problems with tactile activities stated that they initially felt that the drawing tools would help them, but after using them for a time found them *even more* difficult to use, as they still had to use the pen to carefully define and place the shapes. While they had some problems defining things such as

[45] $t[27.323] = -4.032$, $p < 0.01$

straight lines using the freehand tool, they stated that they had very little benefit from the drawing tools.

Concerning the increased use of drawing tools in requests for complete scenes, two contributing factors were identified: Using drawing tools to define background, and using shape tools to quickly add several details to the query. Note that the number of queries in this category in VISI was quite low: 61 queries in total, 28 of these were created using drawing tools. While it is not possible to draw any general conclusions based on this material, some observations can be made.

Queries for complete scenes were generally more *complete* than queries for generic objects and narrative content: More background and contextual elements were used in these queries (Described in detail in chapter 6.1). And, as noted above, the shapes were often used to define major areas of the canvas. Consequently, a correlation between the use of background and the use of shapes would be expected. This was supported by an analysis using Kendall's tau-b correlation coefficient[46].

Some respondents also stated that the tools were helpful when adding several similar details to an image, e.g. such as people in a crowd (Figure 59):

> *The circle tool was useful when I was trying to find those jumping dolphins... It was easy to draw all the people watching [using the circle tool]. That would have taken more time if I had to sketch it.*
> ***Respondent 26***

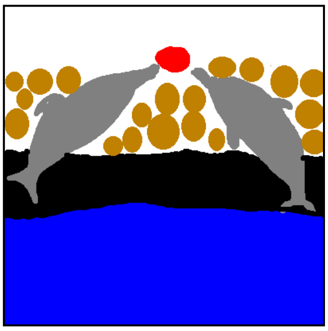


Figure 59 - Use of circles to add query details (Query 287, respondent 26 and query 395, respondent 31).

Another example of this is the two queries represented in Figure 60 (Queries 394 and 395). Both these queries were made by respondent 31 for query task 15 ("Find images of two dolphins

[46] $r(256) = 0.213$, $p < 0.01$

entertaining humans in an aqua park”). The respondent was not satisfied with the results from the first query, and decided to make another attempt, but with more colour in the background.

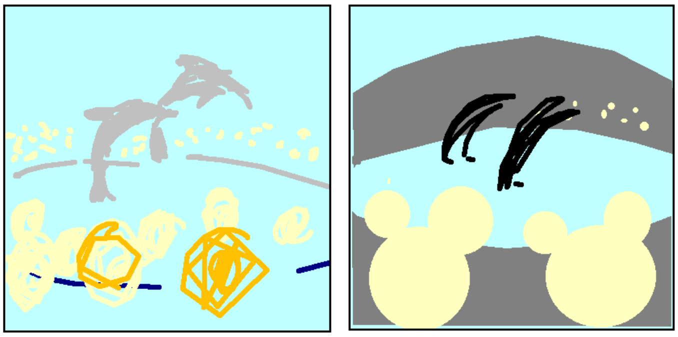


Figure 60 - Two queries for "Dolphins entertaining humans" (Queries 394 and 395, respondent 31)

The following is a transcript from the query session after the second query was completed. “R” represents questions and statements made by the researcher, while “31” represents statements and answers from respondent 31:

> *R: In the previous query, you used only freehand, but here you used the polygon tool and the circles. Could you say something about that?*
>
> *31: It’s... I spent a lot of time on that previous query adding all those details. This time, I wanted to do it faster. Or maybe not faster, but faster to... [Pauses]*
>
> *R: So you used the circle tool?*
>
> *31: Yes. Actually, it wasn’t much faster. It was faster to use the freehand, but I felt I should close the heads. That is, the heads should be filled. That would take too much time and effort with the sketching pen. So I used those filled circles. But.. [laughs] the results were just as good. Or just as bad.*
>
> *R: So, which approach did you prefer? The freehand or the circle tools?*
>
> *31: Depends. It was faster to just sketch the people and add dots with the freehand tool, but... If the VISI computer actually needs the circles to be filled, I think it’s easier to use the circles. But... [laughs]. I guess I was too lazy. I didn’t add as many people in the second query, even though it was easier.*

Some other respondents made similar statements, but there were few examples of query images with this level of detail. Nevertheless, it is possible that an increased level of details in the queries may lead to an increased use of shape tools to lighten the job of adding several identical objects to a query.

Summarized, the respondents were generally more satisfied with the tools available in VISI than in Retrievr. While some of the drawing tools available in VISI were not used at all, the respondents enjoyed having access to the geometric shapes, and found these particularly useful for adding colour to large areas of the canvas. The respondents also reported that they were dissatisfied with the pen sizes available in Retrievr, particularly size of these pens compared to the size of the canvas. Freehand drawing was clearly the most used drawing technique. The respondents generally felt that the tools provided in VISI were not well suited to the retrieval tasks and that freehand drawing was faster, more efficient and more natural. However, the drawing tools were useful when defining large areas of the canvas, particularly when querying for complete scenes. Furthermore, some respondents found the tools useful when adding a lot of detail to the queries. Finally, there were some indications in the interview data that respondents in the IFIM group enjoyed using the drawing tools more than the respondents in the KHIB group, though this was not reflected in the use of tools.

5.1.2 Colours

The questionnaire presented the respondents with two questions regarding colour and the way colours were presented to the user:

- Question 17: How satisfied were you with the choice of colours [In VISI]?
- Question 21: How satisfied were you with the choice of colours [Retrievr]?

Table 27 presents descriptive data of these two questions. The questions were only asked in the final experiment (N=12). A score of 1 indicated “very dissatisfied” 5 indicated “very satisfied”.

Table 27 - Respondent satisfaction with the colours in the two interfaces

Question	Min	Max	Mean	Std. deviation
17: Colours in VISI	1	5	3,08	1,379
21: Colours in Retrievr	2	5	4,5	1,000

The 12 respondents appeared more satisfied with the colours available in Retrievr. The difference was not significant at the 0.01 level[47], and given the low number of respondents little weight should be given to this.

VISI offered the respondents a choice of 40 colours (including black and white), as illustrated in Figure 62 (Rotated):

[47] Wilcoxon signed ranks test: $Z = -1.919$, $p > 0.01$

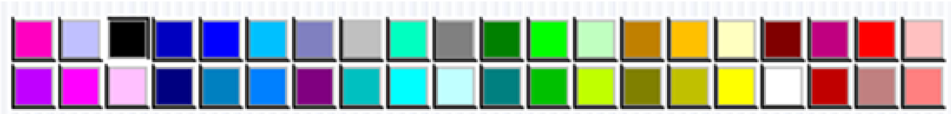


Figure 61 - The colours in VISI. (Rotated to the right)

The respondents generally agreed that they would have preferred more freedom when selecting the colours. Three issues with the choice of colours were identified: The grouping of the colours appeared random, the choice of colour shades was too limited, and they were disappointed by not being able to select additional colours.

First, several respondents mentioned that they were confused by the grouping of the colours, particularly respondents who were used to working with applications such as Photoshop.

> *I think there were enough colours, but the grouping was way off! I couldn't find the colours I wanted when I needed them, and that caused a breakdown in my drawing process. Why were not all the reddish colours grouped together?* ***Respondent 1***

The interface videos also indicated that this might be a problem. In some cases, the respondents spent quite a lot of time looking for the "correct" colour. In some cases, they failed to identify and use a particular colour, even if the colour they were looking for was available in the interface.

Next, the respondents were disappointed by the lack of colour shades. For generic tasks, such as defining "the ocean" or "the sky", they were satisfied with the choice of basic colours. But in some cases these were limited, as illustrated by these statements:

> *Are there more shades of blue than these? [..] No? I'm not sure how I'm going to describe that this [points at the bottom of the screen] is the ocean and this [points at the top of the screen] is sky. I don't think these colours are good representations of the sky. [..] Here, I think I need more colours... More shades of these colours.* ***Respondent 12***

> *I thought [the lack of colours in VISI] was problematic, for example when I wanted to draw skin colour on the humans. She looked like she had shoe polish in her face.* ***Respondent 22***

One issue here is that is seems as if the respondents were more aware of the colour shades when drawing objects they were familiar with, such as "humans" or "the sky". When drawing things such as dolphins and fish, they appeared more than satisfied with using the basic colours:

> *It's OK for the dolphins and sharks and... They're just greyish anyway. But again, the humans got really strange.* ***Respondent 20***

It also should be noted that a few of the respondents were quite happy with the low number of colour shades available, as this meant they didn't have to consider the real world shade of the objects, as illustrated by respondent 7:

> *There were times I thought "No, I should have a slightly lighter shade of blue here" or that I should have used a more "grey" grey. On the other hand, I think that [using more colours] might have influenced the precision of the results. You shouldn't have to hit the exact colour from the millions of possible colours, so...* ***Respondent 7***

Finally, the respondents who used professional photo software (e.g. Photoshop) on a regular basis stated that while they were generally satisfied with the basic colours available, they would prefer to have the ability to select custom colours for particular tasks

> *I think I would have reduced the number of colours. Just keep the basic colours, and provide the possibility of selecting custom colours on demand. Now, it's like... It's like there are too many colours, but most of the colours are not really needed and... They're really just clutter.* ***Respondent 19***

In Retrievr, colour selection is a two-step process. The user first has to select one of 12 basic colours, and then select one of 72 different shades of that colour, as illustrated in Figure 62. This theoretically presents the user with 864 unique colours. However, in practice, for most of the darkest shades and some of the lightest shades appear very similar for the user. *Black and white* are not available as unique colours in Retrievr; the user has to select one of the basic colours and then select the lightest or darkest shade in order to get "white" or "black".

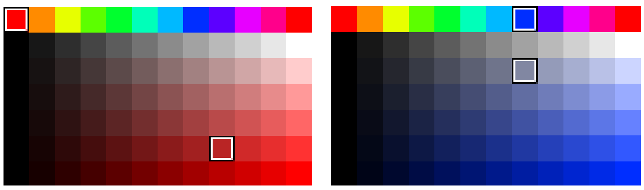


Figure 62 - The colours in Retrievr. Leftmost with "Red" selected, rightmost with "blue" selected.

The respondents were very satisfied with the choice of colours in Retrievr, and had only a few comments regarding colours. Two issues were brought up by the respondents: A lack of a "neutral" colour and that the selection of colours may have been too complicated.

Several of the respondents mentioned that the lack of "black" and "white" forced them to consider the colour of the objects in the query. In some cases, they did not wish to indicate a colour, but rather simply draw the shape of the object. In VISI, they felt that the use of the colour "Black" would allow them to do this; they considered it a "neutral" colour. However, in retrievr they had to select one of the basic colours first. Similarly, the lack of a unique "white" colour forced them to use one of the other colours in order to "erase" drawings using a shade of the basic colours, making them unsure if the system understood this as "erasing" or "adding a colour".

The respondents also felt that the two-step process required too many interface actions. In many cases, they clicked on the basic colour at the top of the colour palette, and expected that the selected colour would be used. If they had selected a very dark shade there would be no difference between the prior colour and the new colour.

Summarized, the respondents were satisfied with the colours available in the interfaces. They were more satisfied with the colours available in Retrievr than in VISI, particularly with regards to the amount of shades available in Retrievr. Some respondents reported that the colours in VISI should be grouped betted, and that selecting colours in Retrievr required too many actions. Several respondents also stated that they would like to have access to "neutral" colours in Retrievr.

5.1.3 Drawing Canvas

The *drawing canvas* is the area in the interface where the users draw the query image. VISI and Retrievr both offered the respondents a square drawing canvas, the only difference between them being size. This was not a topic in the questionnaire, but was brought up by most of the respondents during the query process and in the interview sessions.

The respondents were generally very satisfied with the size of the canvas in VISI. The combination of the large sized canvas and the different pen sizes gave the respondents a high degree of freedom when creating the queries. Only one reservation was made by several respondents in the KHIB group: They would like the ability to resize and re-format the canvas, as illustrated by respondent 16:

> *The canvas [in VISI] was great, there was ample room to draw what I liked. But... It would have been great if I could change the layout of the canvas, particularly for the [tasks] where I was trying to find images of [Skippy] in the fjords. Most images of this type... These kinds of images are normally in a landscape format, not portrait. It would be easier for me to create... I'd be better able to create good landscapes that way.* ***Respondent 16***

Several other respondents in the KHIB group made similar statements, primarily related to the types of images they were attempting to retrieve.

The size of the canvas represented the largest reservation the respondents held towards Retrievr. Most of the respondents stated that the size of the canvas was far too small, particularly when combined with the size of the smallest pen size: It was difficult to compose queries with more than a few objects, and it was difficult to add details to the objects in the queries.

Most of the respondents stated that they found it difficult to add more than a few number of elements or details to the query image. In most cases, they felt that they had to pick *one* object representing the subject matter of the query, even if they ideally wanted to add more elements. One example of this is query 243 by respondent 24 (Figure 63). In this query he was attempting to retrieve images of people practicing sports:

> *I wanted to try an approach with [..] a group of people playing football. But there was simply not enough space in Retrievr. I had to just draw the football on the playing field. There was simply not enough room for the people.* ***Respondent 24***

Figure 63 - Query 243 (Respondent 24): request for images depicting people practicing sports.

When the respondents did create queries with more than a few objects, they were not happy with the level of detail they were able to add to the individual elements in the query. Query 305 (Figure 64) is another example of a request for images depicting people practicing sports. Respondent 27 also tried the approach with a group of people playing football. In this case, he actually drew three persons and a football, but he was very dissatisfied with the results:

> *Well, that looks more like a bunch of mushrooms, or even a group of small elves rocking in the forest. It certainly doesn't look a lot like a football match.* ***Respondent 27***

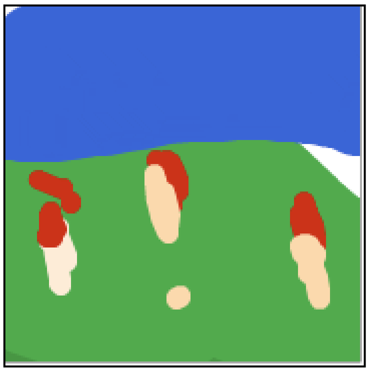


Figure 64 - Query 305 (Respondent 27): Request for images depicting people practicing sports.

Several other respondents also mention problems related to this:

> *That other system [Retrievr]... (sighs) that was too inaccurate, for example for finding a flower. Or rather, finding a particular flower. [..] It's easy to draw a flower, but how can I describe that it's supposed to be a dandelion? It's too coarse. It would have been easier to just type 'dandelion'.* ***Respondent 26***

> *Well, I had a mental image of the Manhattan skyline, and tried to draw large squares. But... It didn't quite work out... It's too dependent on the coloured areas. That is, Manhattan Skyline is... The eye can discern all the details, the houses, buildings... But with the [small canvas] [retrievr] worked with, I couldn't manage to separate the small skyscraper from the big one. They all blended together.* ***Respondent 23***

Most of the respondents agreed that they felt Retrievr should have a larger canvas and smaller pen sizes. However, two of the respondents remarked that, after giving it some consideration, they actually preferred the small canvas, as it forced them to think in *composition* rather than *details*:

> *The small canvas.. It forced you to think in terms of composition. That was probably more effective, as... I don't think the system actually recognizes all those details, and by focusing on the major compositional areas, the computer might be better able to find similar images. So... I guess it was rather nice, in a way.* ***Respondent 30***

Summarized, the respondents were generally satisfied with the canvas in VISI, though some respondents would like to be able to resize and reformat the canvas. All the respondents experienced challenges related to the relationship between the small canvas and the large size of the pen tools available in Retrievr. Most respondents would have preferred a better size-ratio between the two. However, some respondents remarked that this actually forced them to think in composition rather than detailed elements, which they felt might be better for the retrieval systems.

5.1.4 Query Dynamic

Query dynamic relates to the overall query process provided by the two interfaces: Query specification, definition of query parameters and result browsing. This was not evaluated in the questionnaires, but was discussed in detail during the interview sessions.

The two systems represented two very different approaches to the query process. In VISI, the respondents had to first draw the query, then specify the query parameters, and then finally browse the query results. It was not possible to go back and forth between the different stages and modify the drawing or the query parameters. In Retrievr, the respondents did not have the option to specify query parameters; they were bound to the parameters defined by the retrieval system. On the other hand, they could modify the drawing as often as they liked, and the result browsing was integrated into the drawing process, providing a very dynamic query process.

The respondents were not satisfied with the query process in VISI. Several respondents stated that they felt that the different stages in the query process were disconnected and isolated. They often wanted to go back and modify the drawing, something that could not be done:

> *One of the biggest problems was that I couldn't go back and change the drawing. In that one query, I wanted to quickly change the background colour to something else because I saw that the colour I used returned only the wrong results. But this wasn't possible.* ***Respondent 12***

> *Now I've drawn this lousy dolphin, but I didn't get the results I was hoping for. Rather than pressing "new search", there should be an option like "modify". That would have helped me, then I could go like "OK, I've tried this and got that, now how can I manipulate this drawing in order to get something else". That's missing right now.* ***Respondent 4***

Similarly, they wanted to be able to modify the query parameters and see how this affected the results:

> *It would have been easier if the results were updated in real-time when you update these sliders [the query parameters]. It's a bit artificial that you first have to search, then specify the sliders, and then view the results. You don't get the whole picture. The relationship between your input and the results would have been clearer if there was a connection between the sliders and the results. The way it was now, I didn't get a feeling for the differences between setting the slider to 30, 40 or 50, or if it even had any effects at all. So it just became random in the end.* ***Respondent 7***

Several other respondents reported similar issues. They felt that while they enjoyed being able to specify which elements of the query image the system should focus on (Colour, shape, texture or spatial distribution), they were not able to see or understand the effects the sliders had on the results. Consequently, they often decided to just set the weights even and stop using them completely.

Furthermore, none of the respondents who used the query parameter sliders found any use for the “texture”. In most cases, they either ignored the slider or turned it all the way down. Other respondents stated that they would probably use it more if they were using real images as query input:

> *I have some difficulties understanding how the texture slider would be useful. I guess if I uploaded a real image, then the level of detail would be large enough for the texture to matter. But when I was drawing, I primarily made large, flat areas, and I probably shouldn’t have used texture at all. Because there simply isn’t any texture in the simple drawings I’ve been attempting to create.* ***Respondent 10***

All of the respondents enjoyed the real-time presentation of the query results in Retrievr. Three main benefits of this approach were identified: The feedback made the query process more effective, it made it possible to modify the query image based on the results, and made it easier to understand how the system processed the queries.

First, several respondents stated that they felt the dynamic feedback allowed them to draw more efficiently. In VISI, they often did not know when the query would be “good enough”. With Retrievr, they were able to determine this based on the results they obtained. They stopped drawing when they were satisfied with the results, or when they decided they were unable to obtain good results. Additionally, the feedback provided the respondents with assistance:

> *If you’re drawing rough sketches, then many things may appear very similar. The fins [of a dolphin] and the wings [of a seagull], and the beaks and... If I’m trying to find a dolphin, and the first five images shown contain seagulls, I would have thought: “OK, there’s probably something I should do in order to more clearly state that this is a dolphin”.* ***Respondent 6***

> *The nice thing about Retriever was that it updates your search “on the fly”. For example, if you stop drawing in order to think “Hmm, how am I supposed to draw the next detail”, and then you get a feedback, or a preliminary result, which was nice. Then you can suddenly decide that “No, I don’t need to continue drawing”. Even if I’ve only drawn a single tree, I got [an image of an entire forest].*
> ***Respondent 22***

Next, the dynamic approach also reduced the biggest problem with the VISI system: Not being able to change the drawing once it was submitted. Observing the interface videos reveal that the respondents often changed their queries during the process, erasing some elements and added other elements to the query. While this naturally increased the time of the query process, it was nevertheless much faster than creating a new query from scratch. Several respondents also remarked on this during the query process and in the interview sessions.

Finally, several respondents stated that they gained a better understanding of how the retrieval system worked when they were able to see the results of their actions immediately:

> *In [VISI] it was difficult to understand what happened. You drew something, perhaps with details, sometimes without details, and then you specified the weights, but you didn't really see if it had any real impact. But in [Retrievr], with the immediate feedback, in a way you knew and understood what it was after, and you could see if any of the images you created had similarities to the images the engine had stored. You understood the importance of colours and composition, which made if more easy to use. And more fun.* ***Respondent 23***

> *I felt it was easier for me to… in a way to specify what… How did it think? I understood more how [Retrievr] worked with regards to colours and all that. In [VISI] I didn't have a real understanding of what it did with the colours I choose, but here I immediately saw that when I were using only grey, only grey images were returned. Or black.* ***Respondent 24***

To summarize, all the respondents stated that they preferred the dynamic query process provided by Retrievr. While they enjoyed having the ability to define the query parameters, the fire-and-forget mechanism in most cases resulted in the respondents failing to understand the mechanisms and stopped using these. Furthermore, they disliked not being able to modify the drawings once they were defined. The dynamic result presentation provided assistance to some of the respondents by providing images based on a half-completed query. Finally, several respondents reported that they were better able to understand the underlying mechanisms when they immediately saw how their actions and choices were reflected in the query results.

5.1.5 Respondent Query Interface Preference

Based on the above discussion, research hypothesis 1.2 may be evaluated:

> **RH1.2:** The hypothesis must be *rejected*. The respondents reported that there were elements they enjoyed in both interfaces, and would have preferred a combination of the two interfaces.

Summarized, the respondents preferred the tools and the canvas size of VISI, while they preferred the increased number of colours available in Retrievr. Furthermore, they clearly preferred the dynamic presentation of query results available in Retrievr. Finally, they reported that while they preferred being able to specify the query parameters, they found the concepts difficult to understand and use and would have preferred to have the results dynamically update based on the query parameters, rather than having to decide these prior to executing the query.

5.2 Time Spent Creating the Query Images

Time spent creating the query images represent the time the respondents used in the actual drawing process, i.e. excluding all other elements of the query process. The timed measurements were used to as an aid when evaluating the other research hypotheses, and have been used throughout the remaining analysis chapters. The two primary research hypotheses evaluated here are RH1.3 and RH 1.4:

- **RH1.3:** Respondents draw query images more quickly in the interface provided by VISI than in the query interface provided by Retrievr.
- **RH1.4**: Respondents with a visual background will express queries faster than respondents without this background

Time was measured in the number of seconds spent expressing the query. The mean time spent was 99.17 seconds, with a minimum of 4, a maximum of 431 and a standard deviation of 83.36. Note that these times may have been influenced by the think aloud protocol used in the experiment sessions. Accordingly, these measurements might not be representative of *realistic* query times. However, all query sessions were exposed to a similar treatment, giving a sound *internal validity* for the results. Table 28 shows an overview of the time spent by the two respondent groups, the time spent in the two interfaces and the time spent on the three different query categories[48].

Table 28 - Time spent, broken down by group, interface and category.

		Respondent group		Interface		Query category		
	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
Mean	99.18	138,12	72,21	135,05	48,98	81,13	98,68	126,06
Min	4	8	4	12	4	4	4	19
Max	431	431	288	431	241	411	416	431
St. Dev	83.36	96,4	59,8	86,89	42,49	71,74	80,89	95,25

Three main issues were identified in the numbers shown in Table 28: There was a large difference between the two groups, a large difference between the two interfaces and some differences between the three query categories.

[48] Recall from chapter 3.1.5 that the three categories were ”Generic content”, ”Narrative content” and “Scenes”

The difference between the two respondent groups[49] and the two interfaces were significant[50]. The differences between query category 1 and 2[51] and category 2 and 3[52] were not significant, but the difference between category 1 and 3 was significant[53].

It should be noted that the difference between the two respondent groups may have been influenced by the fact that 82% of the queries made by the IFIM group were made in VISI, while only 46.4% of the queries made by the KHIB group were made in VISI. As the queries made in VISI took significantly more time than the queries made in Retrievr, it is likely that the overall mean time of the IFIM group increased. In order to control this effect, the queries were broken down by both interface type and respondent group. The mean time spent in VISI by the IFIM group was 160.97 seconds, while the mean time spent by the KHIB group was 103.1 seconds. This difference was also significant[54], indicating that the KHIB group did use a shorter time on the queries than the IFIM group in the VISI interface. There was also a small difference in the time spent by the two groups in Retrievr (53.6 seconds for IFIM vs. 47.8 seconds for KHIB), but this difference was not significant. Given that only 3 respondents in the IFIM group used Retrievr this has not been given much weight. This indicates that the KHIB group spent significantly less time creating the queries than the IFIM group. A similar result is seen when comparing the respondents who rate themselves as good drawers[55] with the other respondents.

It should also be noted that these measurements only consider the time spent on the actual query expression process. The total time spent for each query was in some cases much longer. The time spent defining the query parameters (VISI) or browsing the results (both systems) was not included in this measurement. However, the differences in the two prototypes with regards to result presentation and the lack of query parameter specification in the Retrievr prototype resulted in a shorter total query time for the queries made in Retrievr. In the first experiment, the time spent on query specification and result browsing in the VISI system was measured in addition to the query specification time. The mean time spent on determining the query parameters was 35.5 seconds, while the mean time spent on result browsing was 59.3 seconds. The total time spent on the queries (excluding loading times between the different pages in the interface) had a mean of 273 seconds.

---

[49] $t(377) = 8.204$, $p < 0.01$
[50] $t(377) = 11.503$, $p < 0.01$
[51] $t(278) = -1.923$, $p > 0.01$
[52] $t(233) = -2.376$, $p > 0.01$
[53] $t(241) = -4.191$, $p < 0.01$
[54] $t(213) = 5{,}406$, $p < 0.01$
[55] These were the respondents who rated themselves as “Highly skilled” or “Very highly skilled” at question 7, questionnaire 2.

Because the query specification process and the result browsing phases were more influenced by the talk-aloud protocol than the query specification process, this was not measured in experiments 2 and 3. However, the total time spent in VISI in the first experiment suggests that the overall time spent on the query process in VISI was considerably higher than the overall time spent on Retrievr. First of all, as there is no specification of query parameters in Retrievr, the total query time is reduced considerably. Additionally, the continuous feedback provided by the result presentation in Retrievr reduces the time required to browse the results. Finally, several respondents commented that the continuous feedback in Retrievr made them use less time than they did in VISI as they could stop drawing if and when they were satisfied:

> *Actually, I think this feedback [in retrievr] made me more efficient. In VISI, I had no idea when to stop adding details and colours and so on... Here, I just stop drawing when... Yes, I stop drawing when I'm satisfied. Here [Referring to query #222, Figure 65], I found the forest at once without adding all that other stuff. In VISI, you first have to draw, import the seed image, and* then *you can search, or see the search results.* ***Respondent 22***

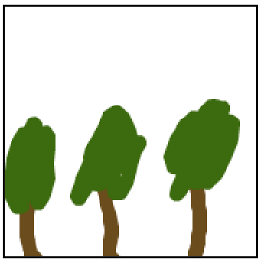


Figure 65 - Query #222, respondent 22.

Based on the above, it is possible to evaluate research hypotheses 1.3 and 1.4:

> **RH1.3:** The hypothesis must be *rejected*. The respondents spent significantly shorter time creating queries in Retrievr than in VISI. This was valid for both respondent groups and all query categories.
>
> **RH1.4:** The hypothesis must be *accepted*. The respondents in the KHIB group spent significantly less time creating the queries than the respondents in the IFIM group.

In order to understand the results from the time measurements in more detail, more information about the actual queries and the query formulation is required. Consequently, the time measurements are not discussed further here, but used in combination with other results in the following sections and chapters.

Summarized, there were some major differences in the query time:

- The KHIB group created queries significantly faster than the IFIM group
- The respondents created queries significantly faster in Retrievr than in VISI
- There were significant differences in the time spent on the different query categories. Primarily, queries for *complete scenes* took significantly more time to create than queries for *generic content*.

5.3 Summary: The Query Formulation Process

Based on the previous sections, some main observations can be made regarding the query formulation process and the respondents' use of the QBD interfaces:

1. Expressing visual queries through drawing appears to be a time consuming process. Skilled drawers are able to express queries significantly faster than people with lower drawing skills. The respondents were able to draw queries significantly faster in Retrievr than VISI, primarily due to a combination of a simpler interface and the real-time result presentation.

2. The respondents preferred to use freehand drawing as their main tool for drawing query images. Freehand drawing represented the most natural way of drawing, and the tools available in the VISI interface were generally not well-suited for the retrieval tasks the respondents were performing. However, having access to a set of basic shapes combined with some supporting tools provided the respondents with an efficient way of adding background and generic details to a query image.

3. The respondents enjoyed having access to a large number of colours, particularly having access to different shades of a given colour. However, the respondents did not appreciate being "forced" to describe an object using colour. In some cases, they would like to be able to create objects and queries without colours.

4. The respondents reported that they enjoyed having a large canvas when drawing the queries. A small canvas combined with coarse pens limited their ability to draw complete and complex queries, and made it difficult to add details to individual query elements.

5. The respondents clearly preferred the real-time nature of Retrievr. Having the result set immediately presented in the same interface as they created the queries made it easier to express the queries efficiently and they felt that they achieved higher understanding of the query process. They also enjoyed being able to specify query parameters, but felt that it was difficult to

understand the *effects* of these parameters when they could not see how these parameters influenced the query results.

The respondents also made several suggestions for improving the interfaces and the query process. These are discussed in detail in chapter 9.

6 Query Image Modality Classification

The second study objective in this work was the *query image modality*, e.g. the *degree of realism* in the query images created by the respondents, as expressed in **research question 2:**

How realistic are the query images drawn by QBD CBIR users?

Four main research hypotheses were suggested to address to this question:

- **RH2.1:** Respondents will create query images with a low degree of visual modality
- **RH2.2**: Respondents with a visual background will create query images queries with a higher degree of visual modality than respondents without this background
- **RH2.3:** Query images created in the VISI interface will have a higher degree of visual modality than queries created in the Retrievr interface
- **RH2.4**: The visual modality of the query images increases with the complexity of the image requests

The framework for visual query image classification described in chapter 4 was used to operationalize these hypotheses. The hypotheses were broken down and analyzed according to the four modality markers presented in the framework:

1. The *completeness* of the query images, e.g. the use of *contextualization*
2. The degree of which *colours* have been used to express the queries
3. The degree of abstraction used to represent the objects in the query images
4. The degree to which *compositional structures* have been used to create realistic images

Sections 6.1 through 6.4 each present a detailed analysis of the query images based on the framework. Each of these sections analysing framework measurements follows the same structure:

i. Breakdown of the research hypotheses into sub-hypotheses
ii. Evaluation of the sub-hypotheses according to the framework described in chapter 4
iii. A discussion of the findings based on material from the interview sessions, questionnaires and the interface videos
iv. A summary of the discussion

Section 6.5 presents a summarized overview of the discussions and a summary of the main research hypotheses described above.

6.1 Query Completeness

Query completeness describes how *complete* the query images are, e.g. the use of *background* and *contextual elements* to create a *complete* query image. This can be represented by the *contextualization* of the query images.

6.1.1 Research Hypotheses

A new set of sub-hypothesis were defined for each of the four major research hypotheses:

- **RH2.1.1:** Respondents will create query images that are *simple*
- **RH2.2.1:** Respondents with a visual background will create more complete query images than user without this background
- **RH2.3.1:** Respondents will create more complete query images in the VISI interface than in the Retrievr interface
- **RH2.4.1:** The completeness of the query images increases with the complexity of the image requests

These hypotheses have primarily been evaluated according to the framework described in chapter 4 using the following measurements:

1. Counting the number of individual objects present in each query image
2. A classification of colour use according to the *contextualization modality marker*
3. A subjective evaluation of the completeness of each query image

According to the modality criteria, a complete query image i an image containing participants, fully detailed backgrounds and contextual elements.

6.1.2 Evaluation of Hypotheses

The first indicator of query completeness was the number of **individual objects** present in the query image. This was a simple indicator of the completeness of a query image: a *complete* image is likely to have more objects than an image that is *simple*.

The mean number of objects created in the queries is presented in Table 29. The respondents generally created few unique objects in the query images, particularly for type 1 queries and type 2 queries. More than half of the query images consisted of a single object and 90.3% consisted of 4 or less unique objects.

Table 29 - Mean number of objects in the query images

Overall	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3

2.10	1.86	2.28	2.14	2.02	1.23	1.87	3.79

While there was a slight difference between the two retrieval systems, this was not significant[56]. The KHIB group had a slightly higher mean number of objects than the IFIM group, but the difference was small and not found to be significant[57]. The number of objects increased in type 2 and type 3 queries (Queries based on requests for complete scenes). The differences between the query types were all significant at a 0.01 level of significance.

The second indicator of query completeness was the **contextualization modality marker** and the corresponding modality criteria. Table 30 presents an overview of the overall classification of these criteria[58].

Table 30 - Contextualization modality criterion for all query images

		Group		System		Query type		
	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
Contextual category	(414)	(181)	(233)	(256)	(158)	(166)	(144)	(104)
Participants (A)	98.3%	98.3%	98.3%	99.6%	96.2%	99.4%	96.5%	99.0%
Background (B)	48.3%	49.2%	47.6%	61.7%	26.6%	32.5%	45.8%	76.9%
Contextual Elements (C)	12.6%	8.3%	15.9%	18.0%	3.8%	6.6%	8.3%	27.9%

Participants represent the inclusion of the objects representing *subject matter* of the image request, e.g. including *humans* in a request for images depicting *humans practicing sports*. Participants were included in almost all query images, and there were no major or significant differences in its use between the two interfaces, the two respondent groups or the three query categories.

Use of background represents any use of background to create a contextual framing for the objects in the query image. The queries which contained a background were classified into three different categories:

1. *Use of a single colour for background.* In these images, a single colour was used to create the background. This was used in 49.5% of the queries including background (Figure 66a).
2. *Use of a multicoloured background*. In these images, the background is created using more than one colour. This was used in 36.5% of the queries including background (Figure 66b).

[56] $T[414] = 0.638$, $p > 0.01$
[57] $T[414] = -2.229$, $p > 0.01$
[58] Note that according to the definition of these criteria in chapter 4, the criteria are not mutually exclusive.

3. *Simple background.* In these queries, colour was not used to create background, but one or more basic elements were used to create a simple background, e.g. using a line to create a horizon. 14.0% of queries with background were placed in this category (Figure 66C).

Figure 66 - Different types of background use. Queries 245 and 415 (Humans and / or animals in nature) and query 236 (A request for dolphins entertaining humans)

There were no significant differences between the two respondent groups with regards to use of background.

The difference in use of background between the two interface types was significant[59], indicating an increased use of background in queries made in VISI. Concerning the way the background was created, *simple background* was used slightly less in Retrievr than in VISI and *single coloured background* was used slightly more, but the differences were small and not significant.

Concerning the difference between the three query types, the difference between query type 1 and 2 was *not* significant[60]. The difference between type 2 and 3[61] and between type 1 and 3 *was* significant[62], indicating an increased use of background in queries based on requests for *complete scenes* compared to queries based on requests for *generic* and *narrative content*. There were no significant differences between the three query types with regards to the way background was created.

Use of contextual elements represents the presence of objects in the query image that provide contextual cues for the query participants. In the framework presented in chapter 4, two criteria were used to describe contextual elements: *symbolic* and *detailed*. However, during the classification work it became clear that it was difficult to distinguish between these. Consequently, these categories were combined into a common category: *Contextual elements*.

---

[59] Mann-Whitney $U[414] = -6.942$, $p < 0.01$
[60] Mann-Whitney $U[310] = -2.394$, $p > 0.01$
[61] Mann-Whitney $U[248] = -4.9$, $p < 0.01$
[62] Mann-Whitney $U[270] = -7.087$, $p < 0.01$

Overall, contextual elements were not used much, but there were some notable differences between the different queries. The difference between the two retrieval systems was significant[63], indicating that contextual elements were used more in VISI than in Retrievr.

There was an observable difference between the two respondent groups: the KHIB group used contextual elements twice as much as the respondents in the IFIM group. However, neither the IFIM group nor the KHIB group used it much, and the difference was not found to be significant at the 0.01 level.

The most notable difference in the use of contextual elements was between the different query types. The slight difference between type 1 and 2 was not significant[64]. The differences between type 1 and 3[65] and between type 2 and 3[66] were significant, indicating that contextual elements were used considerably more in queries based on requests for *complete scenes* than in queries based on requests for *generic and narrative content*.

This shows that almost all the query images contained participants, and about half of the images queries contain background while only 12.6% of the queries contain contextual elements, indicating a low degree of completeness in the queries. Table 31 presents an overview of how the modality criteria are combined in the query images:

Table 31 - Combinations of query contextualization criteria

Category	Overall	Respondent group		Retrieval system		Query type		
		IFIM	KHIB	VISI	Retrievr	1	2	3
A	49.76%	49.7%	49.8%	36.7%	70.9%	65.7%	50.7%	23.1%
B	0.72%	1.7%	0%	0%	1.9%	0%	1.4%	1.0%
AB	36.23%	40.3%	33.0%	45.3%	21.5%	27.7%	37.5%	48.1%
AC	1.21%	1.1%	1.3%	1.6%	0.6%	1.8%	1.4%	0%
BC	0.24%	0%	0.4%	0.4%	0%	0.6%	0%	0%
ABC	11.11%	7.2%	14.2%	16.0%	3.2%	4.2%	6.9%	27.8%
Other	0.72%	100%	1.3%	0%	1.9%	0%	2.1%	0%

Three categories represent 97.04% of the queries:

---

63 Mann-Whitney U [414] = -4.222, $p < 0.01$
64 Mann-Whitney U[310] = -0.571, $p > 0.01$
65 Mann-Whitney U[270] = -4.776, $p < 0.01$
66 Mann-Whitney U[248] = -4.082, $p < 0.01$

- **Participants alone (A)**. These queries only contain the subject matter of the retrieval task, as shown in the three query images in Figure 67.
- **The combination of participants and background (AB).** These are queries in which the object(s) of interests are placed on some kind of background, as shown in Figure 68.
- **The combination of participants, background and contextual elements (ABC)**. These queries represent queries that are *complete* by definition, as illustrated query 4 and query 379 (Figure 69)[67].

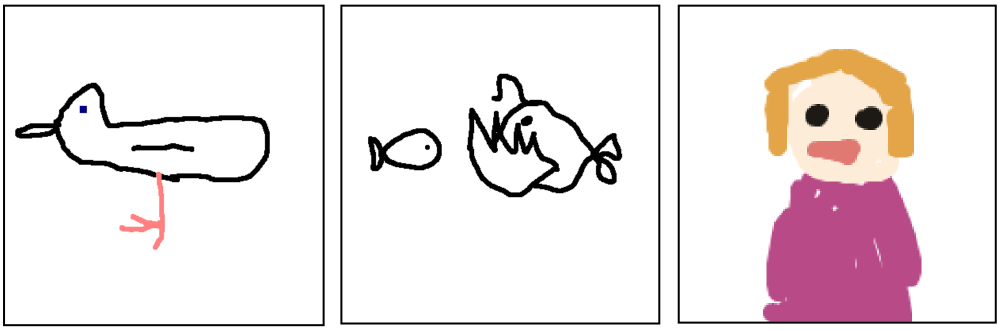


Figure 67 - Queries containing only objects of interest. Query 9 (Seagull), query 337 (a predator attacking a prey) and query 255 (A happy girl).

Figure 68 - Queries with objects of interest and background. Queries 55 (Jumping dolphin) and 408 (People practicing sports) represent objects of interest on a multi-coloured background, while query 3 (Scuba diver) represents objects of interest on a single-coloured background.

67

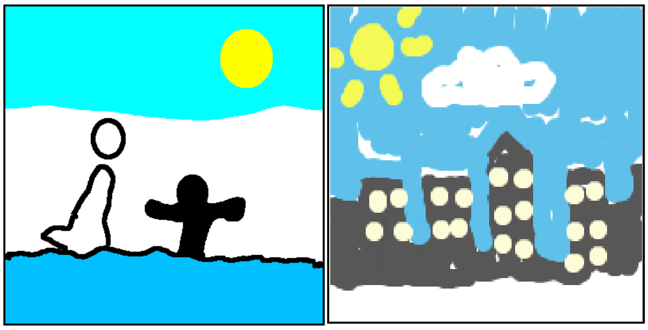


Figure 69 - Use of participants, contextual elements and background (Queries 4 and 379)[68]

The distribution of these combinations is similar if the queries are broken down by respondent group or interface type. The two respondent groups have a similar distribution of the combinations, and no significant differences were found.

There was a significant difference between the two interface types; the three combinations (A, AB, ABC) were the largest in both interfaces, but the magnitude of the categories was different. Queries containing *only* participants were very much more common in Retrievr, while combinations of *participants* and *background* and *contextual elements* were correspondingly less common. The observed difference is significant[69].

Breaking the categories down by *query type* revealed that the use of *only* objects of interest was far more common in type 1 queries for than in type 2 and type 3 queries. Similarly, the combination of *participants* with *background* and *contextual elements* increased in queries based on requests for complete scenes (Query type 3).

The final measurement for completeness was the **subjective evaluation of the overall contextualization**. This was rated on a scale from 1 through 5, where 1 was the lowest score. 50% of the images were rated as 1, 40.1% of the images were rated as 2, while 9.9% were rated as 3. Table 32 presents an overview of the mean contextual score:

Table 32 - Subjective evaluation of contextual modality

Overall	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
1.60	1.56	1.63	1.76	1.34	1.37	1.51	2.08

[68] Note that while the queries represented in Figure 69 are complete by definition, the overall visual modality of these queries must be said to be low.
[69] Mann-Whitney U[414] = -6.832, $p < 0.01$

Generally, the mean score was very low with only minor variations between some of the queries. The slight difference between the two respondent groups was not significant[70].

The queries created in the VISI system was rated higher than the images created in Retrievr. While the difference was small, it was significant[71] .

The most notable query category was type 3 queries. This category obtained a significantly higher mean score than the other categories, indicating that these queries included somewhat more context than the other queries. The differences between query types 1 and 3[72] and types 2 and 3[73] were significant, but the difference between types 1 and 2 was not significant[74].

The descriptive statistics above show that the overall completeness of the query images must be classified as low: There were generally few unique objects present in the queries, the contextual modality was rated low by the evaluators and evaluation of the contextualization criteria illustrated that the query images were generally *simple*, not *complete.* 49.8% of the query images consisted of only participants, 48.3% of the images included background, while only 12.6% of the images contained *contextual elements*. Only 11.11% of the query images were classified as *complete*.

Based on this, it is possible to evaluate the research hypotheses 2.1.1, 2.2.1, 2.3.1 and 2.4.1:

RH2.1.1: The hypothesis must be *accepted*. The respondents created queries that were *simple*, not *complete*, indicating that users prefer creating *simple queries*.

RH2.2.1: The hypothesis must be *rejected*. The KHIB group did not create more complete queries than the IFIM group, indicating that users with a visual background will not create more complete queries than users without this background.

RH2.3.1: The hypothesis must be *accepted*. The respondents created *more complete* queries in the VISI system than in the Retrievr system, indicating that there are characteristics of the VISI system that encourage respondents to create more complete queries than in the Retrievr system.

RH2.4.1: The hypothesis must be *accepted*. There were significant differences in the completeness of the query images in the different query categories, indicating that more complex image requests caused the respondents to create more complete query images. This was particularly true when querying for *complete scenes*.

---

[70] Mann-Whitney U[414] = -0.975, $p > 0.01$
[71] Mann-Whitney U[414] = -6.624, $p < 0.01$
[72] Mann-Whitney U[270] = -8.902, $p < 0.01$
[73] Mann-Whitney U[248] = -6.298, $p < 0.01$
[74] Mann-Whitney U[310] = -2.257, $p > 0.01$

6.1.3 Discussion of the Results

The evaluation raised several questions that needed to be evaluated further:

1. Why did the respondents choose to create simple query images?
2. What caused the observed differences in completeness between the different query types?
3. Why did the images created in the VISI system appear to be more complete than the queries created in Retrievr?

Based on the discussions with the respondents, it seems as if all these three questions seem to be related to three main explanations: The concept of r*elevancy*, problems related to the *drawing process* and the *time required* creating the drawings.

A recurrent theme among the respondents was the concept of **relevancy**. This was primarily used to describe two different, but related, issues: They did not want to spend time and effort to include elements in the query that were not directly relevant for their query task, and they did not want the system to retrieve images with content that was irrelevant for their retrieval tasks. When creating the queries, they wanted to keep focus on the specific task at hand, as illustrated by the following quote from respondent 18:

> *I just thought: focus on the important little core. Birds eating, nothing else. Just keep it simple!*
> ***Respondent 18***

Several other respondents made similar statements. When expressing the queries, they generally wanted to do this as efficiently as possible, and focus on the core of the retrieval task. Several respondents also made comparisons to text based queries, in which they normally only include the subject matter of the query, e.g. when searching for generic images of dolphins, they use simple terms and do not include much contextual information unless this is directly relevant for the retrieval task:

> *If you compare this to text based queries, I don't go about specifying those things [referring to background and contextual elements]. If I want images of dolphins, I just type "Dolphin" and there you go. I never would have considered typing "Dolphin, coral reef, underwater, sea". That's not how it works. So I don't think I'd want to use all those things when just searching for simple objects or animals and so on.* ***Respondent 3***

Similarly, a majority of the respondents discussed the concept of *confusing the system* with additional detail. It appears they felt that including contextual details and background might have a negative influence on the retrieval results: Relevant images might be excluded, or irrelevant images might be retrieved. An example of the former is the following quote from respondent 19. He was asked why he did not include background in a query for a "seagull":

> *It's because... Background might often be of very little relevance. [..] Normally, if the bird is in the mountains, at sea, on an iceberg or in the city, that is basically not relevant.* ***Respondent 19***

In this case, respondent 19 was interested in *all* images of seagulls. He felt that including one particular background in the query might result in images of seagulls in *another* setting would be excluded from the results. Similar explanations of the low use of background and contextual elements were a recurring theme among the respondents. Related to this was that the respondents did not want the system to retrieve images with content *not* relevant to their query task. Several of the respondents mentioned this, stating they did not want to *confuse the system* by including other objects than the object(s) they were attempting to retrieve.

> *I'm thinking... Dolphin is the main thing that gets identified [..]. If I have a dolphin, it's more likely that I find a dolphin with a coral reef, than finding a dolphin if I only draw a coral reef. And it may ruin my results if I include a coral reef. I might find images of a coral reef or even a coral reef and dolphin, but I won't find the images with only a dolphin.* ***Respondent 20***

> *[If I] drew other things than the actual object, there might be a danger of hitting other... That [the system] perceived those objects as other things. It wouldn't know which objects were the main focuses of the image. Respondent* ***11***

In the first quote respondent 20 discusses why he chose not to include contextual elements or background. This quote reflects what many respondents stated: They did not include contextual elements because they did not want to exclude images that *did not* contain these objects, or *include images* that *contained* the contextual elements, but *not contained* the requested objects. Similarly, in the second quote, respondent 11 discusses how he explicitly chose to exclude other objects than the actual objects of interest. In this particular example, he explains why he chose not to include any contextual elements such as clouds or the sun in the query represented in Figure 70. He felt that including these might take focus away from the actual task at hand: Finding dolphins entertaining people in a boat.

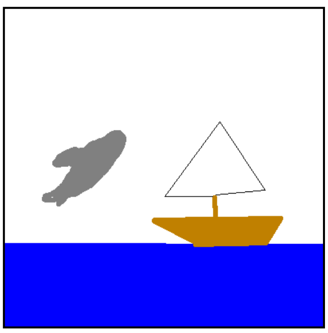


Figure 70 - Query 91, respondent 11 ("Find images of a dolphin entertaining people in a boat")

The above statements might also help explain why there was an increased degree of completeness in the type 3 queries. When querying for images of generic objects, the respondents were mainly interested in retrieving images which contained the objects of interest, regardless of context and composition. This is more or less directly comparable to generic text-based queries: It is unlikely that text-based queries will be very specific in terms of contextual parameters if the requests are generic. A similar effect may be seen here: When asking for dolphins, the respondents drew only dolphins. But in more detailed requests, primarily when requesting *scenes*, the respondents might have a clearer idea of the actual image they wanted to retrieve. If the request was specific (e.g. *dolphins in an arctic landscape* or *animals gathered in a rural setting*), the respondents reported that they had to be more specific when creating the query images. In these cases, the requested objects *should be in a given setting*, and there respondent might want to exclude images in another context.

Another recurring theme was problems related to the actual **drawing process**: The respondents reported problems caused by their lack of drawing skills, they often felt they did not know *how* they should draw the background or *what* contextual elements they should include, and they experienced that there was not enough *space on the canvas* to create complete queries.

First of all, some of the respondents claimed that they found it difficult to draw, or had other difficulties related to the actual drawing process. Some respondents reported that the actual drawing process was difficult. They were not able to draw the objects in a way they were satisfied with. Consequently, they only drew the minimum number of objects they felt were necessary for the query, e.g. the actual dolphin in a request for dolphins, or a seagull and its food in a request for "seagulls eating". It also seems as if the actual minimum number of objects they felt they were required to draw increased with the level of detail defined by the retrieval tasks. Similarly, image requests with a higher level of detail, e.g. "*Find images of one or more animals swimming in an arctic environment*" , provided more specific details, and the respondents often felt that they had to include more when creating query images based on these.

Another issue related to the drawing process was that the respondents often did not have a clear idea of *what types* of contextual elements they should include. In the generic queries, they often did not have more information than "find a dolphin". While they had a mental image of what dolphin looks like, identifying the potential background and contextual elements might be difficult. An example of this is the following quote from respondent 24. When asked why he used few contextual elements in his queries, he stated that he did not know *which* contextual elements he should have included:

> *Yes, I focused more on the actual object. That is, the thing I wanted to retrieve, I didn't necessarily go so much for the actual surroundings. For example, if I'm looking for a dolphin, I'm just going to draw a*

dolphin, rather than everything around it. Because I don't know what should be around it, in a way. ***Respondent 24***

Several other respondents mentioned similar situations. They were unable to properly identify and describe contextual elements and background: they were unable to imagine or visualize *what* this background should be, or *how* they should contextualize their participants. This can also help explain the way the respondents *used* background. 49.5% of the query images that included background used a single colour to describe this background. The respondents often resorted to including a simple background without contextual elements, e.g. using "blue" to illustrate the ocean or "green" to illustrate a forest or a playing field. This is illustrated by the following quote from the think-aloud session with respondent 2 during the process of creating the query illustrated in Figure 71.

I don't know… [..] I think I need a background for this. [..] Blue? I'll fill it with this blue colour. I have no idea what it actually should look like. There are probably a hundred better ways of drawing the background, but… [..] No, I have no clue whatsoever. Blue will have to do. ***Respondent 2***

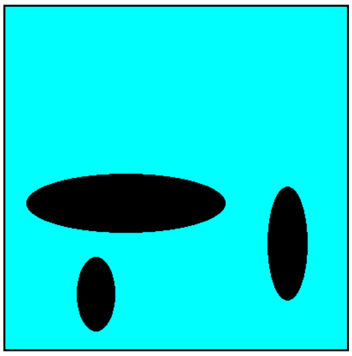


Figure 71 - Query 15 (Find images of an attacking shark), made by Respondent 2.

Several respondents also claimed that there was simply not enough *space* to draw a lot of contextual elements. In many cases, they used the entire canvas to draw the objects of interest, leaving no room for background and contextual elements. This was particularly true for the Retrievr system. Several respondents complained that the size of the canvas was too small and that the pen-sizes were too large for the canvas size. This combination resulted in not enough room to include background or contextual elements. There was no tool for filling large areas of the image, and if they tried to use the freehand tool for this, they often drew over the objects of interest, something they tried to avoid. Respondent 29 discusses this as one of the reasons for not including background or contextual elements in her queries. The query she refers to is presented in Figure 72.

Well, in the rabbit vs. hunter scenario… There were a lot of things [I] should have drawn, because you would like to include things such as green grass and a blue sky, and if you forget, if you draw the hunter *first, and then attempting to draw a blue sky, this becomes difficult, because you can't draw over him. I could have drawn the blue sky first, but… I guess I didn't.* ***Respondent 29***

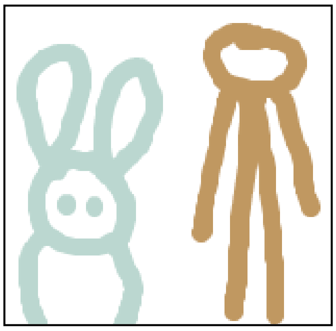


Figure 72 - Query 361 ("Find images of a person hunting a rabbit"), made by respondent 29.

This is likely a contributing factor to the difference in use of contextual elements between the two interfaces.

The final issue covered by the respondents, was **the time required to create the query images**. While the other reasons were mentioned by most of the respondents, this was only mentioned by 4 respondents. However, these four respondents were very vocal about this: They did not want to *spend a lot of time* drawing things they felt were not very relevant for the queries. This is illustrated by the following quotes from respondents 6, 12 and 17:

I don't want to sit and draw in detail in order to retrieve an image. I want to do it quick and easy. I want to draw like this, with few details, and retrieve relevant images. ***Respondent 6***

I certainly don't want to spend time and effort on drawing a coral reef. If this is going to work, I definitely think so, I don't think me or anybody else is willing to draw colourful and detailed backgrounds, if we want to retrieve dolphins. ***Respondent 12***

Well, It's to save time and avoid complicating it. It's like... When you work with drawings and images, you learn to, all the time, you try to simplify things. [Including contextual details] would complicate [The QBD process] further and make it even more time demanding. ***Respondent 17***

This at least indicates that the extra time it takes to include contextual elements and background may result in these details being omitted. Looking at the mean time spent creating queries with different degrees of completeness (Table 33) illustrates this further. Queries containing only objects of interest (A) took 60 seconds less than queries which combined objects of interest with background (AB) and 96 seconds shorter than queries that included *both* background and contextual elements (ABC).

Table 33 - Mean time spent in different completeness categories

Completeness category	A (N=189)	B (N=3)	AB (N=134)	AC (N=5)	BC (N=1)	ABC (N=44)	Other (N=3)
Mean time (In seconds)	65.37	44.00	125.84	215.40	30.00	161.48	7.67

6.1.4 Summary of the Results

The respondents did not create very complete queries, particularly when creating queries based on requests *generic content*. Generally, they reported that they wanted to keep the queries as simple as possible, and added only the minimum level of detail required for expressing their image request. Several reasons for this were found:

- The respondents did not want to add objects that were *irrelevant* for the query and could limit the query images.
- The respondents did not want to *confuse the system* by including contextual details. They were afraid that the system might put too much weight on these, removing focus from the subject matter of the query.
- Some respondents reported difficulties when drawing, and decided to keep the number of objects in the query image to an absolute minimum, not wanting to face the challenge of drawing a large number of different objects.
- Some respondents reported difficulties determining *which* contextual elements they should include. Similarly, they often did not know *how* they should draw the background, or did not *care* what the background was like. This was particularly true for queries for *generic content*. Consequently, they did not add background or contextual elements to these queries.
- Most respondents reported that the combination of a small canvas and a large pen size in Retrievr made drawing background and contextual elements difficult. Consequently, they created less complete query images in Retrievr.
- Some respondents reported that they wanted to keep the effort and time spent on creating the queries as low as possible, and that including background and contextual elements simply took too much time.
- The degree of completeness increased with the level of detail in image requests, particularly when querying for complete scenes. In these cases, the respondents added as much details and background as they felt were necessary in order to provide a minimum specification of

their image requests, e.g. adding *the sea* and *mountains* to a query for a dolphin in order to inform the system that they wanted images of dolphins *in a fjord*.

6.2 Use of Colours

Use of colours describes the degree the respondents used colours in an active way to create query images with a high visual modality. A photographic image consists of a very large number of unique colours. For example, the image in Figure 73 consists of 50.737 unique colours[75]. While one cannot expect that a visual query image would have this many colours, a *realistic approach* to creating query images should include an active use of colours.

Figure 73 - Image of a dolphin. The image consists of 50737 unique colours.

6.2.1 Research Hypotheses

In order to evaluate if colour has been used to create realistic queries, additional sub-hypotheses were created for each of the four main hypotheses:

- **RH2.1.2:** Respondents do not make much use of colours when creating visual query images
- **RH2.2.2:** Respondents with a visual background make more use of colours than respondents without this background
- **RH2.3.2:** Respondents make more use of colours in the VISI interface than in the Retrievr interface.
- **RH2.4.2:** The use of colours in the query images increases with the complexity of the image requests

These hypotheses have primarily been evaluated using three measurements:

1. The actual number of colours used to create the query image
2. A classification of colour use according to the *colour modality marker*

[75] The number of unique colours were counted using a simple software application

3. A subjective evaluation of the colour use in the query

According to the modality criteria, a query image created using an active colour use is an image created using multiple colours, the elements in the image is depicted using varied and realistic colours and illumination effects.

6.2.2 Evaluation of Hypotheses

The first measure used to evaluate colour use was the **number of colours used to create the query images**. Two different approaches were used to count the number of colours used in the query images. For the queries made using VISI, the number of colours was extracted using a software application. This could not be used for the Retrievr images, as the use of screen-capture to extract the query images resulted in an anti-aliasing effect. While this does not reduce the quality of the query images for purposes of classification, it led to an increase in the actual number of colours represented in the digital image. As a result of this, automatic extraction of colours was not possible. Consequently, the number of colours used in each image had to be counted manually by inspecting the interface videos and query images.

The mean number of colours used was 3.62, with a standard deviation of 1.781, a minimum of 2 and a maximum of 12. Table 34 presents an overview of the mean number of colours used by the two respondent groups, the two systems and the three query types. Table 35 shows the frequency distribution of the colours used.

Table 34 - Mean number of colours used in the query images

Overall	Respondent group		Query system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
3.62	3.72	3.45	3.8	3.32	3.09	3.66	4.39

Table 35 - Frequency distribution of colour use

		Respondent Group		Query system		Query type		
Colours	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
2	34,10%	30.90%	37.80%	30.90%	41,10 %	44.60%	38.20%	14.40%
3	23,70%	26.0%	20.60%	21.90%	24,70 %	27.10%	16.00%	26.00%
4	16,50%	12.70%	19.30%	17,60 %	14,60 %	16.30%	13.90%	20.20%
5	10,40%	14.90%	7.30%	10,50 %	10,80 %	6.0%	11.80%	16.30%
6	9,40%	7.70%	10.70%	11,70 %	5,70 %	3.0%	15.30%	11.50%
7	2,90%	5.00%	1.30%	4,30 %	0,60 %	1.80%	3.50%	3.80%
8+	3%	2.90%	3.10%	3,20 %	2,40 %	1.20%	1.40%	7.80%

Few colours were used in the query images. Query images with only 2 colours represent the largest category (34.1%), while 57.8% of all query images had 3 of fewer colours. Overall, the only category which is notably different from the other categories is the higher use of colours in type 3 queries. There was a slight difference between the two respondent groups, but this difference was not significant[76]. The difference between the two retrieval systems was significant[77], but very small. The differences between the three query types were significant: More colours were used for type 2 than 1[78] and for type 3 than 2[79].

The second measure of colour use was the **modality criteria related to colour use**. First of all, note that in the framework presented in chapter 4, the first colour modality criterion was *monochromatic*, indicating that a query image was created exclusively using a single colour on a neutral canvas. However, when classifying the images, this criterion proved difficult to use. Some of the query images were created using a single colour on a white canvas, as illustrated in Figure 74. According to the framework, these queries should have been classified as *monochromatic.* However, it is possible that the respondent did not select these colours on purpose, particularly for queries made in Retrievr. When the initiates a new query process in Retrievr, the colour does not reset to a default colour. Unless the user selects a new colour the query will be drawn using the last colour used. It is possible that the affordance of Retrievr's design for making quick sketches in the already selected colour may have influenced some of the queries created in Retrievr. However, the respondents were not asked directly about this and consequently the classification followed the framework strictly and classified these queries into "simple colour use", as they also satisfied this criterion (i.e. creating objects using a single colour).

Furthermore, several of the respondents stated or implicitly expressed that they used "black lines on a white canvas" in order to create "colour neutral" queries (discussed in detail below). Consequently, the "monochromatic" criterion was replaced with a new criterion; "Lack of colours", indicating that the query objects were drawn using black lines on a white canvas.

---

[76] $t[412] = 1.030$, $p > 0.01$
[77] $t[412] = 2.709$, $p < 0.01$
[78] $t[412] = -3.187$, $p < 0.01$
[79] $t[412] = -3.069$, $p < 0.01$

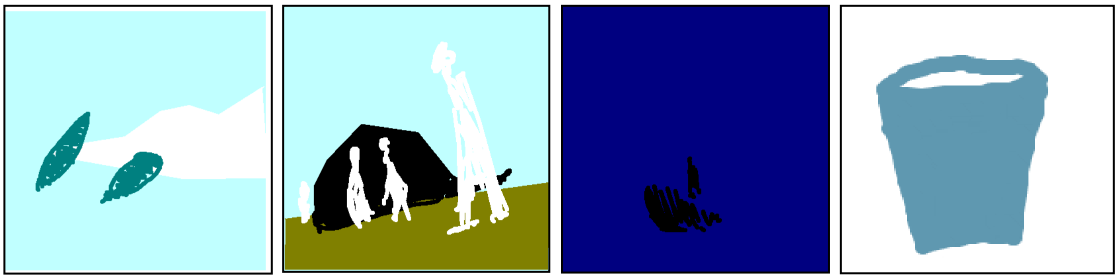
Figure 74 - Three queries created using a single colour (Queries 266, 352 and 375).

According to RH4.2, the degree of *query realism* is determined by the degree to which the query images are created with an active colour use. If colours had been used actively by the respondents, this would have been indicated by using varied colours (C), colour gradients (D) and illumination (E) where applicable, with few query images were created without using colours (A) or with the objects depicted in a single colour (B). Table 36 shows an overview of the different colour modality criteria, and the degree of which they have been used[80].

Table 36 - Colour use classified by colour modality criteria

		Respondent group		Retrieval system		Query type		
Modality criteria	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
Lack of colours (A)	24,60 %	24.9%	24.0%	26.6%	20.9%	28.3%	29.9%	10.6%
Basic colour use (B)	61,10 %	63.5%	59.2%	64.8%	55.1%	50.6%	58.3%	81.7%
Varied colours (C)	25,10 %	23.8%	26.2%	19.5%	34.2%	26.5%	22.2%	26.9%
Colour gradients (D)	0 %	0%	0%	0%	0%	0%	0%	0%
Illumination (E)	0 %	0%	0%	0%	0%	0%	0%	0%

The two criteria which were considered as indicating the highest colour modality (D and E), were not used at all. The criterion used most was *basic colour use* (B). This represents query images where one or more of the query objects were created using a single colour, as seen in Figure 75.



[80] Note that category A is mutually exclusive with other categories. The remaining categories are not exclusive.

Figure 75 - Use of single colours to create image elements. Dolphins and icebergs in query 264, humans and whale in query 393, the boat in query 389 and the glass in query 350.

The next most used criterion was *varied colour use (C)*. This represents queries where one or more of the query objects are depicted using two or more colours, as seen in Figure 76.

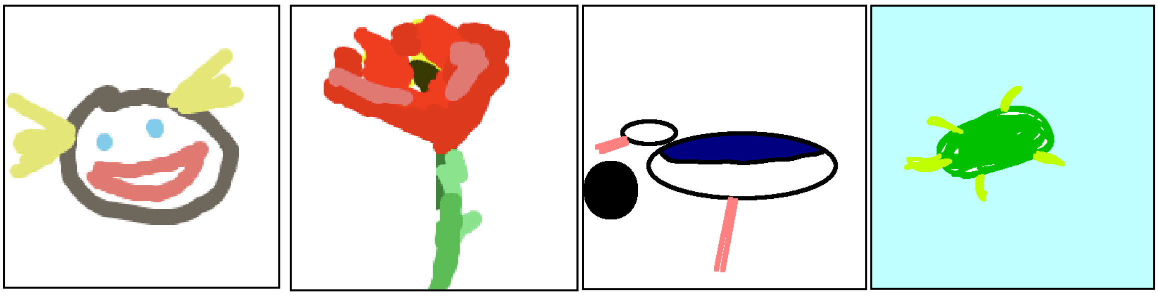


Figure 76 - Use of multiple colours to depict an image element. The happy girl in query 356, a flower in query 296, a seagull in query 12 and a turtle in query 388.

24.6% of the queries were created without colours (A). These are queries which are created without use of colours, as shown in Figure 77:

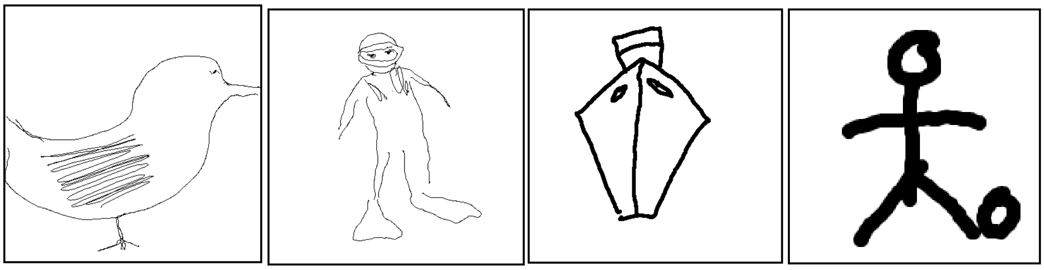


Figure 77 - Queries created without use of colours. A bird in query 62, a scuba diver in query 27, a ship in query 336 and a person practicing sports in query 276.

Table 37 presents an overview of the combinations of colour use used in the query images. Queries created *only* using basic colours are by far the largest group, including more than half of all query images. Queries created using no colours (Category A) represent the second largest group. Together, these two groups represent approximately 75% of the queries. Queries created using some kind of varied colours (C and BC) represent approximately 25% of the queries.

Table 37 - Comparisons of colour modality criteria

Criteria	Overall	Respondent group		Retrieval system		Query type		
		IFIM	KHIB	VISI	Retrievr	1	2	3
A	24.4%	24.90%	24.00%	26.60%	20.90%	28.30%	29.90%	10.60%
B	51.0%	51.40%	50.65%	53.90%	46.20%	45.20%	47.90%	64.40%
C	14.5%	11.60%	16.70%	8.60%	24.10%	21.10%	11.80%	7.70%
BC	10.1%	12.20%	8.65%	10.90%	8.90%	5.4%	10.40%	17.30%

There were no significant differences between the two respondent groups in any of the colour modality markers, indicating that the respondents used colour in a similar manner.

There appeared to be some differences between the two retrieval systems. There were more queries created using colour in Retrievr than in VISI, particularly queries created using only varied colours (Category C). While the only significant difference was for the increased use of varied colours[81], this indicates that the respondents used colours slightly more actively in Retrievr.

The most notable differences are between the three query types. Type 3 queries were more frequently created with colours than type 1 and type 2 queries. Similarly, there was an increased use of *simple colour use* (B) in type 2 and type 3 queries. These differences were significant. While there appeared to be a notably larger use of query images created *only* using varied colours (C) for type 1 queries, this difference was not found to be significant.

The final measure of colour use is the **subjective evaluation of colour modality**. The mean score given to all queries was 1.85, with a standard deviation of 0.772, a minimum of 1 and a maximum of 3. 38% of the queries were rated as 1, 38.5% of the queries rated as 2, while 23.5% were rated as 3. Table 38 shows the modality scores for the two groups, the two interfaces and the three query types:

Table 38 - Overview of colour modality scores

Overall	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
1.85	1.81	1.89	1.85	1.85	1.65	1.87	2.15

The overall score was low, and there were no significant differences between the two respondent groups or the retrieval systems. The main differences are between the query types. Type 1 queries were rated somewhat lower than type 2 queries, but the difference was not significant[82]. Similarly, type 3 queries were rated slightly higher than the other query types, indicating that these were created using colours in a more active manner. The differences between type 1 and 3[83] and type 2 and 3[84] were significant.

Based on the above, it is possible to partially evaluate research hypotheses 2.1.2, 2.2.2, 2.3.2 and 2.4.2:

[81] Mann-Whitney U [414] = -3.334, $p < 0.01$
[82] Mann-Whitney U[310] = -2.421, $p > 0.01$
[83] Mann-Whitney U [270] = -5.207, $p < 0.01$
[84] Mann-Whitney U [248] = -2.825, $p < 0.01$

RH2.1.2: The hypothesis must be *accepted*. The respondents did not use colour very actively when expressing the query images.

RH2.2.2: The hypothesis must be *rejected*. There were no significant differences between the two respondent groups with regards to their use of colours.

RH2.3.2: The hypothesis cannot be fully accepted or rejected based on the data collected. The query images created in VISI used slightly *more* colours than the query images created in Retrievr, while the query images created in Retriever had *varied colours* significantly more than the query images created in VISI. The former indicates that colours are used in a higher degree in VISI, while the latter indicates that colours are used in a higher degree in Retrievr. However, all the queries were created without very active colour use, regardless of retrieval system. No significant differences were found in the *subjective evaluation* of the queries created in the two interfaces. The hypothesis cannot be evaluated fully based on the empirical data available. However, even if there *is* a significant difference between the queries created in the two interfaces, the difference is so small that it is of little consequence.

RH2.4.2: The hypothesis must be *accepted*. The respondents used colours more actively when creating queries based on level 2 and level 3 requests than level 1 requests.

6.2.3 Discussion of the Results

The above evaluation raises some questions than should be evaluated further:

1. Why didn't the respondents use colour much when creating the query images?
2. What caused the observed differences between the different query types?
3. Why didn't the respondents use *gradient colours* or *lightening effects* in any of the queries?

The primary question is why didn't the respondents use colours actively when creating the queries? First of all, all respondents agreed that they wanted to keep the queries *as simple as possible*. Most respondents stated that including colours was necessary in many cases. They stated that they felt that using colours was a very quick and easy way of adding the *required level of detail* to the queries, particularly in order to highlight important aspects of the request. Two examples of this are the queries in Figure 78. When creating query 37 (Figure 78a), respondent 4 stated that she felt that she should add a blue square at the bottom of the query in order to identify that the objects in the image were situated in the ocean. Similarly, when creating query 198 (Figure 78b), the respondent stated that adding "grey" and "blue" to the shape and the background would help both him and "the computer" to understand that it was a dolphin (or at least a fish), and not a bird.

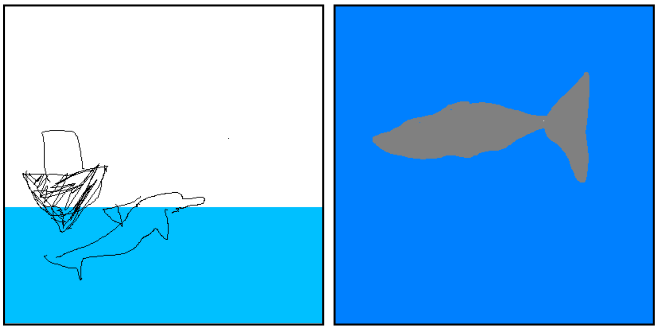


Figure 78 - Examples of colour use (Query 37, respondent 4 and query 196, respondent 21)

However, they also stated that they felt adding colours increased the complexity and time required to create the queries. Table 39 shows the mean number of seconds spent on creating queries using the different combinations of colour. Using colours in an active manner requires spending more time on creating the query image.

Table 39 - Mean time in seconds spent drawing using different colour combinations

Criteria	Images	Mean time
Lack of Colours	101	61.88
Simple Colours	211	98.02
Varied Colours	60	105.38
Simple and Varied	42	172.26

There are several reasons for the increase in time required. Using colours required more actions in the interfaces, increasing the time spent drawing the query. This is particularly true for the Retrievr system, where the user needs to select both a base colour and a colour shade. However, the user also has to determine *which* colour they should use. When drawing without colours, the user only needs to focus on lines and shapes. Introducing colour demands that the user must decide *what* colours to use, and they might not necessarily have a clear idea of what colours a given object has. Using black lines on a neutral canvas may be a very efficient way of expressing the queries, as illustrated by respondent 16:

> *When you do black on white it is a very informative method. It's like cutting it to the bone in terms of efficiency when retrieving something. You just have to sketch something very fast in order to impart specific information quickly and obtain results. It's like, how fast can you do it without using more time... It's like being clear and direct when talking.* ***Respondent 16***

Several other respondents mention this, particularly in relation to simple requests for generic content, as illustrated by respondent 17:

Well, if I'm going to look for something simple, such as a seagull, I'll just draw it as quickly as possible. [Using colours] would only complicate matters further, and it would take a lot more time. ***Respondent 17***

Another very important element in this appears to be the relationship *shapes and objects* and *colours*. Most of the respondents did not consider the drawing canvas in either of the retrieval systems as *white*, but as *neutral*. Similarly, when they drew objects using *black*, they generally did not consider this as the *colour black*, but as a *neutral colour*. Consequently, in queries such as query 338 and query 369 (Figure 79a, b) the respondents did not regard this as black lines on a white canvas, but as lines, shapes and objects drawn on a neutral background.

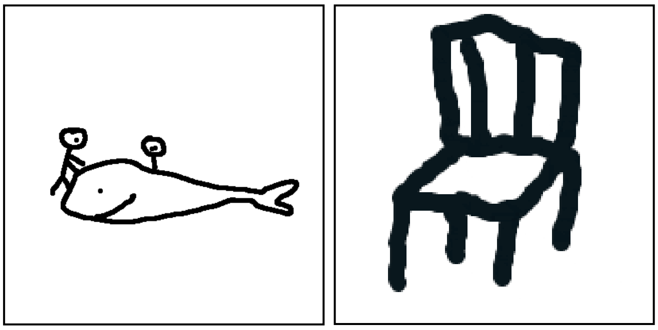


Figure 79 - Query 338 (Respondent 29) and 369 (Respondent 30).

They compare this to a white sheet of paper or a white canvas upon which they draw objects using a pen or pencil, without considering the actual colour of the pen, as illustrated by respondent 16:

I drew using black because I thought "drawing". [..] That is, I consider "black" as if it was a pen. ***Respondent 16***

Several respondents used this approach as a way to express queries in *colour neutral* manner. There were several examples of queries in which the respondents wanted to search for a given object or a given shape, without considering the actual colour or colours of these objects or shapes, as illustrated by respondent 26:

Now, when I was trying to find a chair, I had to... I know what a chair looks like, but I had no idea what colour it should be. [..] I used black, or used a random colour... [A chair] can have any number of different colours, and I didn't want to exclude anything. ***Respondent 26***

This quote also illustrates that a portion of the queries classified as "Simple colour use" may be expressed in a colour neutral manner, as illustrated by the queries in Figure 74 (page 142).

The desire for expressing queries in a colour neutral manner is directly related to the issues of *relevancy* and the desire *not to confuse the system* identified in the discussion of query completeness, as illustrated by respondent 6:

> *Well, I chose to colour the objects I was supposed to retrieve and were mentioned in the text, such as dolphin. And, I know it is greyish, and I think that the sea is blue, and I coloured the boat too, but that was mostly to make it distinct. I chose not to colour the background, because if I had coloured the sky light blue, then I might have excluded images with land in the background. Or, say it was a sunset, and the sky was reddish. There are many things I feel I would have missed if [I used colour].* ***Respondent 6***

Using few (or no) colours is a strategy to reduce the chance of excluding relevant images based on the colours in the query image. Several respondents stated that they would either like to have access to a *colour neutral* drawing tool or be able to instruct Retriever to disregard colour similarly to the *colour query parameter* present in the VISI interface.

The largest observed differences were between the three different query types, particularly between type 1 and 2, and type 3. The subjective evaluation was significantly higher for type 3 queries. A significantly higher number of colours were used in type 3 queries, and there was a significantly more active use of colours in type 3 queries, i.e. fewer query images without colours and more query images with simple and nuanced colour use than for types 1 and 2.

The most likely explanation for these observations is the way the respondents treated the different request categories. The respondents generally preferred to keep the query images as simple as possible, particularly when requesting images with generic contents. As noted in chapter 6.1, query images drawn for requests for *complete scenes* (Type 3) were generally more complete than queries made based on requests for *generic and narrative* content (Type 1 and 2 queries). Comparing the number of colours used in these queries showed that there were differences. In queries without background, the mean number of colours used was 2.88, while the mean for queries *with* background was 4.46, a significant difference[85], and a significant correlation between these was found. Similarly, the mean number of colours used in queries *including* contextual elements was 4.87, compared to 3.44 in queries *not including* contextual elements, also a significant difference[86]. Furthermore, there was a significant correlation[87] between the *number of unique object*s in the query and the number of colours used, as illustrated by the queries in Figure 80. The shark is created using the same (2) colours in both images, but when then query includes another animal (the fish) in another colour, the number of colours increase.

[85] $t[412] = 10.4$, $p < 0.01$
[86] $t[412] = 5.606$, $p < 0.01$
[87] Kendall's tau-b $[414] = 0.265$, $p < 0.01$

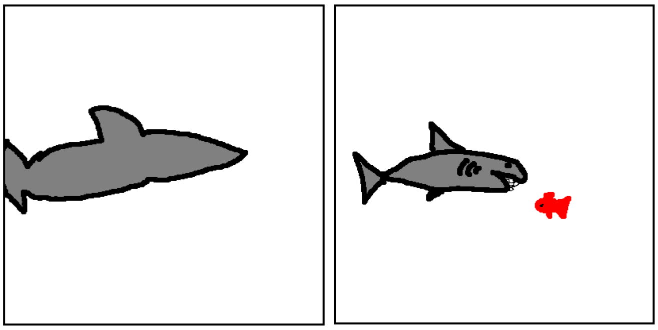


Figure 80 - Two queries containing a shark Query 46 (Find images containing a shark) and query 44 (Find images of a shark attacking another animal).

There are also examples of colours being used in the query images to highlight specific properties of the retrieval task, which may be a contributing factor to the difference in number of colours used in query images drawn for category 1 and 2 requests. Several of the respondents highlight two examples of this: *Injuries and violence* and *emotions*.

Several of the query images drawn for *wounded animals* or *attacking animals* had an increase use of the colour *red*, as illustrated in Figure 81.

Figure 81 - Respondent 17s depictions of seagulls (Queries 134, 136 and 138). Whitespace have been cropped around the central motives.

These queries were made by respondent 17, who generally used few colours in his query images (only 3 of 10 were made using colours). He wanted to express his queries without colours to keep them as simple as possible. All but one of the queries containing animals were expressed using black lines on a white canvas. However, when attempting to draw a "wounded seagull" he included the colour red:

> *[Drawing the wounded seagull] was more problematic. As I said, I wanted to keep things as simple as possible. The bird lies on the ground, maybe it's bleeding. I had to use some blood to illustrate that it's bleeding. Otherwise this would be difficult to show.* ***Respondent 17***

Several other respondents used the colour *red* in a similar manner, illustrated by the query images in Figure 82:

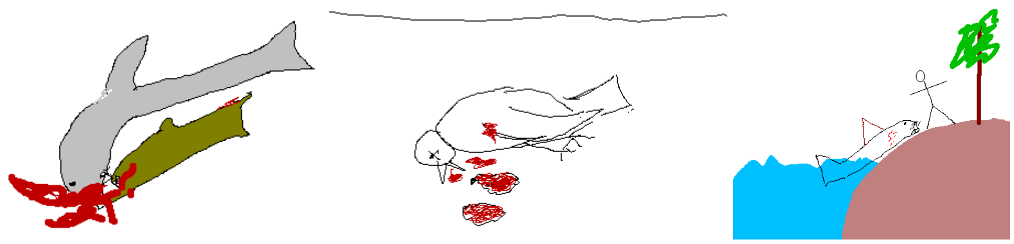


Figure 82 - Some queries using "red" to illustrate injury or violence (Queries 76,148 and 208). Whitespace have been cropped around the central motives.

> *I'm not that good at adding details, so I tried to put more weight on the colours, trying to use the best, the most correct colours. For example, that one with the shark attacking a wounded prey, I tried just to highlight the red, indicating a wound or some sort of blood, meaning that it was injured [..], and I used that spray tool [texture pen] and spread it a bit out, like, you can see that it's... It's running down, in a way.* ***Respondent 21***

It should be noted that these queries (violence or injuries) may represent very particular cases, and there might be very strong connotations to the use of *red blood* in these cases. However, a similar example is in some of the queries for *a happy girl*. *Humans* were often represented using a single colour (or without colours). 66.4% (110) of the query images containing humans were created in this manner, as illustrated by the queries in Figure 83.

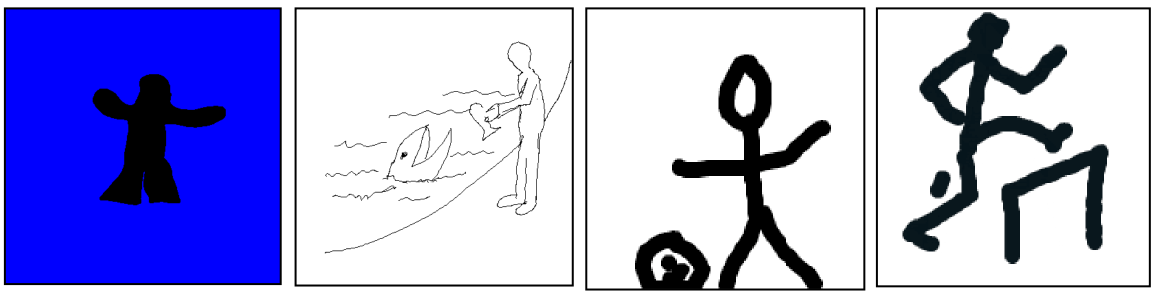


Figure 83 - Humans depicted using a single colour (Queries 3, 147,214 and 373)

However, in the queries for *a happy girl* (Task 27) there was a notable increase in the use of *varied* colour use, i.e. the query objects were created with more than one colour. Note that only 18 queries were made based on this task, and the difference was not significant[88]. However, several respondents stated that they had to use colours in order to express that this was 1) *a girl* and 2) *a happy girl*, as illustrated by respondent 30:

---

[88] Mann-Whitney U[110] = -2.142, p = 0.032

[When querying after the happy girl] I think I was more conscious on the actual colours I used. It's like... I don't know if [the system] has any ideas of what colours are feminine, but I thought that pink was a girl's colour. I felt I had to use [pink colour] to express this. A happy girl is pink, or at least a girl is pink. ***Respondent 30***

Several other respondents made similar remarks. They stated that they had to use *blonde hair* to express that it was a little girl, or that they used light pastel colours to express this query, e.g. they felt that these colours somehow *represented* a "happy girl". Some examples of this are presented in Figure 84. While one should be careful to draw any conclusions based on the relatively thin empirical material, there are at least *some* indications that the respondents used colours to highlight particular requests.

Figure 84 - Some queries for "A Happy Girl" (Queries 192, 245, 255, 356 and 410.

A final question is why there was no use of colour gradients and illumination effects. Two of the respondents stated that they missed a tool for expressing colour gradients, but none of the other respondents mentioned this or said that they would use such a tool. A probable main factor behind the non-existing use of colour gradients and illumination effects is that creating these effects is complicated, even if the respondents had access to a gradient tool. The respondents were reluctant to use coloured backgrounds and use a large number of colours to create the queries. Including illumination effects such as light sources or shadows complicates the drawing process considerably. The queries were expressed using very few compositional structures (described in chapter 6.4) and without regard to the *depth* in the image and the *volume* of the depicted objects. Placing light sources or adding shadows requires that the drawer decides the actual placement of the light sources (inside or outside of the image frame), or has a notion of the *volum*e of the objects for creating shadows. Consequently, introducing illumination effects may be too difficult or too demanding for the user. And, as noted by respondent 16, users might not be very interested in creating very painstakingly made and beautiful images when they're only trying to retrieve images. They might not see the need to, or want to, include illumination effects in the query images:

Well, I could probably have used some more time on colours and adding realistic details, such as colour gradients or illumination. You would normally see a lot of this [in images with animals underwater], a *gradient transition of colours. But it really depends on how much time you really want to spend creating a beautiful image when you're just looking for other images.* ***Respondent 16***

6.2.4 Summary of the Results

The respondents generally used few colours when creating the queries, and the colours were not used very actively. Queries for *generic content* were expressed using fewer colours and simpler colour use than queries for *narrative content* and *complete scenes*, and queries for *complete scenes* were created using considerably more colours and more active colour use than queries for *narrative content*. Several reasons for this were identified:

- The respondents stated that they wanted to keep the queries as simple as possible, and while colours could help them highlighting important elements of the query, having an active use of colours would increase the time and effort required to express the queries.
- Several respondents, particularly in the KHIB group, stated that they often kept the use of colours low, or used "black and white" as a neutral query expression method, not wanting to exclude objects with different colours, or not wanting to consider *which* colours should be used.
- There were some differences in colour use between the two interfaces, but no explanations for these observations could be found.
- There were some indications that colours sometimes were used as a tool for expressing certain types of *narrative* content, e.g. using pastel colours in order to indicate "a young girl", or use the colour "red" to indicate "violence" or "injuries".
- The use of colour was also related to the degree of *completeness* in the queries. Queries for *complete scenes* were generally *more complete* than other queries. The respondents added as much details to the query image as they felt were necessary in order to provide a minimum specification of their image request. As the respondents often used colours to describe or add detail to the background, this resulted in queries for complete scenes having a more active colour use than other query categories.

6.3 Representation of Query Participants

Representation describes the degree of abstraction used when representing the elements in the query image. The elements in a query image can be drawn using any level abstraction; from using simple geometric primitives as representation to highly realistic renditions.

6.3.1 Research Hypotheses

In order to evaluate the degree of abstraction used, an additional set of sub-hypotheses was created based on the four main hypotheses.

- **RH2.1.3:** Respondents will depict query image participants as geometric primitives without using representational components

- **RH2.2.3:** Respondents with a visual background will depict image participants more realistically than respondents without this background

- **RH2.3.3:** Respondents will depict image participants more realistically in the VISI system than in the Retrievr system

- **RH2.4.3**: The degree of abstraction decreases with the complexity of the image requests

These hypotheses have primarily been evaluated using two evaluation measurements:

1. A classification of representational modality according to the *representational modality marker*
2. A subjective evaluation of the representational modality

According to the modality criteria, a query image created with a low degree of abstraction is an image where the elements are depicted as realistic outlines, using both symbolic and detailed visual cues and realistic texture patterns.

6.3.2 Evaluation of Hypotheses

The first measure of representational realism was the **representational modality criteria**. An overview of how the query images satisfied these criteria is shown in Table 40.

Table 40 - Categories of representational criteria

		Respondent group		Retrieval system		Query type		
Category	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
Geometric Primitives (A)	16.4 %	14.9%	17.6%	11.7%	24.1%	6.0%	23.6%	23.1%
Outlines (B)	89.6 %	93.4%	87.1%	95.7%	80.4%	94.6%	82.6%	92.3%
Visual Cues (C)	47.6 %	56.4%	39.9%	59.4%	27.2%	44.6%	60.4%	32.7%
Texture (D)	3.9 %	6.6%	1.7%	6.3%	0%	1.8%	7.6%	1.9%

There were two problems using the framework described in chapter 4.1, identified by all three evaluators. For some queries, it was difficult to determine whether one or more objects should be classified as *geometric primitives* or *outlines*, as illustrated in Figure 85 and Figure 86a. In the former case, the ball is represented by an outline, which is also a geometric primitive (circle). In the latter example, ovals were used to create two whales and a boat, seen from above. In the case of objects that have a real-world shape identical to a geometric primitive (e.g. the ball), these were considered as *outlines*. Objects that *do not* have a real-world shape identical to the geometric shape (e.g. the whales and the boat) were generally considered as *geometric primitives*. However, there were some query images containing some objects on the borderline between outlines and geometric primitives, and these may have introduced some uncertainty in the classification. However, there were few such cases, and considering the large number of images classified as using "outlines", it is unlikely that the overall impact these border cases had on the results was large.

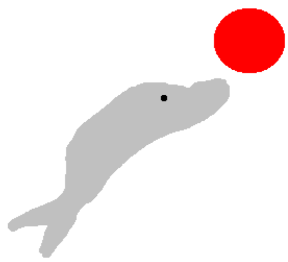


Figure 85 - Outline or geometric primitive? (Query #238). The image has been cropped to the main motive.

Next, the framework used to classify the images used two different categories of representational components: *symbolic* and *detailed*. However, when using the framework all three evaluators had difficulties distinguishing between the two categories. Furthermore, very few representational components were classified as *detailed*. Consequently, the two categories were combined in a single category: *visual cues*.

Geometric primitives were used in 16.4% of the queries, indicating that the respondents did not use this to any degree. For the 68 queries including geometric primitives, 38.2% used a single geometric primitive to represent an object (Figure 86a). In 61.8% of the queries two or more geometric shapes were combined in order to compose one object, such as the "stick man" in Figure 86b.

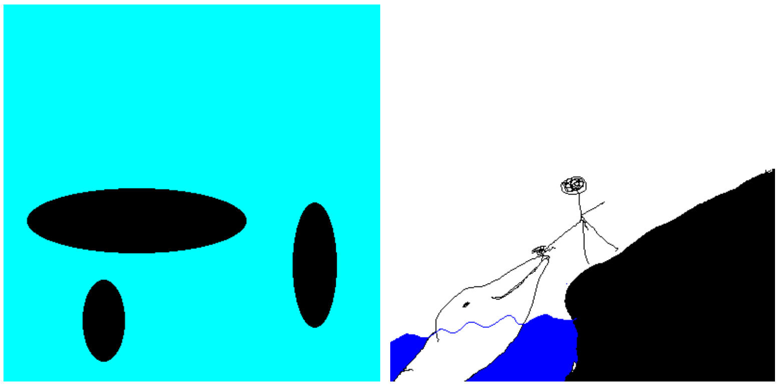


Figure 86 - Use of geometric primitives. Query #15 (Ovals representing a whale and a boat) and Query # 67, (Circle and lines combined to create a "stick figure" representation of a person).

Outlines were used in 89.6% of all queries, indicating that the respondents used *outlines* as the primary tool for representing objects in the query images. Outlines were used in number of different ways to depict an object, from very basic outlines (the shark in Figure 87a), via more detailed outlines (The shark in Figure 87b) to the dolphin's head in Figure 87c.

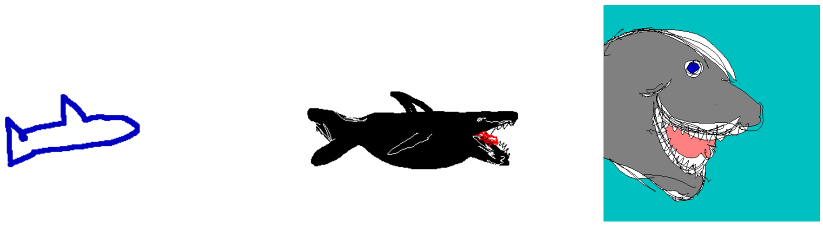


Figure 87 - Different uses of outlines. Queries # 167, 158, and 161

Visual Cues were used in 46.9 % of all queries, and in 51.1% of the queries using outlines. Figure 88A shows a shark without visual cues, Figure 88b shows a shark with such visual cues in the form of an eye, a simple mouth and a set of gills.

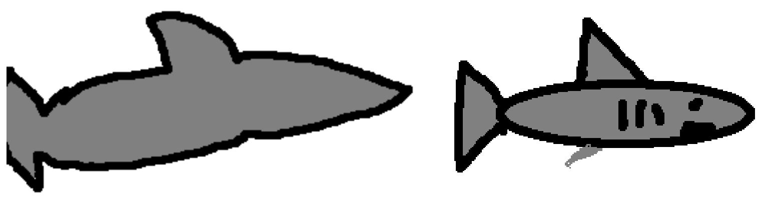


Figure 88a, b - Two depictions of a shark. Query # 46 (No visual cues) and query # 41 (Visual cues included).

Texture was only used in 16 queries (3.9%) and has not been analyzed further.

The above results indicate that outlines are preferred over geometric primitives, and that visual cues were used to some degree.

The small differences between the two respondent groups' use of *geometric primitives* and *outlines* were not significant. However, there were some indications the respondents who rated themselves as having *high* or *very high* skill in drawing used outlines in a different manner than the other respondents, while a majority of the unskilled drawers created objects with outlines similar to those seen in Figure 88, i.e. using continuous black lines to draw the outlines, sometimes filling the outline with colours and adding visual cues. The skilled drawers often drew outlines using several shorter, overlapping lines, focusing more on the real-world shape of the objects, as seen in Figure 89. These query images also tended to obtain a higher score on the subjective evaluation than other query images.

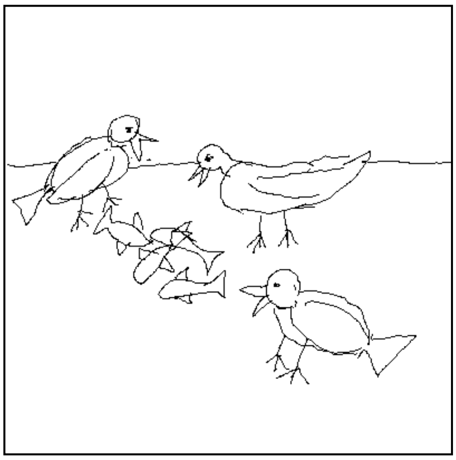


Figure 89 - An example of a using realistic outlines (Query 144, 18)

There was a significantly[89] larger use of *visual cues* in the IFIM group than in the KHIB group. As with outlines, there was a tendency that the respondents who rated themselves as skilled drawers used visual cues in a different manner than the other respondents. While the respondents who rated themselves with low drawing skills tended to use *more* visual cues, these visual cues often held a high degree of abstraction, as illustrated by query 11 (Figure 90).

[89] Mann-Whitney U [414] = -3.320, $p < 0.01$

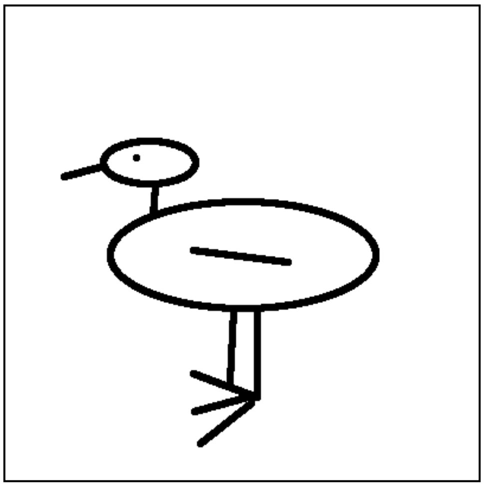


Figure 90 - Visual cues with a high level of abstraction (Query 11, respondent 2)

When the drawers who rated themselves as highly skilled included visual cues, these were often represented in a more realistic manner, e.g. they were more similar to the real-world shape of these elements, as seen in Figure 89. These queries also scored slightly higher on the subjective evaluation than the other queries.

Note that the evaluators had some difficulties describing exactly *why* they found the outlines and visual cues created by the skilled drawers more realistic than those created by the other respondents, and that they had problems pinpointing what these differences were. Accordingly, it was difficult to create any categories based on this, and to determine if there were any significant differences between the drawers. However, all the evaluators remarked this and agreed that there seemed to *be* a difference. Further studies with a larger volume of respondents and query images are required in order to evaluate this further.

Outlines were the preferred method of representing objects in both systems, but there was a significantly higher use of geometric primitives in Retrievr[90], and a significantly higher use of outlines in VISI[91], indicating that the respondents behaved somewhat differently in the two systems.

Comparing the three query types revealed some differences. First of all, *geometric primitives* were used to a very low degree in type 1 queries, but used significantly more in type 2[92] and type 3[93] queries. *Outlines* were used significantly less in type 2 queries[94] than in the other query types. *Visual*

[90] Mann-Whitney U [[414] = -5.001, $p < 0.01$
[91] Mann-Whitney U [414] = -3.286, $p < 0.01$
[92] Mann-Whitney U [310] = -4.418, $p < 0.01$
[93] Mann-Whitney U [270] = -4.102, $p < 0.01$
[94] Mann-Whitney U [310] = -3.350, $p < 0.01$

cues were used significantly more in type 1 and 2 queries than in type 3 queries, and they were used significantly more in type 2 queries than in type 1 queries.

The second measurement of representational realism is the score obtained in the **subjective evaluation of query representation**. This was measured on a scale from 1 through 5, where 1 represents the lowest possible representational modality and 5 representing the highest. 27.1% of the queries obtained a score of 1, 62.8% obtained a score of 2, while 10.1% obtained a score of 3. Table 41 shows the modality scores for the two groups, the two interfaces and the three query types.

Table 41 - Representational modality scores

Overall	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
1.83	1.80	1.85	1.94	1.65	1.83	1.86	1.79

The queries were generally rated very low, independent of respondent group, retrieval system or query type. There were no significant differences between the two respondent groups, and there were no significant differences between the query types. The difference between the two systems was small but significant: queries created in VISI were rated higher than the queries created in Retrievr[95]. While the difference was small, this difference indicated that the respondents were able to represent the query objects slightly more realistic in VISI than in Retrievr.

Based on the above, it is possible to partially evaluate research hypotheses 2.1.3, 2.2.3, 2.3.3 and 2.4.3:

RH2.1.3: The hypothesis cannot be fully accepted or rejected. While the respondents used outlines, and not geometric primitives, as the primary method of representing query participants, only 47.6% of the query images included detailed representational elements, i.e. *visual cues*.

RH2.2.3: The hypothesis must be *rejected*. There were no significant differences in the score based on subjective evaluation of representation, and the KHIB group used less visual cues than the IFIM group. However, it should be noted that there were some indications that respondents who rated themselves as *skilled drawers* created outlines and visual cues with a higher degree of visual modality than the other respondents.

RH2.3.3: The hypothesis must be *accepted*. The respondents created more realistic queries in the VISI interface. The queries expressed in VISI scored significantly higher in the subjective evaluation

[95] Mann-Whitney U[414] = -5.096, $p < 0.01$

than the queries expressed in Retrievr, and there was a much higher use of visual cues in VISI. Furthermore, there was an increase in the use of *geometric primitives* for the queries made using Retrievr.

RH2.4.3: The hypothesis must be *rejected*. There were no indications that there was a decrease in use of abstraction as the complexity of the image request increased. It should also be noted that there were some indications that the respondents used geometric primitives more when drawing query images for narrative content and scenes, indicating that the use of abstraction *increased* slightly for these image requests.

6.3.3 Discussion of the Results

The descriptive analysis above presents answers to the hypotheses, but raises a number of questions that should be examined in order to *understand* these answers:

1. Why have the respondents used a high degree of abstraction when expressing the objects in the visual query images?
2. Why was there an observed increase in the use of geometric primitives in queries for narrative content and complete scenes?
3. Why didn't the respondents use visual cues to any degree, and why did the IFIM group use visual cues more than the KHIB group?
4. Why was there a significant difference between the two interfaces?

In order to be able to answer these questions, an understanding of how people draw and develop their drawing skills is required. Edwards (1999) presents a description of how most people develop their drawing skills, and discusses the differences between people who appear to have a high drawing skill and other people.

According to Edwards, when children first begin to draw they are generally more than happy creating simple drawings, scribbles, doodles and scrawls. However, they soon learn the meaning of *symbols* and *signs*: that a drawn symbol may *represent* something in the real world, such as themselves, their dog or their parents. As they grow older, they become aware of such details such as the buttons on a shirt or the fingers on their hand, and attempt to include these details in their drawings, representing them through symbols. They *create* their own way of representing these details and *refine* them until they are very familiar with them and able to use these whenever they wish to add these details to a drawing. This is often reflected in a similarity in most drawings created by a child, even if the drawings represent very different scenes and motives. When children reach the age of 10, they generally attempt to add as many details to their drawings as they can, hoping to make these drawings as realistic as possible. But as their preference for realistic drawings is at its largest, they also become increasingly aware that the language of symbols they have developed during their early years does not include very realistic representations of real-world objects. Consequently, they experience frustration over the low level of realism in their drawings. Unless they are encouraged to continue to develop their skills they cease drawing at this point. If they try to draw again later in life, without any further training, they most often retrieve their old symbolic language, and experience the same difficulties and frustrations as they had when they were children.

One of the reasons for the use of these symbols lies in the way young children learn to interpret and categorize the world. Most children learn to understand and interpret the world in terms of *words*. Things are named, and we know the properties of these named objects. We learn to identify the items in the world around us based on these properties - the salient characteristics of the objects: “It has two eyes, four legs, a tail and a pointed nose: It’s a dog”. These properties, or characteristics, of the real-world objects are reflected in the set of symbols used when creating drawings, e.g. when drawing a face, we “know” that it consists of two eyes, a nose, a mouth and two ears. The symbols for these objects are retrieved from memory and used to create a face. However, as these are *symbols* of the real world objects, they most often have a low realistic resemblance to their real world counterparts.

The ability to *draw realistically* is closely related to the ability to *see* the world in a particular way. People who either are naturally skilled in drawing or have been taught to draw have the ability to *see past* the symbolic concepts they have learned, *see* the “real” shape of objects and translate the edges, contours and spaces formed by these shapes into a drawing. Consequently, creating a realistic drawing of a real-world object requires that a drawer can *observe* the real world objects in this manner, or has drawn a particular object so many times that they *know* how these objects are drawn. Being able to draw something from memory is a difficult task unless the object have been studied thoroughly previously, as the mental image of an object is dominated by our semantic labels and our abstractions of the properties of these objects.

Finally, the ability to *see* and *create realistic drawings* of an object is something that can be learned and mastered by most people, given knowledge about the approach and time spent practicing to transfer these concepts to a drawing.

With this in mind, it is possible to answer the questions raised by the descriptive analysis. The primary question is: why were the query objects depicted using a high degree of abstraction? The respondents seem to belong to two different categories: Those who wished to express the objects as realistically as possible, and those who wished to express the objects using a high degree of abstraction.

The respondents belonging to the former category generally stated that they wanted to create realistic depictions of the query elements, but were unable to. The primary reason stated for this was their perceived lack of drawing skills. These respondents often started to create a realistic depiction of the objects, but ended up with queries with a high degree of abstraction. This is illustrated by respondent 12. He often stated that he *wanted* to draw in a realistic manner, but when asked why he *didn't* do this, he stated that he found it too difficult:

> *[Because] it was difficult to draw. It was difficult to draw as precisely as the [mental image] I had. I imagined a fish playing with a red ball, and I have a rather good image in my mind. But then the results do not look like this at all. It's dishevelled, crooked and… It doesn't represent [the mental image] I had and wanted to use as a query. So I guess [my drawing skills] were the biggest hindrance.* ***Respondent 12*** *(Referring to query 102, represented in Figure 91):*

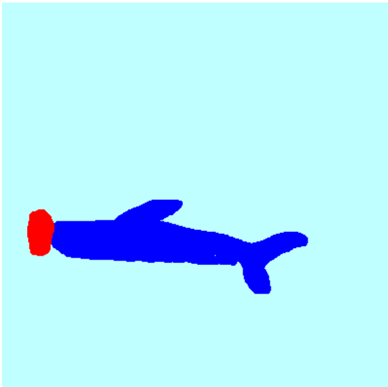


Figure 91 - Query 102, respondent 12 (Find images of a dolphin playing with a ball). The image has been resized.

Several of the respondents in this category claimed that they were disappointed or frustrated with the queries they made, and compared their query images to children's drawings. Respondent 22 presents an example of this when discussing some of his queries, illustrated by query 204, (Figure 92):

> *I think the biggest problem was my lack of drawing skills. When I make something, I want it to be perfect. And those drawings look like they could have been created by a 3 year old child. So it's not something I was very content with.* ***Respondent 22***

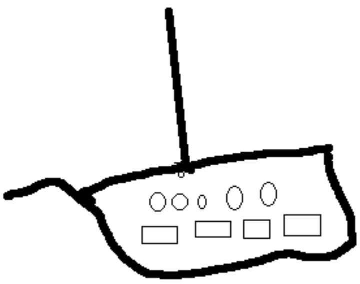


Figure 92 - Query 204, respondent 22 (Find images of a ship). The image has been cropped and resized.

Other respondents in this category stated that they ideally would like to create realistic drawings, but didn't attempt to do this, as they felt their drawing skills would not be high enough, as illustrated by respondent 16:

> *With my drawing skills, creating realistic drawings wouldn't lead to anything, I would have spent a lot of time on something that might not have worked, so the most logical thing to do was to create the queries [with a high degree of abstraction].* ***Respondent 16***

The respondents expressing these frustrations were generally in the IFIM group, and primarily those who rated themselves low with regards to drawing skills and drawing frequency. The frustrations and experiences of these respondents reflect the findings reported by Edwards. They stated that they felt their queries could have been made by children and that they ended up having a low degree of visual similarity to their real world objects. Furthermore, even though the data material is too small to provide anything but anecdotal evidence, it appears as the feeling of inadequacy becomes stronger when they draw objects they are very familiar with, such as humans and boats. These respondents appeared more satisfied with their representations of dolphins, seagulls and sharks. This is also reflected in the way *humans* are represented in the queries. In many cases, this group represented humans by combining geometric primitives into "stick figures", or as very basic outlines vaguely resembling a human, as illustrated by the queries in Figure 93. When asked about the level of detail and realism of the different elements in query 68, respondent 8 stated that she was generally very satisfied with the dolphin and the boat, but that it was too difficult to draw the humans:

> *I think I nailed the dolphin pretty well, and the boat. Those were quite [realistic]. But the humans were too difficult. The previous search, I.. I've already tried to draw the humans looking realistic, but they were just… I didn't manage it at all, so I just drew those stupid stick people. It's rather poor, isn't it?* ***Respondent 8***

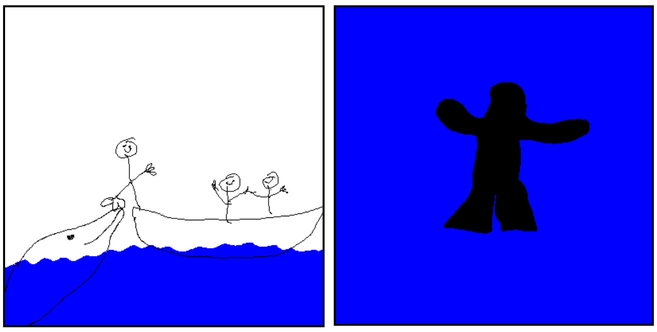


Figure 93 - Humans represented as "Straw figures" (Query 68, respondent 8) and as a simple outline (Query 3, respondent 1)

One possible explanation is that the more familiar the respondents were with the object they were drawing, the more they attempted to draw these using their existing symbolic language, and became frustrated when this resulted in “poor and stylistic” representations. These respondents usually resorted to one of two strategies: Continuing to attempt to draw realistically or resorting to pictograms. *A pictogram is a pictorial representation, an iconic sign which represents complex facts, not through words or sounds but through visual carriers of meaning (Abdullah and Hübner 2006) (Definition 15).*

The respondents who continued to attempt to create realistic drawings generally reported that they were dissatisfied or embarrassed with the resulting query images. The respondents who resorted to pictograms were generally more satisfied with the results, but they were often doubtful if the chosen strategy would work with the CBIR system:

> *The humans they were... That one on the beach [Figure 94] was too difficult to draw. I had to use a stick dude. I doubt if the system would be able to understand it (Laughs). I guess I would have wanted it to. That it could understand that those lines and circles and stuff represented a human. But I know it doesn't. So I guess it was a stupid thing to do, but... It was the only way I imagined.* ***Respondent 22***

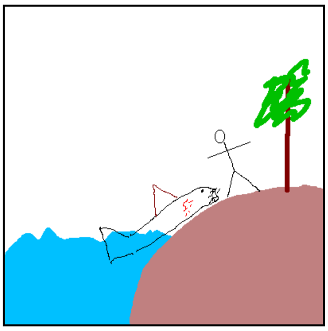


Figure 94 - A human represented as a stick figure (Query 208, respondent 22)

> *I would prefer that [the retrieval system] understood what I drew even if I drew in a stylistic manner. I'm not a good artist, and neither are most of the people I know. So, if I could draw a stylistic shape, a blue banana with some extending lines, I think that would be better than thinking about what "Flipper" actually looked like on TV and attempting to draw something similar.* ***Respondent 12***

The respondents stating that they explicitly wanted to create the queries with more abstract representations or pictograms were primarily found in the KHIB group. The primary reason stated was that they felt that abstract representations, such as *icons* or *pictograms* would be much more efficient than attempting to draw in a realistic manner. Several of them stated that, given time and possibly an example image to draw by, they would be able to draw something that at least would have a resemblance to the real-world counterparts. However, this would take too much time, and was not something they would be interested in doing when searching for images. They preferred to express the objects as simply as possible, using an approach based on pictograms:

> *Pictograms are the way to go. If you're after a chair, then you just draw, or ideally select, the pictogram for a chair. I just drew pictograms, if I wanted a person, I drew a pictogram for it, or a dolphin or a house or a bike, just the pictogram. It's fast and efficient. Easier than realistic!* ***Respondent 23***

For these respondents, creating queries with a low degree of realism represented a quick and easy way of expressing the queries, while still giving them the freedom to use the benefits of the compositional nature of the visual queries. This was probably the major reason for the increased use of geometric primitives in type 3 queries. As already noted these queries were generally more complete and contained more unique objects, in many cases several humans or animals. When creating these queries, the time required to create outlines and add a lot of detail grew with the number of objects, and a common strategy was to simplify the way the objects were represented. In most of the queries containing more than two people or a crowd of people, these were often drawn using very simple representations, as seen in Figure 95. When compared to the queries containing one human, the queries with a single object were generally more detailed (Figure 96). While there were some examples of the opposite, this generally appeared to be the case. Similarly, type 3 queries generally had a larger scale, requiring that the depicted objects were relatively smaller, presenting the respondents with less space for adding details.

Figure 95 - Queries containing several humans (Queries 383, 394, 257 and 306)

Figure 96 - Queries containing a single human (Queries 115, 408, 132 and 308)

The low use of *visual cues* is probably related to all of the issues discussed above. First of all, the overall low use of visual cues is at least partly explained by the fact that including visual cues takes time. The mean time for queries not using visual cues is 64.65 seconds, while the mean time for queries *including* visual cues is 92.1 seconds, a significant difference[96]. Some respondents also stated that they felt the extra time required to add these details would be "wasted", as they did not think that the retrieval system would benefit from them.

> *Of course, I could have spent a lot more time on [creating the queries], adding more details and so on, but the reason I drew this in a very rough manner with little details, is that I don't think it would help much, and that those details [eyes, fingers, hair] would have helped the system. Spending time creating eyes, or begun drawing detailed fins or having the mouth of the dolphin open with a lot of teeth… It would take too much time, without much benefit.* ***Respondent 6***

Similarly, the difference in use of visual cues between the different query types is probably related to the fact that queries for scenes are generally more complete, giving less space to add details, as well as increasing the time required to add details to the increased number of objects.

It should also be noted that several respondents stated that when they *included* visual cues, these were often added in order to add a *required level of detail* to the queries, e.g. identifying that a given shapes is a "shark" and not another fish by adding a set of sharp teeth and a fin (Query 64, Figure 97) or adding a snorkel and a flipper in order to indicate a "scuba diver" (Query 95, Figure 97). Several respondents stated that they felt these details were necessary, both in order for them to be satisfied with the drawing and in order for the retrieval system to correctly identify the elements ("It's not just *any* human, it's a scuba diver").

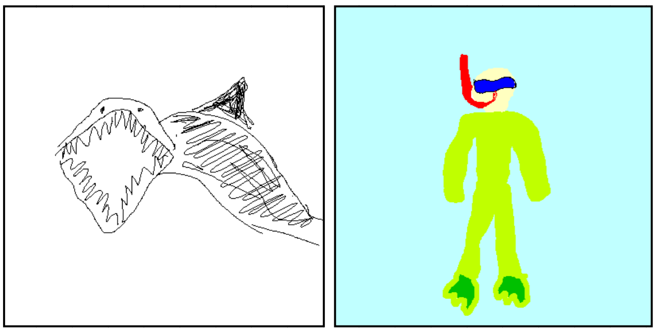


Figure 97 - Illustrations of visual cues. A shark (Query 64, respondent 8) and a scuba diver (Query 95, respondent 12)

The difference in use of visual cues between the two respondent groups was very likely influenced by the higher level of drawing skill among the respondents in the KHIB group. As most of these had at least some training and experience in drawing, they were most likely aware the relationship between

---

[96] T(377) = 7.118, p < 0.01

drawing and *seeing*. Consequently, they may be less inclined to add "symbolic" elements to their drawings, but focus more on contours and edges when drawing, as illustrated in Figure 98.

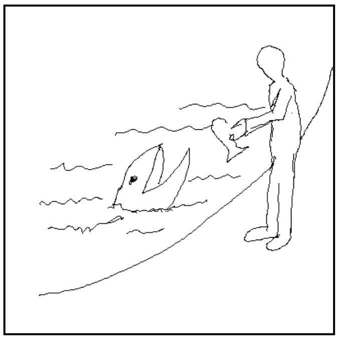


Figure 98 - A human represented using edges and spaces (Query 147, respondent 18)

Several of the respondents made claims related to this during the query process, as illustrated by respondent 29.

> *Oh, now I'm headed straight into that old trap of using symbols and stuff when drawing the people. (Laughs). Using this [The pen and tablet] in this interface is just like when I was young and used paint, and I'm kind of drawing in the same way. I can't draw humans like this, the system would never share my... stereotypes. If I'm going to.... I have to draw more like we've been taught. This... I don't think the computer will understand any of this... strange mess.* ***Respondent 29***

Concerning the difference between the two retrieval systems, a major influencing element was the different size of the canvases. The canvas available in Retrievr is much smaller than in VISI, combined with a very thick pen tool. This forced the respondents to create less detail, leading to the lower use of visual cues, and much simpler representations of the objects depicted in the query, using geometric primitives to represent animals or humans. Some illustrations of this are the stick-figures representing humans in Figure 99a and the flock of birds represented as small dots or circles in Figure 99b.

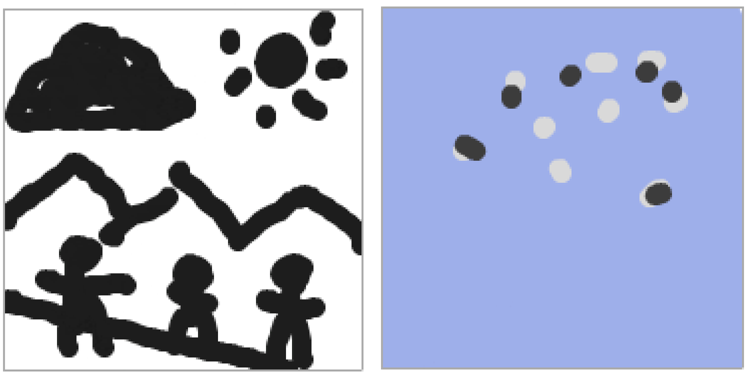


Figure 99a and b - Some representations of humans and animals in Retrievr.

Concerning the lower use of *outlines* in type 2 queries and the higher use of *visual cues* in type 2 queries, no explanations were found in the empirical data. Further studies would be required in order to determine this.

6.3.4 Summary of the Results

The respondents generally represented the query objects with a high degree of abstraction. While outlines were used as the primary method for drawing the objects, these outlines were often very simple, and the overall modality of the queries were rated low. Visual cues were used to some degree, particularly among the IFIM group. However, these visual cues were often also very abstract, i.e. held a low visual modality. Apart from the use of visual cues, there were few significant differences between the two respondent groups, but there were some indications that the respondents with a higher drawing skill created *outlines* and *visual cues* that held a slightly higher visual modality than the other respondents. The queries created in VISI held a slightly higher representational modality than the queries created in Retrievr. There overall representational modality of the queries did not change much between the three query types, but there was an increased use of *geometric primitives* in queries for *complete scenes*. Some possible explanations for these observations were identified:

- Some respondents stated that they found it difficult to create realistic representations, even though they wanted to do this. It is very likely that this is related to the way these respondents have learned to draw. They have not developed the ability to *see* the objects as they really are, and resort to using symbolic elements learned in their childhood.
- Some respondents stated that even though they "knew" the visual appearance of an object or entity, it was difficult to draw these without having an example to study. This may be related to the concept of *drawing is seeing*: Without a proper example or very much experience in drawing a particular object, it is difficult to create a realistic representation of these objects.
- Some respondents, particularly in the KHIB group, stated that while they probably were capable of creating relatively realistic representations, this would take too much time and effort, and was not something they would prefer to do in order to create an image query. Consequently, they preferred to express the query objects using sketches, abstractions or simple pictograms.

6.4 Use of Compositional Structures

Use of compositional structures describes the degree to which the respondents used a realistic composition and structure when creating the spatial layout of the query images. An image is a two-dimensional representation of a three-dimensional space. As a result, images are, in a sense, "flat". However, the human mind is capable of interpreting this 2-dimensional image and understanding elements such as the relative scale of the depicted objects, their order (from the viewpoint of the observer) and depth in an image.

6.4.1 Research Hypotheses

In order to evaluate the degree to which compositional structures have been used in the query images, additional sub-hypotheses have been created for each of the main research hypotheses:

- **RH2.1.4:** Respondents do not use compositional structures when creating query images
- **RH2.2.4:** Respondents with a visual background use more compositional structures than respondents without this background
- **RH2.3.4:** Respondents use more compositional structures in the VISI system than in the Retrievr system
- **RH2.4.4**: The compositional modality of the query images increases with the complexity of the image request

These hypotheses have primarily been evaluated using four measurements:

1. A classification of representational modality according to the *compositional marker*
2. A subjective evaluation of the use of depth, scale and perspective
3. An analysis of the placement of the query objects
4. An analysis of the sequence the query objects were drawn in

According to the modality criteria, a query image with a high degree of compositional modality is an image where compositional structures such as realistic scale, overlapping elements perspective techniques have been used to create a high degree of realism.

6.4.2 Evaluation of Hypotheses

The first measurement of compositional realism was the **compositional criteria** described in the framework in chapter 4. It should be noted that classification of the compositional modality criteria was found to be very difficult by all three evaluators. The query images generally held a very low degree of completeness and determining whether the respondents had used any compositional measures when composing and creating the queries was in many cases not possible. '

Determining whether the objects were drawn to a realistic scale or not, was in many cases difficult, and this was the criterion where the evaluators disagreed most. The queries that were classified as having a "realistic scaling" were the queries where scaling was obviously used. There may have been some queries that *should* have been included in the "realistic scaling" category. Similarly, determining if “central perspective” was used also proved difficult. During the evaluation it became clear that this criterion would be too strict, and as a consequence, it was redefined as just “perspective”, encompassing any use of perspective in the queries. This was easier to use and determine, but still represented an area where the evaluators disagreed over some of the query images. This indicates that the results obtained using these criteria should be used with some reservations. However, given the low overall completeness of the query images it is not believed that this had a major impact on the overall results.

Table 42 shows an overview of the use of the various compositional modality criteria.

Table 42 - Overview of compositional modality criteria

		Respondent group		Retrieval system		Query type		
	Overall	IFIM	KHIB	VISI	Retrievr	1	2	3
Scaling (A)	9,70 %	6,10 %	12,40 %	12,90 %	4,40 %	2.40%	8.30%	23.10%
Overlap (B)	16,40 %	14,40 %	18,00 %	23,00 %	5,70 %	7.80%	17.40%	28.80%
Perspective (C)	7,00 %	0 %	12,40 %	8,60 %	4,40 %	5.4%	4.2%	13.5%

While 96 of the query images included one or more compositional structure, none of the three structures (scaling, overlap or perspective) were used in a high degree. The only compositional structure that saw any major use was the use of overlapping objects. Of the 96 query images containing compositional structures, 51 were drawn using only overlapping elements.

Comparing the two groups reveals that the KHIB group seemed to use all three categories more than the IFIM group, but only the difference in use of *perspective* was significant[97]. Next, comparing the two retrieval systems reveal that there was a higher use of composition in the queries made in VISI than in Retrievr. The differences in *scaling* and *overlap* were significant. Finally, there was a higher use of compositional structures in queries for complete scenes than for queries for generic and narrative content. These differences were significant. There was also a significantly higher use of *overlap* in type 2 queries than in type 1 queries.

The second measurement of compositional realism was the **score obtained in the subjective evaluation** of query composition. A majority (74.64%) of the queries obtained a score of 1, 23.43% obtained a score of 2, while 1.93% scored 3. Table 43 shows the mean compositional score broken down by respondent group, retrieval system and query type.

[97] Mann-Whitney U[414] = -4.916, $p < 0.01$

Table 43 - Overall mean composition score by group, interface and query category

Overall	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
1.27	1.18	1.35	1.36	1.14	1.13	1.26	1.53

These results show that the respondents generally used very little compositional structures to add depth, scale and perspective to the queries. There were some differences between the different query types, and all the differences were significant[98]. While the overall score was low for all categories, the analysis indicates that the queries made in VISI were made using slightly more compositional structures than the queries made in Retrievr and that the respondents in the KHIB group used slightly more compositional structures than the queries made by the IFIM group. The most notable difference was observed for type 3 queries*;* these obtained a slightly higher score than the other queries.

The third evaluation measurement was an analysis of **object placement** in the query images. This was not initially planned, but given the low compositional score and the low use of scaling, overlap and perspective, it was included in order to get a better understanding of how the query images were composed. In order to evaluate this, a simple grid-based classification was used. The placement of the objects was categorized based on two axes: Horizontal and vertical placement. Where possible, the placement of the query elements was categorized as belonging to “Left, centre, right” and “Top, middle, bottom”, dividing the query image into 9 “grid squares”. The placement of the objects on the two axes was classified separately from each other. A query image object was classified on the axes categories based on where the dominant part of query contents was placed. 324 of the 414 images were classified according to this. The remaining 90 images could not be classified, primarily because of the low level of completeness in these queries. Table 44 presents the percentage of query images classified for each query category, while Table 45 presents an overview of the classification.

Table 44 - Query images classified according to object placement

Category	Respondent group		Retrieval system		Query type		
	IFIM	KHIB	VISI	Retrievr	1	2	3
Evaluated	84.5%	73.4%	77.3%	79.7%	92.2%	86.1%	45.2%

[98] All differences in Table 43 were significant at the 0.01 level in Mann-Whitney U tests.

Table 45 - Query object placement

	Left	Centre	Right	Vertical
Top	2.78%	9.88%	0.00%	12.65%
Middle	7.10%	63.27%	2.47%	72.84%
Bottom	1.85%	12.35%	0.31%	14.51%
Horizontal	11.73%	85.49%	2.78%	100%

The classification was done manually by the primary evaluator. Consequently, the results in Table 45 may be prone to the subjective nature of this classification, and one should be careful when generalizing from these results. Nevertheless, even with a large error margin, the results indicate that a majority of the queries were created with the query objects in the centre of the canvas, with a slight tendency to be in the left part of the canvas. There were no significant differences between the respondent groups. Comparing the two retrieval systems showed that the overall results were similar, but there was a slightly higher number of queries created in the middle and lower part of the canvas in Retrievr. These overall results are similar for type 1 and type 2 queries. For type 3 queries, the middle category was still the largest (with 42.6% of the classified queries). However, the remaining queries had the query objects evenly distributed across the 9 categories, indicating that the respondents used placement more actively in these queries.

The final measurement was an analysis of the **sequence the query objects were drawn** by the respondents. As with the object placement measurement, this was not initially planned, but included in an attempt to gain more information about how the respondents composed the query images. This evaluation was also done by the primary evaluator alone. *Query Sequence* represents *the order* the respondents drew the objects in the query images. This evaluation was based on an analysis of the query interface videos, complemented by the interview sessions. The analysis was based on the 200 images that included *background*, *contextual elements* and combination of these.

It was difficult to find any clear tendencies from this material. The only general tendency that was identified was that in 67.1% of the 73 queries including detailed backgrounds (e.g. using more than just a single colour for background, as seen in Figure 100), the objects of interest were drawn first, followed by the background and other elements. This was particularly true for queries for *complete scenes*, where 88.89% of the 36 queries including a complex background were drawn in this order. It should be noted that in some cases, the respondents started by filling the canvas with the main background colour (e.g. “blue” for the ocean), then drew the objects of interest, and finally detailed the background.

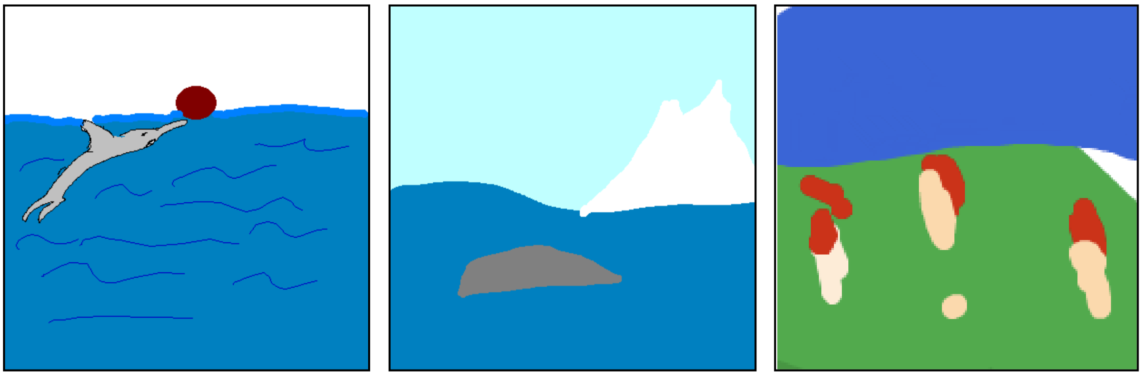


Figure 100 - Queries with detailed background (Queries 78, 237 and 305)

During this analysis, another observation was made. The respondents tended to *complete* each object before moving on to the next object, e.g. when drawing the dolphin in query 78 (Figure 100a), the respondent *completed* the dolphin before drawing the other elements. This was true for 74.23% of the 214 queries containing more than 1 object.

Based on the above, it is possible to evaluate research hypotheses 2.1.4, 2.2.4, 2.3.4 and 2.4.4:

RH2.1.4: The hypothesis must be *accepted*. The respondents *did not* use realistic scaling or overlapping elements and *did not* make use of one or more perspective techniques when creating visual queries. However, it should be noted that there seems to be an *increase* in the use of these when expressing queries based on requests for *complete scenes*.

RH2.2.4: The hypothesis must be *accepted*. The respondents from the KHIB group *used* more compositional structures than the IFIM group. However, it should be noted that this was a small difference, and the only significant difference was in their use of *perspective*.

RH2.3.4: The hypothesis must be *accepted*. The respondents used significantly more compositional structures in the VISI system than the Retrievr system, particularly for *scale* and *overlap*. However, it should be noted that while there was an increased use, compositional structures were still not used very much in the VISI interface.

RH2.4.4: The hypothesis must be *accepted*. There was a significant increase in the compositional modality of the query images as the complexity of the image requests increased.

6.4.3 Discussion of the Results

Given the challenges related to using the compositional modality criterion on the query images, the evaluation of these hypotheses should not be given very much weight. However, eight issues with the query image composition were found:

1. The respondents used few compositional structures when creating the query images

2. The KHIB group used slightly more compositional structures than the IFIM group
3. There was a tendency to use more compositional structures in VISI than in Retrievr
4. There was a difference between the query types. Compositional structures were considerably more used in queries based on requests for complete scenes
5. The respondents generally drew the queries in the middle of the canvas
6. There appeared to be a tendency that the
7. When creating *complete* query images, there was a tendency among the respondents to *first* fill the background, draw the important objects and finally add contextual details
8. The respondents tended to *finish* each object before moving on to the next

The use of compositional elements must be seen in relationship with the discussion of completeness in chapter 6.1. For many queries, and particularly queries for *generic objects*, the respondents were often not interested in the background, setting or compositional structures of the images they were requesting, and in some cases they were even determined *not* to specify these elements. For example, if a respondent was interested in generic images of dolphins, including compositional structures in the query would not be necessary or even desired. Given the large number of queries including only participants, including compositional structures other than perspective might not even be possible. The increased use of compositional measures in type 3 queries strengthens this explanation. In many of these queries, the respondents were actually interested in the compositional structures of the image, which may have caused them to include these elements.

Regarding the use of *scale*, there is some evidence that the respondents used this to determine the *importance* or *weight o*f the different query objects, rather than aiming for a realistic relationship between the objects depicted, as illustrated by the two queries in Figure 101. When discussing query 130 in the interview session, respondent 16 stated that he deliberately oversized the dolphin compared to the boat and the people, in order to give it more weight:

> *[The larger size] was to give it more weight in the image, as a relevant object. It was like, you know, movie posters, where the main character is in front while the other characters are in the background, much smaller, even if the relative sizes are all wrong, or they would have to be much farther back than they are, in order to give the main character more salience. In this case, that it's a dolphin in Norway.*
> ***Respondent 16***

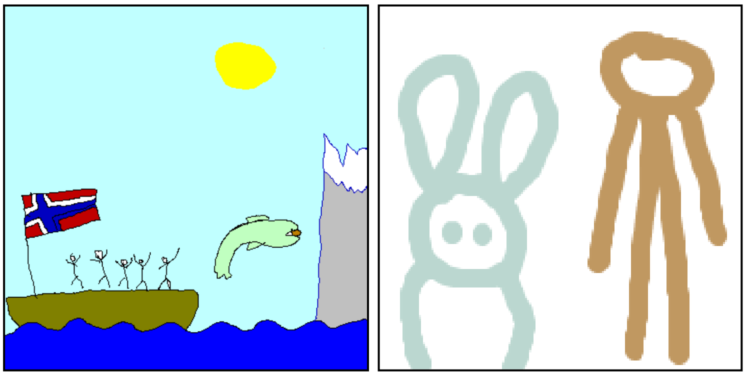


Figure 101 - Two queries illustrating the use of *value scaling.* (Query 130, respondent 16 and query 361, respondent 29)

When creating query 361, respondent 29 stated that ideally, the rabbit and the hunter should be placed much further from each other, and that the relative size of the rabbit according to the hunter was way off, but that she felt that she had to represent both participants as large as possible in order to “tell the system” that they were equally important.

Several other respondents made similar claims: They increased the size and detail of the main object in order to emphasize its relative importance in the query. Rather than depicting the query objects in a realistic scale, the visual objects are scaled based on their relative *perceived importance*, e.g. a *value scaling* was used to determine the relative size of the objects.

Concerning the use of perspective, there was little material describing this in the interview data. However, two different reasons are indicated by the respondents. First of all, some of the respondents in the IFIM group stated that they found this very difficult. During the query sessions, at least two respondents explicitly indicated that they would like to include perspective in the queries, but found it too difficult:

> *Now, I’d really like to add some kind of... What is it called? Perspective? Central perspective? I think that real images have these kinds of lines running... Or that it should be possible to indicate that some these objects are in the foreground and those in the background, but I’m really not sure how... It’s too difficult, I’ll just pop these in here and hope for the best.* ***Respondent 5, while creating query 45 (Figure 102****)*

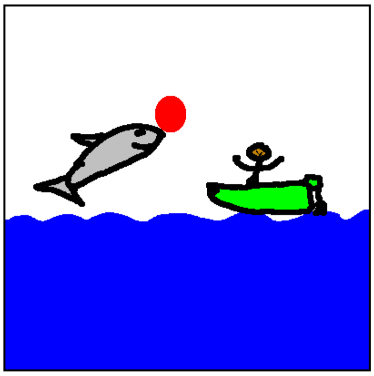


Figure 102 - Query 45, respondent 5

Next, some of the respondents from the KHIB group stated that they explicitly wanted to avoid adding perspective to the queries. They felt that it would take too much time, and that they believed that the algorithms would be confused by such structures:

> *Adding details such as horizon lines, vanishing perspectives and similar things... That would take too much time and... I'm thinking it would just confuse the system. Make the results worse. Keep it simple.* ***Respondent 18***

Finally, when discussing perspective with the KHIB respondents who included some degree of perspective in their queries, they generally stated that this was not something they did on purpose or explicitly tried to do, but that it felt "natural" for them, and that it was the way they normally drew:

> *I don't think I… No, I didn't actively use any perspective techniques, but I guess it just came natural, particularly in that query with the beached whale [Figure 103], I just drew the people in front larger. I was kind satisfied with that one, actually, because of the perspective. But no, I don't think that I made a conscious choice, using perspective.* ***Respondent 31***

Figure 103 - An example of a query image created using perspective (Query 393, respondent 31).

Some of these respondents also stated that the size of the canvas and the pen tools in Retrievr made use of scaling, overlap and perspective difficult. This was most likely the main cause of the differences between the two interfaces; no other explanations for this were found in the material.

Concerning the actual placement of the query objects, a majority of the query images had the main objects either centred in the image or slightly below the middle of the image. This was particularly the case in type 1 and 2 queries.

Some respondents stated that they started to draw in the middle of the canvas in order to make sure they had enough space to add all the details they wanted:

> *It is because I draw very simple. I start in the middle, and think "What am I going to get now"? I have to make sure I have enough space on all sides, and I usually tend to draw off in one of the directions.* ***Respondent 5***

Other respondents reported that they chose the middle of the canvas as they either did not have a clear opinion of where the objects should be placed, or found it difficult to imagine where the objects would be located in a photograph. Placing the objects in the middle of the canvas represented a simple way out of this problem.

Some of the respondents reported that they made a conscious choice to draw the object s in the middle of the canvas as they felt it was natural, and that they felt that the important objects should be given the most prominent position of the canvas, i.e. "the middle":

> *That I'm drawing in the middle is... Usually, the important things are in the middle of the image. When I tried to draw the dolphin and the diver, I placed the dolphin, then I placed the diver a little to the side, and that's probably what a photographer also would have done. That is, from what I've seen on BBC and the Animal Planet, then it's like the shark in the middle and the diver next to it. It's that way of thinking - it's where the interesting things are.* ***Respondent 2***

Other respondents stated that they, either consciously or subconsciously tried to follow conventions from photography, possibly explaining the tendency to use the lower part of the canvas more than the upper part:

> *I placed a lot of the images, not necessarily in the middle [..], but tried to place them two thirds towards the lower part. It would be like the "Rule of thirds" in photography. I imagined that is where a photographer would have placed them, so that it hits with the "golden mean". That's where I'd like my object, because I imagined that's the way a photographer would have made the image.* ***Participant 12***

Generally, the respondents reported that they did not give much consideration to where they placed the objects. In most of these cases, they were only interested in retrieving images with the particular content, and did not have any strong opinions on *where* in the resulting images these objects were placed. This was particularly true for type 1 and 2 queries. This was directly related to the way they used colours and contextual elements: They did not want to be too specific when creating the queries in order to avoid excluding content. The respondents did not put much weight into the actual placement of the objects, and did not want the retrieval system to focus on it either, e.g. they wanted to retrieve *all* images containing the depicted object(s).

> *[Drawing the objects at a particular position of the canvas] doesn't really mean much. Fair enough, it is possible that you would like to have images with objects in that particular position, but it might not necessarily be the best way, or... Your placement might not be the best, and there may be a lot of images with better placements, that you could not have imagined [..] I'm thinking that it's more a kind of recognition process. Ok, here I've drawn a happy face, and now I want all images containing a happy face. Independent of where the happy face is located in the image.* ***Respondent 24***

Query 113 by respondent 14 (Figure 104) is an extreme representation of this type of object placement. The query image represents a request for images containing a dolphin entertaining one or more happy people in a boat, by playing with a ball:

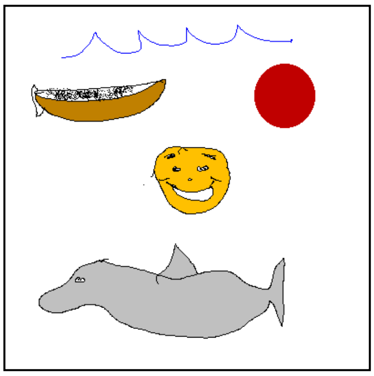


Figure 104 - Query 113: Several objects of interest.

The participant stated that she didn't think that the system would actually be able to retrieve images based on this, as it was unlikely that there were any spatially resembling the structure created in the query. But she felt that this query was the best way she could manage to express that she wanted one or more of these objects in the image, without having to make up her mind about the actual spatial placements of these objects.

When creating type 3 queries, or when the respondents had a very clear mental image of the composition they wanted in the query image, they were more conscious about the spatial arrangement of the query objects, as illustrated by the 3 queries in Figure 105, all made by respondent 3:

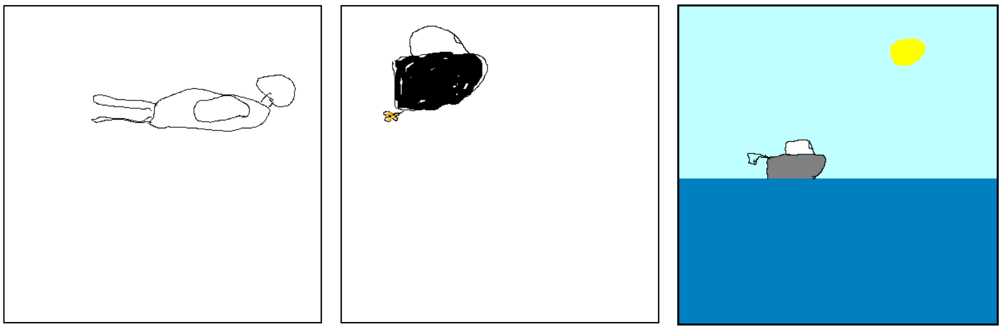


Figure 105 - Different ways of composing query images (Queries 18, 23 and 24)

The two first queries were requests for generic content (“Find images of a scuba diver” and “Find images of a boat”), the final query was a request for a complete scene (Based on newspaper article scenario). Respondent 3 stated that in the case of the two first queries, he had no particular opinion about where the objects should be placed in the image. The last query was more complex, and he tried focus more on the actual composition and placement of the objects:

> *No, I didn’t have a conscious notion about it, I just drew it somewhere it could fit. I didn’t think that the things I was trying to retrieve would be placed in the lower right corner or anything like that. I really didn’t think about it. But then, it might be a little different depending on what you’re trying to retrieve. If you’re just going to search for a dolphin, I don’t consider where the dolphin is, but in the end, when I was trying to retrieve a boat in the sea, with the sun and things, then things were a little different. Naturally, I placed the boat on top of the surface.* ***Respondent 3***

Concerning the sequence the objects were drawn, little material was found in the interview material. Generally, the respondents reported that they did not have a particular strategy when creating the queries. Some of the respondents in the KHIB group stated that they normally preferred starting with the background in order to provide a context and composition to the image, but that given that they were *searching* for images rather than *composing* or *creating* images, they did not focus much on this when creating the query images. Some respondents stated that they drew the background first, particularly in Retrievr, as the size of the canvas forced them to do this: If they started with the query participants, they had no easy method to add background without compromising the already-drawn objects. Similarly, little evidence was found for the tendency to complete each object before moving on to the next object. For queries containing complex background (e.g. backgrounds consisting of more than one colour), the respondents had a tendency towards completing the background first, then adding the other elements. A likely explanation for this was that it may have been difficult to add background *after* drawing the query participants and contextual elements, particularly in Retrievr. Further investigation is required in order to determine if these results are generalizable.

6.4.4 Summary of the Results

The respondents used few compositional structures, and only overlapping objects were used in any degree. The query images were also given a low subjective score. There were some differences between the two groups; the KHIB group created query images that held a slightly higher compositional modality, and used the compositional structures slightly more. However, the general compositional modality was so low for both groups that the observed differences were minor, even if they were significant. The query images created in VISI held a higher compositional modality than the images created in Retrievr. Finally, type 3 query images held a significantly higher compositional modality than the other images. Some explanations for this were identified:

- Adding compositional structures may be difficult. Several respondents in the IFIM group stated that they would like to include such structures, but had no idea how they should approach this.
- Adding compositional structures may take too much time. Several respondents in the KHIB group stated that while they were able to create these, they did not want to spend time creating this, particularly when they were just searching for generic content.
- Generally, there were a low number of objects in the query images. Consequently, there were few objects which the respondents could work with.
- *Value scaling* was in some cases used as a method for highlighting important elements in the image. Several respondents stated that they felt this was more important than representing the objects in a realistic scale.
- The small the canvas combined with a large pen size made it difficult to add very much compositional structures in the Retrievr interface, compared to the VISI interface.
- The respondents often did not put much weight on the placement of the query objects, particularly in queries for *generic* and *narrative* content. They wanted to retrieve images *containing* the depicted objects, without making any statements concerning *where* the objects were located in the retrieved images
- Several respondents stated that they believed that placing an object in the centre of the query gave it more weight than placing it in other parts of the query
- There were some tendencies among the respondents to draw important objects first, and that they tended to *finish* each object before starting to draw the next object.

6.5 Summary: Query Image Modality

Based on the above sections, it is possible to present some general perspectives on the modality of the query images and answer the three research hypotheses discussed in this chapter. The most immediate observation is that the query images generally held a very low visual modality. Table 46 presents an overview of the subjective evaluation of the four modality markers for the different query categories, along with an average of the modality score for each category. Note that the "Mean" row is not the "mean of the means", but the overall mean of all query modalities for all queries in each query category. The overall average modality score given to the queries were 1.64, on a scale from 1 to 5, indicating that the overall naturalistic modality of the query images was very low.

Table 46 - Overview of evaluation of modality markers

Marker	Overall	Respondent group		Retrieval system		Query type		
		IFIM	KHIB	VISI	Retrievr	1	2	3
Contextualization	1.60	1.56	1.63	2.14	2.02	1.23	1.87	3.79
Colour	1.85	1.81	1.89	1.85	1.85	1.65	1.87	2.15
Representation	1.83	1.80	1.85	1.94	1.65	1.83	1.86	1.79
Composition	1.27	1.18	1.35	1.36	1.14	1.13	1.26	1.53
Mean	1.64	1.56	1.68	1.72	1.49	1.49	1.62	1.89

Figure 106 presents the frequency distribution of the mean modality score. Note that while this figure (mean overall modality) may represent a useful indicator of the overall query modality, a high degree of caution should be exerted when using it, as it is represents a mean of *the mean* of the four subjective evaluation scores, for all queries. However, comparing this figure with the results obtained by the various evaluation tools used in the previous sections shows that it reflected these results.

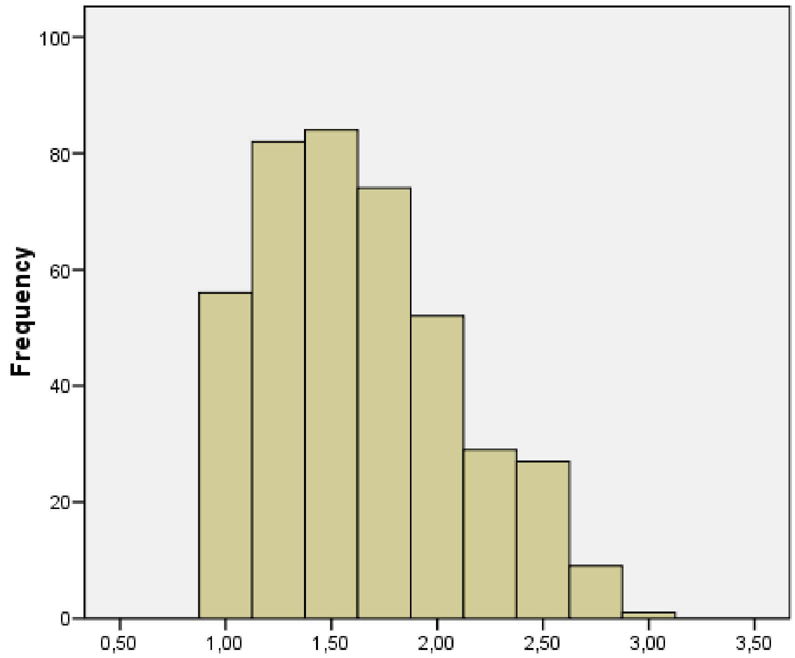


Figure 106 - Frequency distribution of query modality mean

Based on this and the discussions in the previous sections, it is possible to answer the three main research hypotheses. Each hypothesis is answered and discussed in the following sections.

6.5.1 Research Hypothesis 2.1: Overall Query Image Modality

Research hypothesis 2.1 stated that the respondents would create queries with a *low visual modality*. The previous sections showed that this was true:

> **RH2.1:** The hypothesis *must be accepted*. The respondents created queries with a low degree of visual modality.

4 major reasons for the low query modality were identified:

1. Keeping the queries simple using a minimum level of detail
2. Difficulties related to the drawing process
3. A desire to avoid "confusing the system"
4. A desire to secure relevant query results

Each of these is summarized in the following sections.

6.5.1.1 Keeping the Query Images Simple

First, almost all respondents stated that they wanted to keep the queries as simple as possible. When looking for images, they wanted to spend as little time and effort as possible. They did not wish to spend time and effort creating very good, detailed and realistic images, but wished to focus on specifying the *important elements* of their image request. For example, when requesting images of a dolphin, they did not want to spent a lot of time adding contextual details, colours, adding very specific details to the dolphin, or to spend time creating good composition or use a realistic scale in the query image. They wanted to focus on the task at hand: How do I create something that is *similar enough* to the objects in my request? This is also reflected in the time spent on the different queries, as seen in Table 47. With the exception of queries with a mean modality score of "2.75", there was an increase in the mean time spent for each increase in mean score. The correlation between mean modality score and mean time was significant[99].

Table 47 - Mean time spent on queries, categorized by mean modality score.

Mean modality score	1.00	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
N	56	82	84	74	52	29	27	9	1
Mean time	34.41	65.18	84.44	121.78	133.53	168.46	170.58	107.38	200

While few of the respondents stated this explicitly, these results indicated that there was a tendency among the respondents to treat the visual queries as what can be called *visual keywords*, i.e. rather than expressing the query using a single textual keyword, they expressed the queries by creating a very simple representation of the object they were attempting to retrieve. For example, when expressing queries with a relatively low detail, e.g. "Find images of a dolphin", "Find images of a seagull" or "Find images of an interior object" the queries were often very simple, containing a basic representation of a dolphin, a seagull or an interior object, as represented in Figure 107.

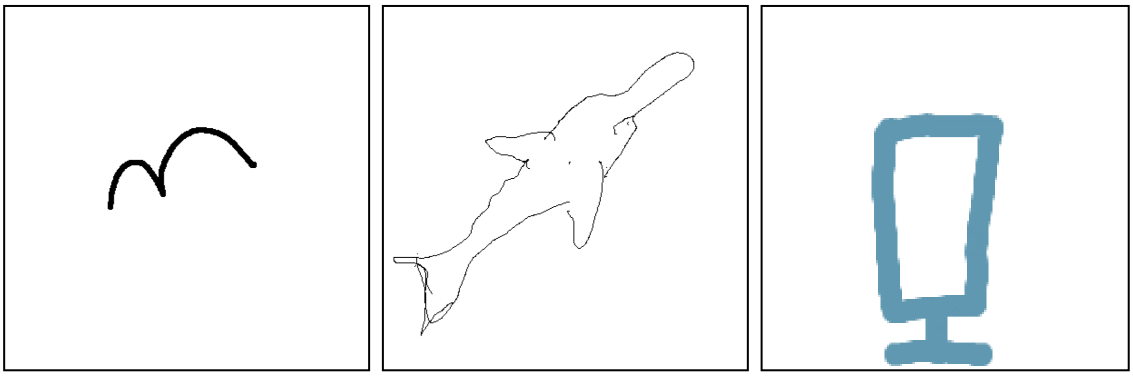


Figure 107 - Illustrations of "Visual Keywords" (Queries 1, 38 and 349)

The desire to keep things simple and only add the required level of detail may also help explain the observed increased modality in type 3 queries. The generic requests and some of the requests for narrative content used in this project were generally very *low in detail*, e.g. "Find images of a

[99] Kendall's Tau-B r[379] = 0.478, $p < 0.01$

dolphin", "Find images of an interior object" or "Find images of a happy girl". In these cases, the respondents were free to determine the level of detail in the queries, and consequently tended to use the "visual keywords" approach. However, in the case of more detailed requests, such as "Find images of several people and / or animals gathered in a rural setting" or "Find images of humans practicing sports", these were more defined, and "demanded" that the respondents included more details in the queries in order to fulfil the query specifications. For example, a common strategy used to indicate "forest" was to include large green areas in the query image, as illustrated by the three queries in Figure 108.

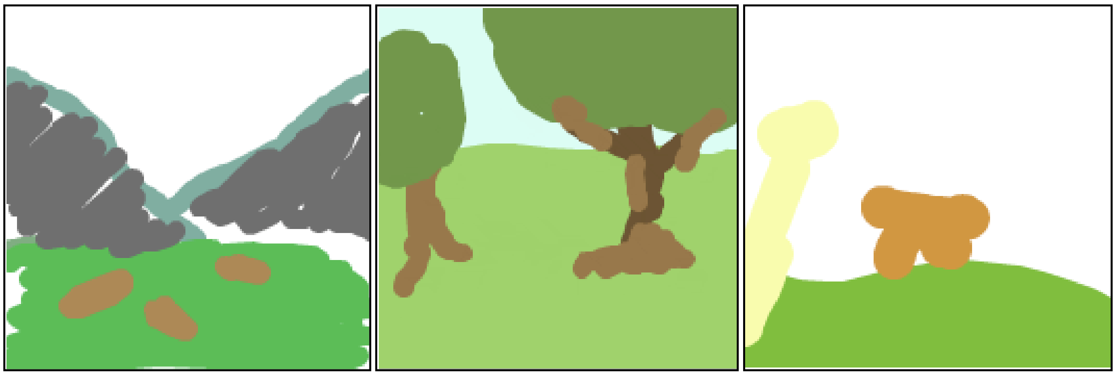


Figure 108 - Requests for "Humans and / or animals gathered in a forest" (Query 360, 401 and 185).

Similarly, when attempting to find images based on the scenario based tasks (i.e. the newspaper article about the dolphin visiting Norway), the queries often included mountains, the sun, a boat and a dolphin, as illustrated in Figure 109. While the modality of these queries varied between the respondents, the modalities of these queries were generally higher than other queries, particularly with regard to *completeness*. On the other hand, the *representational* and sometimes *colour* modality often *became lower* as the completeness of the queries increased, primarily due to the increased time and effort required to create these additional details.

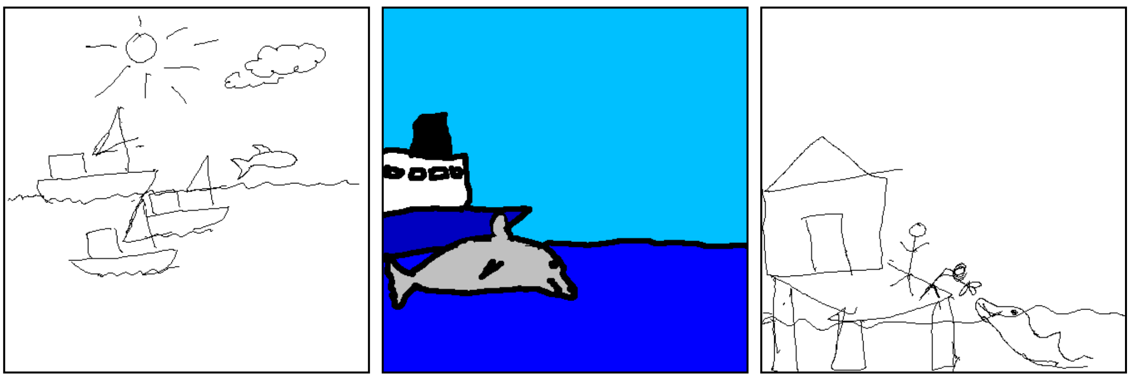


Figure 109 - Queries based on the "flipper" scenario (Queries 140, 120 and 69).

6.5.1.2 Difficulties Related to the Drawing Problems

Second, a large portion of the respondents experienced several problems related to the actual drawing process. These challenges are discussed in further detail in chapter 7, but summarized the respondents found it difficult to *draw the different objects*, they experienced problems *initiating* the drawing, i.e. determining *what to draw* (e.g. context and background) and they experienced difficulties in creating a *realistic composition*.

6.5.1.3 Avoid Confusing he System - Assumed Behaviour of CBIR Systems

Third, several respondents stated that they did not want to *confuse the system*. They felt that by including details such as contextual elements, background and colour, the system might focus on the wrong elements, e.g. retrieve any images with a blue background rather than retrieving images containing a dolphin, or retrieve images of trees when they were looking for images of forest animals. Related to this, some respondents chose to *highlight* important elements of the query image by using particular colours they felt carried certain meanings (e.g. “red” indicating “blood” or “injury”) or exaggerating the relative size and scale of the important objects related to less important objects, e.g. contextual elements or background.

6.5.1.4 Seeking Inclusivity in the Query Results

Fourth, several respondents stated that they explicitly kept the contextual elements and colours low in order to not *exclude relevant images* from the result set, e.g. when querying for *interior objects* they did not want to exclude objects based on their colour, or when querying for *seagulls* they did not want to exclude images based on contextual details.

6.5.2 Research Hypothesis 2.2: Differences between Respondent Groups

Research hypothesis 2.2 stated that there would be a difference between the two respondent groups. However, while there were some differences that suggested that the KHIB group created queries with a slightly higher modality than the IFIM group, these differences were very small and most of them were not significant. Accordingly, the hypothesis must be rejected:

> **RH2.2:** The hypothesis must be *rejected*. There were no major significant differences in the visual modality of the query images drawn by the two respondent groups.

Despite this, there were some observations that should be discussed. First of all, it can be argued that the classification used (KHIB vs. IFIM) might not have been ideal. While the respondents from the KHIB group all had some degree formal education related to visual arts, not all of these respondents rated themselves as very skilled drawers. Similarly, while the respondents in the first experiment were asked to rate themselves with regards to drawing skill, they were not asked whether they had attended drawing classes or had any formal or other informal education related to visual arts and drawing. Consequently, while the two groups represent populations with different *formal educational background*, they might not represent populations with *different drawing skills*.

In order to examine if there were differences between people with varying drawing skills, the respondents were classified according to how they rated their drawing skills in question 7 in the first questionnaire, separating the respondents into two categories: Those who rated themselves with "Very low", "low" or "medium" drawing skills (1, 2 and 3), and those who rated themselves as "high" and "very high" (4 and 5). The 4 respondents who rated themselves as "medium" were included in the first category in an attempt to ensure that the respondents in the second category were different from the respondents in the first category. Table 48 shows how the respondents are classified according to their drawing skills.

Table 48 - Respondent classified by drawing skill

Category	Total	IFIM	KHIB
High drawing skill	10	1	9
Medium and low drawing skill	20	16	4

Using this classification of the respondents rather than IFIM vs. KHIB indicated that the respondents who rated themselves as having a *high* or *very high* drawing skill created queries with a significantly higher modality than the other respondents, for all the measurements of query modality, as illustrated by the mean score obtained on the subjective evaluation of query modality (Table 49).

Table 49 - Mean score obtained on the subjective evaluation, according to drawing skill.

Modality marker	Overall	Low and medium skill	High skill
Colour	1.85	1.79	1.93
Context	1.60	1.53	1.68
Representation	1.83	1.68	2.0
Composition	1.27	1.17	1.38
Mean	1.64	1.54	1.75

All differences observed in Table 49 except colour[100] were significant at the 0.01 level, indicating that the respondents who rated themselves with a high drawing skill created queries with a higher visual

[100] Mann-Whitney U [414] = -1.804, p < 0.01

modality than the other respondents, particularly for *representation* and *composition*. These results are reflected for all the modality evaluators described in the previous sections. Similarly, these respondents were also able to create the queries *faster* than the respondents with a low drawing skill, with a mean time of 62.01 seconds vs. 95.51 seconds, a significant difference[101]. This indicates that the users with higher drawing skills are able to create queries with a slightly higher visual modality, and are able to create these faster. However, while the differences in visual modality between the two groups were significant, they were very small, and it is doubtful that the differences have any impact on the overall quality of the query images.

Additionally, there were also some indications that the drawers who rated themselves with *high* or *very high* drawing skill created *outlines* and *visual cues* that held a slightly higher representational modality than the other respondents.

Finally, it should be noted that there were some indications that the respondents in the KHIB group were better able to utilize the canvas and drawing tools available in VISI than the respondents in the IFIM group. This indicated that, given suitable tools, respondents with a visual background are able to create query images with a higher modality than respondents without this background. Comparing the respondents based on their self-rated drawing skills presents even stronger indications towards this. Table 50 presents the score of the subjective evaluation of the query images created in VISI, according to respondent type:

Table 50 - Differences between the respondent groups' modality scores in VISI

Modality marker	Overall	IFIM	KHIB
Colour	1.85	1.79	1.94
Context	1.76	1.63	1.94
Representation	1.94	1.86	2.06
Composition	1.36	1.20	1.56
Mean	1.64	1.54	1.75

All differences were significant at the 0.01 level except from colour use. While caution should be taken with these results based on the small empirical material, it may be an indication that the size of the canvas and the level of the detail provided by the pen sizes may influence the respondents'

[101] $t[377] = 4.564$, $p < 0.01$

ability to express queries with a high modality, and that this is particularly true for respondents with a visual background.

6.5.3 Research Hypothesis 2.3: Differences between Retrieval Systems

Research hypothesis 2.3 stated that the queries made in VISI would have a higher visual modality than the queries created in Retrievr. The evaluations in the previous sections revealed this to be true:

> **RH2.3:** The hypothesis must be *accepted*. The queries created in VISI held a higher visual modality than the queries created in Retrievr. This was true for all modality markers. However, it should be noted that while there *was* a significant difference between the query images, all images generally held a very low modality.

First of all, one should not put too much confidence in these results, as the structure of the evaluation may have introduced some uncertainties in the data material. 80% of the queries created in Retrievr were created by the KHIB group; the 3 IFIM respondents who used Retrievr only created 33 queries. Consequently, it is possible that differences between the interfaces may be influenced by the differences between the two respondent groups. Additionally, different query tasks were used in the two interfaces. It is possible that the different query tasks used may have caused changes in behaviour, which might falsely be identified as differences caused by the interface. However, based on the material from the interviews it was found very likely that most of the differences observed between the queries made in the two interfaces could be attributed to two factors: The size of the canvas and the dynamic result presentation provided by Retrievr.

As already noted, the respondents were generally much more satisfied with the canvas and pen sizes available in VISI than in Retrievr. This difference forced the respondents to use different drawing strategies in the two interfaces. For the queries created in VISI, the respondents could add all the details they wanted. The small size of the pen allowed them to add minute details if they wanted, or they could use drawing tools in order quickly draw major compositional elements such as background. In Retrievr, they were forced to focus on the general composition of the image, draw very few objects or use very abstract representations of the objects, as illustrated by the three queries in Figure 110.

Figure 110 - Level of detail in the queries. Queries 144 (VISI), 366 (Retrievr) and 281 (Retrievr)

In addition, there were some indications that the dynamic result presentation offered by Retriever system caused the respondents to add *less* detail than in VISI. As noted, the respondents generally wanted to keep the queries as simple as possible. In VISI, they had to rely on their "gut feeling" in order to decide when enough details were added, while in Retrievr they could keep adding details until they either were satisfied with the results, or decided that they had to use another approach for the query. Consequently, these queries often had fewer details than the queries made in VISI.

6.5.4 Research Hypothesis 2.4: Differences between Query Categories

Research hypothesis 2.4 stated that the visual modality of the query images would increase as the complexity of the image requests increased. The previous sections showed that this was *partly true*:

> **RH 2.4:** The hypothesis can be *partially accepted*. There were significant differences in the visual modality of the query images created for the different image requests. This was particularly true when querying for *complete scenes*. There was generally an *increase* in *completeness, use of colours* and *use of compositional structures* as the query complexity increased, but there was a *decrease* in the *representational modality*.

The decrease in *representational modality* was most likely related to the increase *level of detail* in these queries. As the complexity of the queries increased, the number of elements included in the query images increased proportionally. This had two implications. First, more time and effort was required by the respondents in order to add details to each object, and the respondents generally wanted to keep the queries as simple as possible. Next, the increased number objects combined with a limited canvas size led to a decrease in size of each object, giving less space for details for each object. Combined, this led to an increased level of abstraction in the objects in the query image.

The *increase* in the other modality markers may be related to the actual nature of the retrieval task. The *generic* retrieval tasks were simple in nature, e.g. "Find images of a dolphin". When faced with such tasks, the respondents often did not see the need to include additional details, or did not *want* to add additional details in order to avoid *confusing* the system or to ensure *relevancy* of the results. In many cases, the respondents treated the visual queries as *visual keywords*, e.g. rather than typing “Dolphin”, they drew a simple representation of a dolphin. As the complexity of the queries increased, particularly with the requests for *scenes*, the respondents had to include additional details in order to be able to fully articulate the query images according to the retrieval task.

7 Query by Drawing: Major Challenges

The third objective in this study was to identify the major challenges the users face when they draw visual query images, as expressed in research question 3:

What are the major challenges encountered when users draw visual queries?

An important element here is that the focus is on the challenges *experienced by the users* when drawing the query images. The scenarios in section 1.1 (page 3) introduced three potential challenges, which are addressed in the following research hypotheses:

- **RH3.1**: Lack of drawing skills is a major challenge when respondents draw visual query images
- **RH3.2:** Drawing visual queries is too time-consuming to be an efficient tool for image retrieval
- **RH3.3:** Lack of usable interface tools is a major challenge when drawing visual query images

In addition to these hypotheses, two additional major challenges were identified in the previous sections: *Expressing narrative content* and *problems related to initiating the drawing process*, i.e. the "Page Zero" problem. The three research hypotheses and these two additional challenges are examined and discussed in detail in the following sections.

7.1 Challenges Related to the Users' Drawing Skills

Challenges related to the users drawing skills refer to potential challenges that arise because the user lacks training and / or experience in drawing, as expressed in research hypothesis 3.1:

- **RH3.1:** Lack of drawing skills is a major challenge when respondents draw visual query images.

This hypothesis was evaluated using two methods:

- The respondents' answers to questions in questionnaire II
- An analysis of the interviews and the query sessions

The first method was an evaluation of **the respondents' answers in questionnaire II**. Two questions were directly relevant for this challenge:

- Q2: How easy was it to express a search using Visual Queries? (Table 51)
- Q15: To what degree did you feel that your own drawing skills influenced your ability to create good queries? (Table 52)

Table 51 -Questionnaire II, Q2: Ease of using QBD

Answer	Overall (N=30)	IFIM (N=17)	KHIB (N=13)
Very difficult	6	4	2
Difficult	16	10	6
Neither	7	2	5
Easy	1	1	0
Very easy	0	0	0

Table 52 - Questionnaire II, Q15: Influence of drawing skills

Answer	Overall (N=16)	IFIM (N=3)	KHIB (N=13)
Very low influence	4	0	4
Low influence	3	0	3
Neither	2	0	2
Much influence	3	2	1
Very much influence	4	1	3

Note that since question 15 was not asked in the first experiment, 13 of the 16 answers to this question were from respondents in the KHIB group, and one should be careful to generalize from this material alone. However, based on the structured interview sessions, the overall impressions was that this was true for the 14 IFIM respondents taking part in the first experiment.

These answers seem to indicate two things: The respondents found it difficult to express the requests through visual queries, and the respondents were divided on the relationship between their own drawing skills and their ability to express visual queries.

The results were reflected in **the analysis of the interviews and the query sessions**. 14 of the 17 respondents in the IFIM group stated that their own drawing skills represented the largest challenge when expressing the visual queries. They reported two problems related to this: They did not possess enough *basic drawing skills*, and they experienced problems *translating their mental image into a drawing*:

> *It is that I simply can't manage to draw a dolphin. Not even something resembling a dolphin.*
> ***Respondent 1***

> *It was difficult to draw because of a lack of drawing skills.* ***Respondent 11***

> *The most difficult thing was to create a precise drawing of the image I'm visualizing. I visualize a fish playing with a red ball, and have a rather clear image in my mind. And then the results don't look very much like this. It is dishevelled, stylistic and slanted.* ***Respondent*** *12*

> *My biggest challenge is that I have this inner image, and recreating this visually is difficult when you're not a very skilled at drawing.* ***Respondent*** *15*

While the respondents in the KHIB group often reported problems related to drawing the queries, they generally felt that this had very little to do with their own drawing skills, as illustrated by respondent 25:

> *I don't think [my own drawing skills] had any influence whatsoever. I don't think these drawings reflect that I'm actually quite good at drawing.* ***Respondent 25***

These respondents often stated that the problems they experienced were not related to their *drawing skills*, but were more often related to their ability to *visualize* the objects they were attempting to draw:

> *No, I don't think my drawing skills influenced this much. Drawing is very much... It's seeing.. Drawing skills don't mean very much, I guess. It's more related to the ability to visualize things that is important. If you're going to draw dolphins, you either need a lot of experience in drawing them, and I guess that would count as some kind of drawing skills... Or you would need to be able to see an example of a dolphin, and draw according to that. If you haven't got much practice in drawing dolphins, or don't have a dolphin to look at, drawing a good dolphin will be very difficult.* ***Respondent 27***

Two considerations should be taken when discussing these statements. First, it is quite possible that the respondents in the KHIB group underestimated the benefit they have from their background and their training, particularly compared to people who do not have similar background. Next, it appears as if the two groups refer to different concepts when they talk about "drawing skills". Most of the respondents in the IFIM group describe "drawing skills" as a single skill which they do not possess, while the KHIB respondents considers "drawing skills" as a specific skill belonging to a larger set of competencies.

When the respondents in the IFIM group discussed drawing skills, they claimed that if they had more experience in the "skill of drawing", they would be able to create more realistic representations in their query images. They said that they often had a clear mental image of the objects they were going to draw, but were unable to translate this into a drawing due to their lack of drawing skills.

When the respondents in the KHIB group discussed drawing skills, they talked about at least two different concepts. Most of them had substantial drawing experience, and some of them reported that they had formal education fields related to drawing. Despite this, most of the KHIB respondents acknowledged that they were not able to, or willing to, create realistic representation in the query images. Some of them even reported experiences similar to the respondents in the IFIM group, e.g. they realized that their images resembled children's drawings and were embarrassed over their drawings. But most of these respondents explicitly referred to the theory of drawing reported by Edwards (1999): In order to create realistic object representations, they either have to draw things they have prior experience drawing, or have to spend a substantial amount of time studying the object. The domain of the retrieval tasks in this study was unfamiliar for most of the respondents; drawing sharks, dolphins, seagulls or happy girls was not something they did on a daily basis. Consequently, according to their own statements, some of the respondents resorted to their prior symbolic representations of these objects, either subconsciously or by active choice. This resulted in representations that often had a very "childish" appearance. Other respondents explicitly chose to express these unfamiliar objects in an iconic manner or by using various pictograms.

In addition to this, most of these respondents stated that they considered their ability to create "realistic depictions" as an ability or skill independent on the actual physical act of drawing. Most of the respondents in the KHIB group reported that they were fairly skilled in the physical act of drawing, and that actually drawing the objects they had decided to draw did not represent any major challenges for them.

This is reflected in the results reported in the previous chapters. The analysis in chapter 6 showed that there was little actual difference between the visual modality of the query images created by the two groups. Accordingly, there was little evidence that the respondents in the KHIB group were able to, or willing to, create more realistic representations than the respondents in the IFIM group. This was also more or less true for the respondents who rated themselves with high or very high drawing skill, even though these respondents created slightly more realistic representations. However, they were able to create the query images significantly *faster* than the other respondents, indicating that the main benefit from their visual background was in their ability to express the queries in an *efficient* manner.

Despite this, one should not underestimate the fact that the respondents in the IFIM group *experienced* that their lack of drawing skills was the most significant challenge when drawing the query images. If these respondents are to take advantage of the potential benefits offered by the

QBD CBIR approach, effort should be made to reduce the challenge, and encourage the users to express the queries in a way that can provide them with meaningful results.

In summary, the respondents in the IFIM group reported that they experienced their lack of drawing skills as a one of the major challenges facing visual query specification by drawing. The KHIB group generally reported that they did not feel that their drawing skills influenced their ability to create good visual queries.

Based on the above discussion research hypothesis 3.1 can be evaluated:

> **RH3.1:** The hypothesis is *accepted* for the respondents in the IFIM group, while it is *rejected* for the respondents in the KHIB group. The respondents in the IFIM group found their perceived lack of drawing skills to be a major challenge when drawing visual query images. It should be noted that it is likely that this is *experienced* as a larger challenge than in actually is. Most of the respondents in the KHIB group *did not* consider a lack of drawing skills as a major challenge when drawing visual query images.

7.2 A Time Consuming Process

A time consuming process refers to the time required to express image requests through drawing visual queries, particularly compared to the time it takes to express these requests through text. This was expressed in research hypothesis 3.2:

- **RH3.2:** Drawing visual queries is too time-consuming to be an efficient tool for image retrieval

This research hypothesis was evaluated using three methods:

- The respondents' answers to questions in questionnaire II
- An analysis of the time spent drawing the query images
- An analysis of the interviews and the query sessions

The first method was an evaluation of **the respondents' answers in questionnaire II.** Three questions were related to time:

- Q12: Terms selected to describe visual queries
- Q13: How time consuming did you experience this form for image search? (Table 53)
- Q14: How problematic did you find the time required by this form of image search? (Table 54)

In question 12, 18 respondents selected the term “Time consuming”. 13 of the 17 respondents in the IFIM group selected the term, while 5 of the 13 respondents in the KHIB group selected this term. This indicates that the respondents in the IFIM group found the process more time consuming than the respondents in the KHIB group.

Table 53 shows the answers to question 13, while Table 54 shows the answers to question 14. These answers also indicate that the respondents found the process time consuming, but that the respondents did not necessarily find the time required to be a major problem.

Table 53 - Questionnaire II Q13: How time consuming is QBD?

Answer	Overall (N=16)	IFIM (N=3)	KHIB (N=13)
Very time consuming	0	0	0
Time consuming	10	2	8
Average	5	1	4
Little time consuming	0	0	0
Very little time consuming	1	0	1

Table 54 - Questionnaire II Q14: How problematic is the time required by QBD?

Answer	Overall (N=16)	IFIM (N=3)	KHIB (N=13)
Very problematic	1	1	0
Problematic	3	0	3
Average	5	1	4
Little problematic	6	1	5
Very little problematic	1	0	1

The second evaluation method was **analysing the time spent to create the queries**. The overall mean time required to create the queries was 99.18 seconds, with a mean of 138.12 in the IFIM group and 72.21 in the KHIB group (Discussed in chapter 5.2, page 120). With the mean time just above one and a half minute, creating visual queries by drawing takes considerably more time than expressing the requests through text based queries, at least for generic requests (e.g. “Find images of a dolphin”). The major question is: do the respondents consider this a major problem?

All respondents agreed that drawing the visual query images took considerable time, particularly compared to expressing generic text based queries. However, several respondents stated that they would probably be able to draw considerably faster if they were more experienced using QBD CBIR systems. Some stated that the pen and tablet were unfamiliar tools, and that they felt it very likely that they would be able to utilize these better with some practice. Other respondents claimed that more experience with QBD CBIR systems would allow them to create the queries, particularly with regards to drawing the images in a way that the system could interpret. This claim is supported by the significant difference in time spent in the two interfaces: Several respondents stated that they were better able to understand how Retrievr worked, and were able to adapt their queries to this, spending less overall time on the query process.

While most respondents agreed that it *was* time consuming, there were differing views of the how *problematic* they found this. Eight of the respondents in the IFIM group reported that they had *fun* creating the queries, and would not mind spending time if they were using a QBD CBIR system for leisure or entertainment. However, they were not convinced that they would be willing to spend this time if they were in a professional situation or required a high degree of efficiency when retrieving images:

> *It was fun. It was innovative, but time consuming, took at lot of time. If you are going to use it as a toy, then it is fine. You don't care much about how long it takes. But if you are short on time, or you are looking for valuable information, then I don't think this is as good as it should be.* ***Respondent 5***

> *It was fun doing this as an experiment, but if I was at work, and needed to get images quickly, I don't think I would think that [Query by drawing] was very cool. While it's fun to see how the algorithms work and compare the different systems, it takes too much time. It's not efficient!* ***Respondent 22***

Most of the other respondents in the IFIM group and all the respondents in the KHIB group were more positive towards the question of time. These respondents stated that they did not mind the time it took to create the queries, and reported several reasons for this.

First, they stated that while drawing queries took considerably longer time than creating textual queries this might not be true for *all* types of image requests. The respondents mentioned several types of retrieval tasks where text based queries would be difficult to articulate and that their requests could be expressed more efficiently using QBD CBIR:

> *I found [drawing visual queries] more difficult and more time consuming than text. But... As I mentioned, there are a lot of situations where text simply isn't good enough, or... There are some things that are so difficult to express using text, things that I can express very easily when drawing. In*

> *these cases, I don't think I would mind spending all that extra time, or rather, I would spend less time.* ***Respondent 20***

> *If you're looking for some types of flowers, and you're not interested in them as a botanic, but more in the general visual impression of the flower, then this would be considerably faster and more efficient than [using text]. You won't have to spend a lot of time in Google just finding the name of the [..] flower, but just draw the colours and... For these [image requests], this is going to be so much faster!* ***Respondent 16***

Next, several of these respondents also stated that they didn't mind spending the extra time specifying the queries, if this could improve the level of precision in the query results. Several respondents reported that they often spend little time expressing a textual query, but spend quite a lot of time browsing through the query results:

> *I actually think it is much better to spend 5-6 minutes on obtaining concisely the things you wish to have in the image, than the frustration of not hitting the correct terms and searching through several pages. I would rather spend more time on the query. Definitely. But I think I would have started using text. And if this had seemed like a chore, I would have drawn.* ***Respondent 14***

> *[The extra time required] might be part of the drawback, but I think that if I get a lot of relevant hits, I'd rather spend time drawing than eventually searching through a lot of images later, if I had searched using text. So in that regard I don't think I actually spent so much time on the search. But of course, if you see it in relation with typing a single word, it took some time.* ***Respondent 6***

> *Actually, it didn't take that much time, if you compare it to browsing through 30 pages of results on Google Images just in order to find a suitable image. That takes a lot longer time. While the groundwork [Drawing the query] takes much longer, but you may improve the results, if you are successful with the groundwork.* ***Respondent 17***

The difference in opinion between the two groups may be related to whether the respondents are able to identify areas where they might benefit from using visual queries. Most respondents in both groups were able to identify situations where they might use QBD CBIR systems. However, only some of the respondents in the IFIM group were able to identify real-life situations they could *personally* gain benefit from using QBD CBIR. On the other hand, all the respondents in the KHIB group could identify such situations. Those respondents who *could* identify such situations were the most positive towards the QBD CBIR approach, while the other respondents were far less positive towards the time required.

Summarized, the respondents agreed that drawing visual query images was a time consuming process, particularly compared to expressing image requests through text. Some respondents stated that while they found the approach entertaining, they would not be willing to spend the required time if they were in a real-life situation. However, a majority of the respondents, including all the respondents in the KHIB group, reported that time was not a major issue, particularly if visual queries could reduce the time required to browse through the query results.

Based on this, research hypothesis 3.2 can be evaluated:

> **RH3.2:** The hypothesis is *rejected*. Most of the respondents did not find the time required to express visual queries to be a major challenge. However, it should be noted that a minority of the respondents stated that they were most likely not willing to spend this much time creating queries. The respondents who request images regularly on a professional basis were positive towards spending time using QBD CBIR systems.

7.3 Lack of Usable Tools

Lack of usable tools refers to the tools available current QBD CBIR interfaces. As suggested in chapter 1, these tools might not be sufficient for drawing visual queries. This potential challenge has been addressed in hypothesis 3.3:

- **RH3.3:** Lack of usable interface tools is a major challenge when drawing visual query images

This research hypothesis was evaluated using three methods:

- The respondents' answers to questions in questionnaire II
- An analysis of the tools used by the respondents in the query sessions
- An analysis of the interviews and the query sessions

The analysis and evaluation of the use of tools was presented in section 5.1.1 (page 95). Based on that evaluation, it appears as if the selection of tools available in the two interfaces were sufficient: The respondents preferred to draw using freehand drawing. The respondents in the KHIB group stated that the tools did not influence their ability to draw to a high degree. This is reflected in the answers in question 16 in questionnaire II ("To what degree did you feel that the tools available in the interface influenced your ability to create good queries?"), shown in Table 55. Note that this question was only asked in the 2nd and 3rd experiments. The answers seem to indicate that the respondents in the KHIB group were less influenced by the choice of tools than the respondents in the IFIM group. While only two of the IFIM respondents actually answered this question, these attitudes were reflected in the structured interviews.

Table 55 - Questionnaire II, Q16: Tool selection and drawing ability

Answer	Overall (N=15) [102]	IFIM (N=2)	KHIB (N=13)
Very low degree	2	0	2
Low degree	6	0	6
Some degree	3	0	3
High degree	3	1	2
Very high degree	1	1	0

A majority of the respondents reported that they would have enjoyed *additional* tools, particularly more domain specific tools. This is discussed in detail in chapter 9. However, lacking such tools did not seem to present the respondents with major challenges when drawing. Accordingly, research hypothesis 3.3 can be evaluated:

> **RH3.3:** The hypothesis must be *rejected*. The tools available in the two interfaces did not present a major challenge when drawing visual query images.

7.4 Expressing Narrative Content

During the interview sessions, almost all respondents reported that one of the most challenging elements of the retrieval tasks was to express *narrative content*, e.g. *actions, interactions* and *conditions*.

An image always represents a *snapshot of time*, e.g. while it may depict different types of narrative content, this content is always represented as a moment frozen in time. While humans are capable of understanding the narrative content of an image by interpreting the structural characteristics of the image (Kress and van Leeuwen 2006; Hove 2007), being able to *express* this when *drawing* seems to be considerably more difficult. Most of the respondents reported that they struggled to express different types of *actions*. Respondent 12 discusses problems related to drawing an *attacking shark* or a *seagull eating:*

> *I visualized an image, and even if an action is depicted, it is a moment of time which has been stopped. So how does the shark look while it is attacking? It is likely to have its mouth wide open, and there is probably something edible nearby, such as a dolphin or a scuba diver, or something else. I tried to stop the action where I felt it was descriptive according to the text.* ***Respondent 12***

[102] One of the 3 respondents in the IFIM group did not answer this question.

> *What does a feeding bird actually look like? It's not like it's a process with a knife and a fork, it's a rather quick affair… It's difficult. You might say that an image expresses an action, but it is difficult to draw it the other way. "Hey! There's a lot going on in this image" one might say about an image or an illustration, but it is difficult to do it the other way around.* ***Respondent 12***

A related problem is illustrated by the following statement from respondent 16, where he discusses query 130 (Figure 111):

> *One of the more difficult things was to make sure that the system didn't interpret this as a "a shark attacking" the humans. I figured that the best way of expressing that the dolphin was happy was by having it jump in front of the boat, with a ball in the mouth. But when I started to draw it, I didn't want to create a "Jaws" moment by having the dolphin jump towards the people. So I drew it jumping away from the boat. It probably didn't have any effects, but... I didn't want to see images of attacking sharks when I made that [query].* ***Respondent 16***

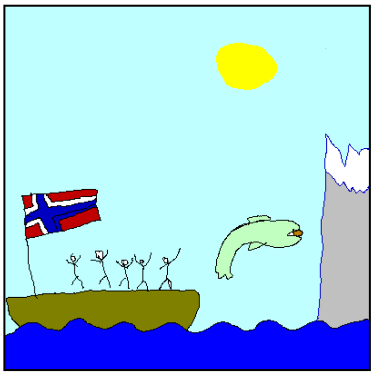


Figure 111 - A request for images of a dolphin entertaining people (Query 130, respondent 16)

Similar problems were reported by the respondents when they were attempting to express *conditions.* As described in chapter 6.3, the respondents often struggled to represent query participants in a realistic manner. Despite this, most respondents managed to create some sort of representation of the image participants. However, describing the *conditions* or *states* of these participants was in some cases very difficult. One example is respondent 4. She reported that she was generally comfortable when expressing the "basic concepts" (e.g. the query participants), expressing the state of these objects, e.g. illustrating that a dolphin was *hurt*, was considerably more difficult:

> *It [the dolphin] is going to have an injury on its side, how do I draw that? There are some, these basic concepts… I more or less know how a dolphin looks like, I know what a seagull looks like, but I don't know what a seagull looks like when run over, or a seagull feeding, because it just isn't in my head.* ***Respondent 4***

The respondents used different strategies when solving these problems. When expressing actions such as *movements* or *interactions*, some of the respondents included, or wanted to include, abstract representation of movement. Respondent 10 draws parallels to cartoons when describing how he wanted to express that a child is petting or playing with a dolphin:

> *It wasn't easy for me to indicate that the dolphin should [..] box children in the snout (sic). [..] I placed the child near the snout to try to illustrate this. But in cartoons [..] one often uses lines, smoke clouds or similar to express movement. But this wouldn't be included in a real image, and it wouldn't… It would not matter if I drew such lines, because it would not have been any on the image I was trying to retrieve. You can't express actions - in cartoons you use aids which are not really there.* ***Respondent 10***

A similar example is respondent 24s query for images of people practicing sports ("A man running"), illustrated in Figure 112.

> *I felt that I someway had to describe that the person was moving, and I felt [adding speed stripes] was relevant. But I know that real images aren't like that. If it was a real image of a running man, I guess the person would have been in focus, but with a blurred background. But it's just a kind of symbolic language you have become used to from cartoons. And I guess [query 242] is something I might have drawn if I was playing Fantasy, just to illustrate that he, the person, was moving forward.* ***Respondent 24***

Figure 112 - Query 242, a request for people practicing sports, illustrated using movement lines.

There were only a few examples of query images that included such “cartoonish” elements, but several respondents considered using these in lack of other way of expressing movement. Respondent 31 made one attempt at representing “realistic movement” in a query for a jumping dolphin. He added several long, vertical stripes indicating the flow of water, as seen in query 397 (Figure 113). This was the only image where an attempt to create “realistic” movement was included.

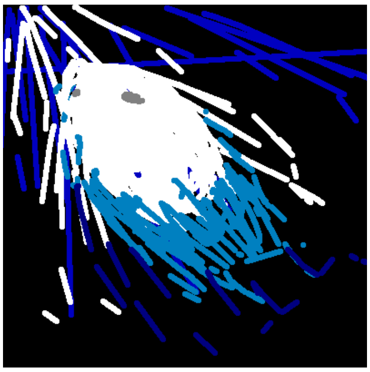


Figure 113 - A dolphin jumping out of the water (Query 397, respondent 31)

The most common strategy for expressing *interactions* was simply to juxtaposition the interacting objects in the query image, as illustrated by the two queries in Figure 114. The first image represents a request for images of people helping a beached whale. When discussing this image, respondent 31 stated that he was very uncertain about how he should indicate that the people were attempting to *help* the whale. He claimed that his query image might just as well represent four whalers butchering a whale, but could not find a viable strategy for differentiating between the two different situations. In the second example, respondent 29 was attempting to find images of a hunter hunting a rabbit. She was not very satisfied with the query image, as they would never be that close in a realistic image, but she felt that the only way she could indicate that the hunter was hunting the rabbit was to place them side by side.

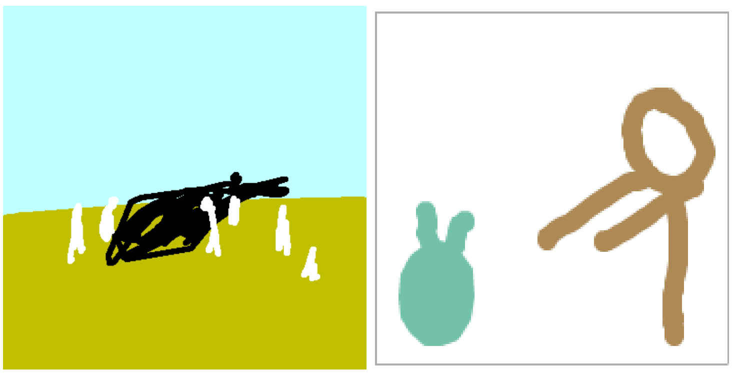


Figure 114a and b - Interaction is indicated by juxtaposition of objects (Queries 392 and 362)

Finally, as noted in chapter 6.2, *colours* were in some cases used to indicate conditions such as *injured* or *happy,* actions such as *attacking,* or more general concepts such as *violence*. One example of this is query 80 (Figure 115), where respondent 9 stated that he wanted to indicate injuries using “red”, even though he was looking for dolphins with old injuries, and that it was unlikely that these would be “red” in a real image.

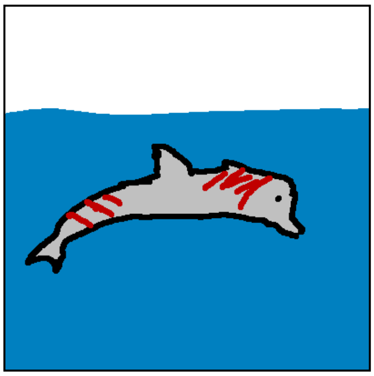


Figure 115 - A request for images of "an injured dolphin"

Summarized, most of the respondents reported that they often found expressing narrative content difficult. This was particularly the case when attempting to express *actions, interactions and conditions*. Two different aspects to this were identified: The respondents reported problems determining *how* they should express this type of content. Some respondents stated that while they felt they were able to indicate that there were actions taking place in their query images, they found it difficult to indicate *the type* of actions performed, e.g. differentiating between people *petting a dolphin* and people *hurting a dolphin*. Some respondents stated that they *would like to* include some sort of *iconic elements* to the queries in order to indicate various types of *narrative contents* and a few respondents *included* such elements in their queries. Finally, a majority of the respondents simply felt that *interactions* could be best indicated by a juxtaposition of two or more elements in the query image.

7.5 The Page Zero Problem

The final challenge identified in the material was that several respondents had some problems *initiating* the drawing process. While only 1 respondent explicitly reported these problems, several other respondents indirectly expressed this during the query sessions. Accordingly, the interface videos were the main source for this challenge.

Several respondents had some difficulties when faced with a new retrieval task, e.g. when they first tried to retrieve images of *a jumping dolphin* or *people practicing sports*. They often hesitated when starting the query process, restarted the drawing process or made several attempts at creating the query. Two main issues related to this were identified: Difficulties to determine *how* they should draw the objects they desired, and difficulties related to how the images should be *composed*.

The problem of determining what the object should look like was probably the most major problem. As mentioned in chapter 6.3, most respondents stated that they had a mental image of the general visual characteristics of most objects, but found it very difficult to determine how they should translate this into a query. This is very likely related to the prior discussions on *drawing skills* and *realistic drawings*: Creating a realistic drawing of an object is difficult unless the drawer has considerable experience in drawing that object from memory, or has access to some visual representation of the object.

The problem of determining the composition of the image did not appear to be as severe as the previous problem. This problem was primarily experienced by few the respondents who had prior experience with CBIR technology, and were aware that queries consisting of black lines on a white canvas might not return very precise results. The respondents who treated the query images more like *visual keywords* often did not include background, contextual elements, nor did they consider the composition of the query very much.

The "page-zero" problems were primarily experienced when the respondents *first started* to draw query images for a particular topic. After they had done an initial query and got some results, they were less hesitant when drawing similar or related images. Some respondents explained this by stating that they felt more comfortable expressing the queries when they had seen images similar to the images they were retrieving. The example images might not necessarily be relevant to the query process, but just by looking at real images when creating the query allowed the respondents to overcome some of the problem of determining what the image should look like. It should also be noted that this problem was primarily experienced when expressing queries in VISI. Once the respondents were familiar with the Retrievr system, they generally solved this by just starting to draw something, modified their results or started a new query based on the results they obtained.

Similarly, the respondents seemed to experience the "page-zero" problem less when attempting to retrieve images of complete scenes, both when they determined these scenes on their own, or when trying to retrieve images based on any of the scenarios. This indicates that the "page-zero" problem is primarily related to queries for *generic content*, e.g. when the respondents are interested in any type of image depicting some sort of object or content, and the level of detail in their request is low.

The "page-zero" problem may not be very large or relevant in all cases, but the frustrations and challenges experienced by the respondents appeared substantial, and the problem should not be ignored. However, as the results indicate, the problem may easily be counteracted by providing a small set of example images directly in the query interface, or using a dynamic presentation of the query results.

In summary, several respondents seemed to experience problems when initiating a query process for objects they were unfamiliar with, particularly when they had generic requests, e.g. requests where the context of the request is less important than the main image contents. These findings related to the *page zero* problem reflect some of the issues discussed in (Lai, McDonald et al. 1999; Lee, Jeong et al. 2004).

7.6 **Summary: The Major Challenges**

Based on the above discussion and the evaluations in the previous chapters, the main challenges the respondents faced when expressing visual queries by drawing can be summarized as:

- Problems related to creating realistic representations of the query image participants. The respondents in the IFIM group attributed this primarily to a perceived lack of drawing skills. The respondents in the KHIB group generally did not report this as a major challenge, but acknowledged that their representations may not have been very realistic.
- Problems relating to expressing *narrative content*. A majority of the respondents stated that they experienced problems when attempting to describe *actions, interactions* and *conditions*. They experienced problems deciding *how* to include narrative structures in the queries, as well as problems related to *describing what the narrative structures indicated*, e.g. differencing between different kinds of interactions or conditions.
- Some respondents reported problems deciding how to start composing a query, e.g. when starting a query with a blank canvas. This was primarily the case when expressing queries based on *generic* requests, and when the image requests were of a *low level of detail*. The respondents experienced fewer problems when querying objects they were *familiar with*, when they had seen images *similar* to their image requests, or when the image requests had a *high level of detail*.
- Some respondents, particularly the respondents not working with images and image requests on a daily or on a professional basis, stated that the time required to create the queries might be a very significant challenge towards expressing visual queries through drawing.
- Most of the respondents stated that they preferred drawing using freehand, and that they did not feel that a lack of tools was a major challenge in the QBD CBIR process. However, they indicated that they *enjoyed having access* to drawing tools, particularly some basic

geometric shapes and supporting tools such as a tool for filling the canvas. They also indicated that they might find expressing the queries somewhat easier if they had access to more domain specific tools, and had access to drawing tools that allowed them to express their queries on at their desired level of detail, e.g. having access to a small pen tool and a relatively large canvas.

8 Do They Like It? User Opinions and Attitudes

The fourth study objective in this work was to determine how the respondents felt about expressing image retrieval tasks through drawing visual queries, and if they were willing to use this in a realistic setting. A prerequisite for any query method is that the users feel comfortable using it. CBIR literature suggests that users might not want to use query by drawing as a tool for image retrieval because of challenges related to the query formulation process. However, little actual empirical data has been collected in order to determine this claim. This chapter presents an evaluation of how the respondents felt about using this QBD CBIR as a means for expressing visual queries. Attempts are made to determine if there is any validity to the claim that users are not willing to use QBD for image retrieval tasks, as expressed in **research question 4:**

> How do users feel about expressing image requests by drawing visual queries?

Three hypotheses were suggested as answers to this question:

- **RH4.1:** Respondents do not like to express image retrieval tasks by drawing visual query images
- **RH4.2:** Respondents with a 'visual background' are more positive towards expressing image retrieval tasks by drawing visual query images than respondents without this background
- **RH4.3:** Respondents do not prefer to use drawn visual queries over text based image queries

The main approach used to answer this research question was an analysis of the data collected in the interview sessions, supported by the respondents' answers to questionnaire II. In order to evaluate the hypotheses, three issues were explored:

1. Do the respondents *enjoy* using QBD?
2. Are the respondents *willing* to use QBD?
3. What *types of uses* do the respondents see for QBD?

The first issue was used to evaluate how the respondents felt about drawing visual query images. If the respondents *disliked* expressing image requests in this manner, it is unlikely that they would use this approach in a real-world setting. This is described in section 8.1.

The second issue was used to evaluate if the respondents *were willing* to express image requests by drawing visual queries. Having a positive experience with a retrieval method does not necessarily mean that the method will actually be used. This is described in section 8.2

The third issue was included to identify *what types* of retrieval tasks the respondents could imagine themselves performing by drawing visual queries, given that they were willing to do this. This is described in section 8.3.

Section 8.4 summarises the discussion and presents answers to the three research hypotheses.

8.1 General Attitudes towards Drawing Visual Queries

Five questions from the 2nd questionnaire were directly related to the respondents' general attitudes to the QBD CBIR process:

- Question 1: How well did you enjoy searching for images using Visual Queries?
- Question 2: How easy did you find it to express a search using Visual Queries?
- Question 10: If a system such as this was publically available, is it likely that you would use it over a text based search?
- Question 11: If a system such as this was publically available, is it likely that you would use it in *addition to* a text based search?
- Question 12: In the table below, please mark those words you feel best describe image retrieval using visual queries.

Table 56 presents descriptive data from questions 1, 2, 10 and 11 (Questionnaire II). "1" represents the lowest possible score (Not well / very difficult/ very unlikely), while "5" represents the highest possible score (Very well / very easy / very likely).

Table 56 - Answers from questionnaire II for questions 1, 2, 10 and 11.

Question	Minimum	Maximum	Mean
1	1	4	3.0
2	1	4	2.1
10	1	5	2.5
11	1	5	3.77

There were some differences between the two groups, as illustrated in Table 57. The respondents in the KHIB group seemed both more positive towards QBD, found it easier to use, and were more willing to use QBD than the IFIM group. However, none of these differences were significant[103].

Table 57 - Differences between respondent groups for questions 1, 2, 10 and 11

	IFIM	KHIB

[103] *Mann-Whitney U [30]*, $p > 0.01$ for all questions

Question	Min	Max	Mean	Min	Max	Mean
1	1	4	2,94	2	4	3,08
2	1	4	2,00	1	3	2,23
10	1	4	2,12	1	5	3,00
12	1	5	3,59	2	5	4,00

Question 12 presented the respondents with a set of terms they could use to describe their experience with QBD. The frequency of the terms selected is presented in Table 58:

Table 58 - Respondents choice of terms in question 12

Term	Overall		IFIM		KHIB	
Enjoyable	26	86,67 %	16	94,12 %	10	76,92 %
Creative	22	73,33 %	13	76,47 %	9	69,23 %
Time-consuming	18	60,00 %	13	76,47 %	5	38,46 %
Usable	10	33,33 %	3	17,65 %	7	53,85 %
Useful	8	26,67 %	2	11,76 %	6	46,15 %
Toy	7	23,33 %	2	11,76 %	5	38,46 %
Insufficient	6	20,00 %	2	11,76 %	4	30,77 %
Demanding	5	16,67 %	3	17,65 %	2	15,38 %
Easy	4	13,33 %	4	23,53 %	0	0,00 %
Complicated	3	10,00 %	0	0,00 %	3	23,08 %
Difficult	2	6,67 %	1	5,88 %	1	7,69 %
Effective	1	3,33 %	0	0,00 %	1	7,69 %
Quick	0	0,00 %	0	0,00 %	0	0,00 %
Useless	0	0,00 %	0	0,00 %	0	0,00 %
Efficient	0	0,00 %	0	0,00 %	0	0,00 %
Boring	0	0,00 %	0	0,00 %	0	0,00 %

- *Overall* represents all respondents who selected the term
- *IFIM* represents the number of respondents in the IFIM group who selected the term
- *KHIB* represents the number of respondents in the KHIB group who selected the term

Based on the terms selected, the respondents appeared to be positive towards using QBD to express image requests. A majority of the respondents found the approach *enjoyable* and *creative*, while none of the respondents selected *boring* or *useless*. There were some differences between the two groups choice of terms. A substantially higher portion of the KHIB group selected *usable* (54% vs. 17%) and *useful* (46% vs. 12%). Furthermore, the IFIM group selected *time-consuming* more than the KHIB group (76.47% vs. 38.56), while *complicated* was used more in the KHIB group than the IFIM group (23.08% vs. 0%).

The above results indicate that the respondents were generally positive towards using QBD, but found the process somewhat difficult. They indicated that while they were not willing to use QBD *instead of* text based queries, but might be interested in using it as an *addition to* text based queries. It also appears as if they found the QBD process *enjoyable* and *creative*, but that there might be some challenges related to its use.

These observations were reflected in the data from the interview sessions. With a few notable exceptions, most respondents reported that they were positive towards the QBD approach. Most stated that they had fun and enjoyed the process and some of the respondents compared it to a game. Several respondents stated that they clearly saw a potential use for the approach, particularly among the KHIB group. The following quotes illustrate the initial responses from the respondents regarding the QBD process:

> *It took a lot more time than using text. But it was much more fun! I mean, it's more pleasurable, it's like a game.* ***Respondent 2***

> *It was very fun. It's unfamiliar. Not just drawing like that, but searching based on shapes rather than words. It's a whole different way of thinking. [..] You had to use your creativity a lot more. But I think it's a very useful approach!* ***Respondent 8***

> *I think it's very interesting. Since I haven't used it before, I would probably use it as an addition to regular retrieval engines. But [..] if it becomes more widespread and connected in a similar manner to Google Images, I would probably use it a lot, since shapes often are much more interesting than the function of the object.* ***Respondent 16***

> *I think it has a very high potential, absolutely. In the beginning, I thought "no, why [QBD]?". But after trying it, I actually think it was very clever.* ***Respondent 29***

> *Exciting. Very large potential. It was very interesting to describe [the queries] using drawings instead of words. For the time being there were very many weaknesses, but with a well-functioning tool, then... There are absolutely many possibilities.* ***Respondent 31***

This indicated that most of the respondents had a positive and pleasurable experience expressing image retrieval tasks as visual queries. However, it should be noted that this may have been influence by the experiment setting. For most of the respondents, QBD represented a novelty. Having an opportunity to play with a novelty may have put the respondent in a positive mood. Similarly, the actual process of drawing and expressing something visually may be associated with something fun or be related to leisure activities. Furthermore, the respondents may have felt a desire to "please" the researcher, either consciously or subconsciously, as they may have felt that the researcher might have vested interest in the approach. This may have made the respondents adverse to describe the approach in a negative manner. Finally, most of the respondents were unfamiliar with the problems related to the CBIR approach. It is possible that they might be less positive if they had used the QBD approach in a realistic setting based on their own needs.

8.2 Willingness to Express Image Requests through Drawing

The respondents seemed positive towards the QBD process, and most of them had a pleasant experience when drawing the visual query images. However, they were divided in their opinions on whether they would actually use QBD in a real-life setting. Some were very enthusiastic, while others were more reserved. While there were different reactions towards using QBD in both groups, there was a notable difference between the two groups with regards to their willingness to *use* QBD. The respondents in the IFIM group seemed less willing to use QBD than the respondents in the KHIB group. This was indicated by their different answers in questionnaire II and by the impressions from the interview sessions. As noted in chapter 7.2 (page 194), this may be related to whether the respondents were able to imagine real-life situations where they *personally* would benefit from the QBD approach. The respondents who were able to identify some concrete cases where they would benefit from using QBD were also the respondents who seemed most willing to use QBD.

Concerning the major reservations *against* using QBD, four categories of reservations were identified based on the analysis of the interview sessions and the videos from the query sessions:

1. Frustration over the *systems' abilities to understand* the visual query images
2. Challenges related *to creating the queries*
3. The *expressive convenience* of QBD compared to QBT (Query by Text)
4. A *lack of content* in the collections.

Several of the respondents stated that they were dissatisfied and frustrated with the results they were presented by the two retrieval systems. These responses were fairly common, and are illustrated by the following quote from respondent 12:

> *I think it's fun that it works to a certain degree, but it is frustrating that when I was retrieving this dolphin with the red ball in the mouth, where I was very sure of what I wanted to retrieve and was very prepared to find, nothing was retrieved, only a lot of other things such as turtles and sharks and a lot of other stuff I didn't require at all.* ***Respondent 12***

When asked to identify *why* they achieved poor retrieval results, the respondents identified two potential causes. Some respondents blamed the poor results on their own inability to draw "good" visual query images. As noted in chapter 7.1 (page 190), a majority of the respondents in the IFIM group felt that they would be able to get better results if they had better drawing skills. This is illustrated by the following excerpt from the interview with respondent 20. “R” denotes the researcher, while “20” denotes respondent 20:

> ***R:*** *Why do you think the results didn’t match your expectations?*
>
> ***20:*** *It was too difficult... Drawing... I didn’t manage to draw properly, and the results were... They were not very good.*
>
> ***R:*** *Do you think you would be able to get better results if you were better at drawing?*
>
> ***20:*** *Yes... It probably had a major influence. Well, probably the system isn’t very good, but... I don’t think that’s where the main problem was. I definitely would have gotten better results if I was better at drawing.*

Several others of the respondents who rated themselves with low drawing skill made similar comments. They primarily directed their frustrations towards themselves, not towards the retrieval system. Some of these respondents also believed that the QBD approach might be primarily of use to “professional drawers”, and that they felt too inadequately skilled to use these systems. Some of these respondents stated that they felt this as a larger problem when expressing queries in VISI than when using Retriever, as they were better able to see how their actions were reflected in the query results. It is possible that this reservation could be reduced if the retrieval system could help the users to understand *how* they could draw queries without having to draw realistic queries, for example by using a dynamic presentation of the results.

Other respondents stated that they did not *trust* the two systems’ abilities to interpret and process their query images. This was particularly true when they failed to see any similarities between their query image and the images returned by the retrieval system, illustrated by respondent 4:

> *Well, of course the system doesn’t search on what I think it should search for. I believe that... at least for me, it’s obvious that I’ve drawn a dolphin, but the stupid system doesn’t understand that it’s a dolphin, it thinks that it’s just a lump with something sticking out, there in the middle of the image, right? [..]It has no understanding of what I’m drawing, and for me, as a user that is very problematic. I don’t trust the system’s ability to find anything resembling what I’ve drawn.* ***Respondent 4***

This lack of trust was primarily expressed when trying to express type 1 and type 2 queries, particularly when the respondents created the visual queries as “visual keywords”. When the respondents had spent considerable time trying to create a realistic representation of an object, and the system returned images that held no obvious similarities to the query image, these respondents often stated that they felt that the system behaved strangely. These reactions were less common when the respondents created more *complete* query images. In these cases the respondents were often better able to identify the similarities between their query image and the resulting images. Again, the lack of trust were more commonly observed when the respondents were using VISI than when they were using Retrievr: The dynamic presentation of Retrievr gave the respondents more feedback towards how they should create their queries in order to direct the system towards the type of images they were looking for.

Another important category of reservations was related to the *challenges* the respondents faced when expressing the queries, as noted in chapter 7: Challenges related to the drawing process, the time required to create the queries, the page-zero problem and problems related to expressing narrative content. These reservations were primarily voiced by the respondents in the IFIM group, and were mostly related to their drawing skills and the time required to draw the images. The respondents in the KHIB group expressed fewer reservations towards using QBD based on these challenges.

The third group of reservations were related to the *expressive convenience* of QBD, particularly when compared to text based queries. Several respondents stated that while they felt that QBD was *fun*, they believed that text based queries would be much faster, easier and more convenient:

> *It was fun, in a way, but it wasn't any revolution for me. But it was fun, it was a little entertaining, but I don't know if it will be very useful for me to do it. I guess it won't be.* ***Respondent 5***

> *[..] But I think it would have been much more effective to use Google and use keywords, or search using tags at Flickr, or using other ways of finding dolphins. I think that would have been more easy and effective. But it was fun!* ***Respondent 12***

This reservation was primarily voiced by respondents who were unable to identify situations where text based retrieval might present challenges, i.e. some of the respondents in the IFIM group. None of the respondents in the KHIB group voiced this reservation, indicating that it may be related to the respondents' normal needs and uses for image retrieval.

The final reservation voiced by the respondents was concerns about the lack of content in the retrieval systems, particularly in the VISI system. While they understood that the system was a prototype system, several respondents claimed that the QBD approach would be most beneficial for collections with a large and varied image collection. Several participants stated that they would very much like to see a similar approach used for Google images:

> *As I said, I definitely can see the potential benefit for this approach. But... I can't see myself using any of these systems. I'm not that interested in dolphins, and I don't think the Flicker system [Retrievr] has that many images. Now, if I could use this for Google Images, I'd be very happy.* ***Respondent 31***

> *I discussed this with some of my colleagues [before the experiment], and they wanted to know when this could be used against Google Images. I don't think these prototypes are very useful, they are too limited in their contents. But with all the images and shapes available through Google, this would be great. And very useful.* ***Respondent 16***

8.3 What uses do the Respondents see for Query by Drawing?

Despite the reservation some of the respondents had, all of the respondents could identify several situations where QBD might be useful. Several respondents also identified several situations in which they would *prefer* to use QBD, or at least use it in *combination* with text-based queries. Four categories of suggestions were identified:

- Searching for *specific images* and *images with specific motives*
- Searching for images with a *particular composition* or images depicting *a particular scene*
- Using QBD *in combination with text* to refine or limit the number of images retrieved
- Using QBD to express requests based on visual structure, i.e. overcoming *the problem of explicability*

Several respondents stated that QBD might be very helpful in situations where they were looking for *particular images or images with a particular motive.* Several of the respondents referred to prior experiences where they were looking for an image they knew existed on the Web, but were unable to find. Similarly, several of them referred to situations where they were looking for a certain object, but unable to express this verbally, as expressed by respondent 10:

> *[QBD might be very useful] if there were objects I had seen previously, and tried to draw it, rather than trying to, well, If I don't know the name of it, it's not possible for me to search using text.* ***Respondent 10***

While respondent 10's situation was primarily hypothetical, respondent 2 mentions a concrete example from his time in the military:

> *[It would be useful] if I was looking for something I didn't know what was, but knew what it looked like. It might be a strained example, but when I was in the military, they had these tank-recognition tasks.* ***Respondent 2***

The respondent continued to talk about these tasks. He claimed that that an approach based on QBD might have been very useful in this situation. If he could draw a silhouette of the tank, and have this compared to existing silhouettes, this would have been very useful for him. Several other respondents described similar situations in which text-based queries were difficult, as illustrated by this quote from respondent 22:

> *Well, it can be useful to draw an image when you don't see how… for example, an animal, you know what it looks like, but have no idea what it is or what it is called. Then it could be useful to draw the animal, and try to retrieve images of it or identify it. But if I'm looking for something simple, just an image of a tractor or things like that, I never would use such a tool. It would be much easier to use a text-based search.* ***Respondent 22***

A similar approach was described by several other respondents, illustrated by respondent 12:

> *Well, there are situations like:"Oh, I can't remember [the name of the film], but I remember the cover, it looks just like 'Jaws'" [..] and it would typically be the same for similar things you have seen: CD covers, a movie-poster, a traffic sign and what not.* ***Respondent 12***

Respondent 12 then went on to talk about how QBD might be very useful for a lot of similar cases, like finding signs and determining their meaning. Another example was made by respondent 14:

> *Text can be terribly difficult at times, because you know exactly what you're looking for, and you know that such an image must exist on the internet, but you just can't seem to find it. [..] So I think it would have been wonderful if I could search visually.* ***Respondent 14***

Aside from using QBD to identify objects or entities where they were unable to express textual queries as described above, the respondents were generally not interested in using QBD for generic retrieval tasks such as "finding a dolphin" or "retrieve images of a happy girl". They felt that these tasks were better covered using text based queries, and that they believed that QBD might not be very well suited for these tasks.

Another area the respondents highlighted was the use of QBD *to search for images with a particular composition*. First of all, several of the respondents in the IFIM group volunteered several situations where they *thought* QBD might be interesting, as illustrated by this quote from respondent 3:

> *If you're interested in finding an image you could use in a particular context, for example [an image of] a rural landscape, [QBD] could be very nice to do. You could say that you would like it in a particular way, with the trees in a particular place, and so on.* ***Respondent 3***

Note that while this and a large number of similar comments from the IFIM group were hypothetical situations and not based on real life situations or their previous experience, these examples illustrate that these users at least see a potential use for QBD. However, some of these respondents were able to offer some more concrete examples. One example is respondent 2 who suggested that QBD might be very useful for him when looking for very specific types of visual content:

> *For example, if I was looking for a house, [..] and if there was a toolbox of elements I could combine, that I would like three doors in front, and I could have drawn a square, and used two towers, and combined this to [A catalogue of houses]. That would have been useful.* ***Respondent 2***

Similarly, respondent 7 stated that he often used image retrieval system such as *Google Images* when retrieving images he could use when designing web pages, or when attempting to find existing web pages with a particular compositional structure. He felt that QBD might be a very useful tool in these situations:

> *[QBD could be useful], particularly for typically graphical elements, such as "I would like a red background with yellow dots, or blue stripes". It's difficult to use text to… describe accurately what I want.* ***Respondent 7***

> *It might have useful for, like, web design.[..] "I want to find pages that use these colours", and then draw schematically, describing what a webpage looks like.* ***Respondent 7***

The most concrete examples were presented by the respondents in the KHIB group. These respondents were generally very enthusiastic about the possibility of querying for images based on shapes and colours. Several respondents expressed dissatisfaction with text-based queries for such purposes. One example is the way respondent 16 talks about querying based on shapes:

> *The function I would use most is to query based on pure shapes, and not shapes that have a particular function, which you often have to do when using text based queries. Then you might not have to go through all the generic things you get when query by text, and rather go directly to the things you were looking for, [..] without having to know the names of the objects you are looking for, without naming anything [..] I've discussed this with some of my colleagues, and this is exactly what is required. So I'm very interested in this type of search.* ***Respondent 16***

The above examples highlight two important potential uses for QBD: Using QBD to retrieve images with a particular spatial composition, such as particular scenes, and using QBD to retrieve images based on the shape of an object rather than its semantic label, i.e. the name or its function. While neither of these is "new", i.e. they have already been described in literature, these tasks types generally have been rated low in prior studies of image retrieval tasks. While the number of respondents in this study was too low to draw any generalizable conclusions, the strong consensus among the respondents in the KHIB group of the importance of these tasks indicated that QBD may represent a very useful approach for image retrieval for users with such needs. Furthermore, it indicates that these types of tasks may possibly have been underrated in prior studies.

Another area highlighted by the respondents was the *possibility of using QBD in combination with text based queries*, particularly when requesting images from various web-based retrieval systems such as *Google Images*. Most of the respondents stated that one of their main reservations with using query-by-text for image retrieval was the enormous amount of images returned by these systems.

Literature in the CBIR field often claims that a major problem of the text-based approach is that there are many images that never get retrieved because of the problems of volume and subjectivity. However, the respondents claimed they were not very bothered by this. The respondents were usually capable of finding one or more relevant images unless they had very specific needs. Their main problem was that it often took a substantial amount of time to *sort through the retrieval results* in order to find these relevant images. Text-based queries often returned images with absolutely no relevance to their query tasks, and they often had to spend a lot of time to identify which images might be relevant, often by browsing through numerous pages of image results.

Related to this, some of the respondents who gave an impression of understanding the problems of the semantic gap stated that they felt that they would be able to get much better results by combining QBD with textual tags. Respondent 14 illustrated this when discussing his earlier experiences with text based queries and how he felt QBD could be used to narrow the results obtained through these queries:

> *[..] and then you have to go through terribly long searches in order to find [the relevant] images. Now, I haven't actually done this very often, but the main reason for not doing this is that the results are so large, and that the tags and words describing [the images] never seem to fit what I would have used to describe the images. So I think [using QBD to narrow the results] would have been absolutely wonderful!* ***Respondent 14***

Similarly, while discussing some of the problems with QBD, particularly in relation to the semantic gap, some respondents stated that using QBD might be more interesting if it could be combined with text:

> *I don't think it is very likely that I would replace text based queries with [QBD]. But, as a supplement, [QBD] would be very interesting, and that is because the text could be used to filter out irrelevant images. You would get more relevant images in the collection [the query] is compared to. For me, and the way I've used [QBD] now, it would be as a potential addition to text-based queries, or as a mechanism for filtering the results.* ***Respondent 15***

> *Yes, it could. What I think could have been positive by combining this with a text based search, is that you could have had both the situation and… You might get the motive, the actual motive, in the text search, then drawn a little different, or tried more to get the actual situation [..] It could have given more precise search, then. Not see seagulls when looking for a dolphin.* ***Respondent 11***

Several other respondents presented similar views. This was particularly true for the KHIB group, where several respondents had prior experience with similar problems. One of the designers

(Respondent 17) discussed a problem he often faced when trying to retrieve images of depicting specific objects:

> *When we're designing, we often seek inspiration, or wish to see if what we've designed is too similar to what others have made. In these cases, you definitely wish to [find images] belonging to a certain category. In these cases, you could use this to make a thin outline of the product you're considering. And then you might wish to say, "Now I'm making a lamp, and I'm envisioning a singular lamp screen", or I'm envisioning 4 different lamp screens. In these cases you have a clear mental image of the type of lamp you wish to retrieve. And you only wish to find [images] of visually similar lamps. And not have to browse through the enormous amounts of lamp images available on Google Images.* ***Respondent 17***

These examples highlight two important issues for the role of CBIR in image retrieval. First of all, the problems related to a lack of proper annotation might not be very important for the users, particularly when querying very large collections: The users will often be able to find one or more images that are *good enough* from the vast number of images available. Furthermore, while CBIR might not presently be able to do proper segmentation and object recognition, it may still prove useful for *filtering* image results, particularly when the domain of the results have been narrowed down by a textual query.

Several of the respondents in the KHIB group discussed how QBD could be used to solve the problems of *explicability*, such as finding images of flowers in a particular colour or finding images with more abstract content. An example of the former is the following excerpt from respondent 27s query session. He interrupted the query tasks in order to discuss an earlier situation where he was looking for images of red sofas:

> *I was looking for an image a red sofa. Do you have any idea of how many images there are of red sofas at Google? There are millions! And I just wanted to find sofas with three pillows, and in that particular shade of red. I think I spent a quarter of an hour just browsing through those [..] results. I f I had this tool [referring to Retrievr], it would have been much easier. That's why I tried to find a red sofa for that previous assignment. I want to see if I could find a red sofa. And I found an exact image of that red sofa in under a minute! I think [QBD] would be very useful for this type of task.* ***Respondent 27***

While this particular example was most likely a result of a lucky incident, it further illustrates one the potential benefits of QBD in terms of filtering results - the ability to precisely define the query parameters of a query using visual techniques. And while this result is also not "new", the empirical data from this project indicates that the users' ability to precisely define their queries using QBD may present them with real benefits over text based queries, indicating that there is a demand for these systems even if the current technology is lacking in some regards.

8.4 Summary: Respondent Opinions and Attitudes

First of all, most of the respondents stated they had a pleasurable experience using QBD, and were generally positive towards the process of expressing QBD queries. Only 1 respondent explicitly stated that he did not enjoy the experience. They were not directly negative towards using QBD as a tool for image retrieval. Furthermore, it seems as if the KHIB respondents were more positive towards QBD than the IFIM respondents.

However, the respondents were more divided on whether they were *willing* to use QBD. The biggest reservation seems to be directed at the challenges related to the QBD process, particularly difficulties related to the actual drawing process. Furthermore, several of the respondents stated that they did not have complete trust in the CBIR systems ability to properly interpret and process their queries, either because they were incapable of expressing good queries, or that they did not trust that the system would be capable of processing the queries in a meaningful way. The respondents would also like to try the approach on a large scale collection.

Moreover, few respondents seemed willing to use QBD as an alternative to text based queries, except for some specific retrieval tasks. They were more willing to use QBD as an addition to text based queries.

Summarized, four main categories of potential useful areas for QBD were identified based on the interview sessions:

- Searching for particular images and images with particular motives
- Searching for images with a *particular composition*
- Using *QBD in combination with text* to refine or limit the number of images retrieved
- Using QBD to express requests based on visual structure, i.e. overcoming *the problem of explicability*

Based on the above discussion, research hypotheses 4.1, 4.2 and 4.3 can be evaluated:

RH4.1: The hypothesis must be *rejected*. Most of the respondents stated that they had a pleasurable experience expressing image queries using QBD. However, this needs to be taken with two reservations:

1. The experiment setting may have influenced the results in a positive direction. It is possible that the respondents might have been less positive if they had used QBD in a real-life situation.
2. While the respondents felt the actual process of expressing the queries pleasurable, there were several reservations towards their willingness to actually *use* visual queries.

RH4.2: The hypothesis must be *accepted*. There was a definitive difference in the way the two groups felt towards using QBD. While both groups were generally positive towards using QBD, the KHIB group seemed more positive. They were better able to express *in what situations* they would benefit from QBD than the IFIM group. The KHIB group also had fewer reservations towards using QBD, and did not see the major challenges (time and drawing difficulties) as a major obstacle towards using the QBD approach.

RH4.3: The hypothesis must be *accepted.* While most of the respondents saw some potential uses for QBD, very few of the respondents were willing to use QBD instead of text based queries. However, most of the respondents saw a number of areas in which they felt QBD could be a *complement* to text based queries, particularly with regards to filtering the results obtained from a text-based search. In addition, some of the respondents, primarily in the KHIB group, identified some particular areas where they claimed QBD might be *better than* text.

9 Respondent Suggestions for Improvements

The final study objective in this work was to identify potential improvements to visual query interfaces in order to better support users when drawing visual query images, as expressed by research question 5:

> *What improvements can be made to CBIR systems in order to better support users when drawing visual query images?*

This was a major topic during the interview sessions, and answering this research question was primarily based on these sessions, combined with observations made during the query sessions and the analysis of the interface videos. A large number of suggestions were identified based on this, but five topics stood out as the most important suggestions:

1. The ability to modify and deform the basic geometric shapes provided by the drawing tools
2. A more varied set of drawing tools more suited for the image retrieval tasks
3. The possibility to express queries using pictograms and icons
4. A colour-neutral canvas and a colour-neutral drawing tool
5. A fully dynamic and integrated query process

These categories are discussed with regards to *what problems* the improvements might solve for the respondents, the *feasibility* of adding these suggestions and the *consequences* these suggestions might have for the way the CBIR systems process query images. The remaining suggestions are briefly presented in section 9.6, and a summary of the suggestions is presented in Table 59 (page 235).

9.1 Deformable Shapes and Objects

The most frequent suggestion was a request for more flexible drawing tools. There were several suggestions related to this that can be placed in three major categories:

1. The ability to *modify* and *deform* the shapes provided by the drawing tools
2. The ability to *manipulate* objects and elements already added to the canvas
3. The ability to *group* objects and elements added to the canvas

One of the primary reasons why the respondents were reluctant to use the shape tools provided by VISI was that these tools were not well suited to the retrieval tasks they were performing. However, several respondents claimed that these tools would have been much more useful if these were *deformable*, i.e. the users could alter the appearance or the dimensions of the shapes created by the tools:

> *[The shapes would be better] if you could deform them. I'd like to use an oval circle as a starting point, but I'd like for it to be a bit taller, so I'd just like to grab the back of the circle and lifted it up [in order to create a dolphin].* ***Respondent 2***

> *In some applications you have the option to draw a line, and then you can pull and twist the line [and] manipulate [the shapes], for example a straight line can become an arc, right? Then you could join two arches and make the body of a dolphin, or whatever you're making. That's something I'd use if it was there.* ***Respondent 6***

> *I would love to have access to vector tools [..]. If I could quickly draw a fish [using the shapes] and then use these... What are they called... points to adapt, drag or move it around.* ***Respondent 19***

Related to this was the fact that once something was *drawn* on the canvas in either of the systems, it immediately became *a part of* the canvas. For example, if the respondents used tools to draw circles or lines, these were simply drawn on the canvas and could not be manipulated any further. The respondents reported a desire to be able to modify the objects already placed on the canvas similarly to deforming the shapes: They would like to deform, rotate and resize the objects on the canvas:

> *[The objects you have drawn] could have some nodes attached to them that you could pull. You could modify it, like "Now, that fin looks totally weird, I could change the curvature of the line", and I'd get a shape that I'd be able to process after I've put in on the image, that I could move about and deform after it's been drawn.* ***Respondent 12***

> *The thing with a more vector-like approach is that once you've drawn something like dolphin, you'd be able to pull it, move it around and adjust the way it's twisting when it's jumping up there in the air.* ***Respondent 22***

> *What would have been very practical was if it was more like a vector-like tool. [..] You could be able to change the shapes, so that they're not static once they're drawn. You should be able to change the things, move them around, kind of what you can do in Illustrator*[104]*.* ***Respondent 24***

Several respondents also reported that they sometimes were disappointed with the appearance and quality of the lines and shapes they created using the freehand tool. Some of these respondents suggested that the drawing interface could *smooth*, *correct* or *improve* their drawn elements:

[104] Adobe Illustrator - A vector based computer drawing tool

> *In some drawing applications they have this thing, when you draw something it straightens it for you. It would make the drawing process more comfortable, like, you're writing all pointy and rough, and then it just... it smoothens the lines for you... To help those of us who are not very precise with that [digital pen].* ***Respondent 7***

Finally, several respondents expressed a desire to be able to group different objects and use layers when manipulating these object groups. This might provide them both with an easier way of manipulating the objects and a possibility to indicate *spatial structures* in their query images:

> *If you had [vector based drawing], then you could manipulate dolphin. Or just enter it as a unique object, like in Photoshop, where you just could use a layer, and pull the entire layer into the ocean. Then it's swimming up towards the left part, rather than just being there.* ***Respondent 22***

> *If I could use layers, that would be great. I'd be able to describe that the lousy dolphin is actually op top of the sea, not inside the sea, or… I could just drag the dolphin layer down into the sea, so I could show that the dolphin is actually coming out of the sea, not just partly above and partly below the surface.* ***Respondent 4***

> *The problem was that once I had drawn the buildings, I couldn't go back and edit the background, because I hadn't started with the sky. But if I had more layers, I could just move all the buildings around, and then add the background or the details. But now, it was just too much of a hassle.* ***Respondent 30***

All these suggestions were variations of a single theme: Using a *vector* based approach rather than *bitmap drawing*. By creating *objects* of the drawings rather than *bitmaps*, the users would be able to manipulate and deform the shapes and the objects they had added to the canvas. Based on their statements, it is possible that this approach may increase the versatility and flexibility of the drawing tools, and allow users to use these tools more when representing real-world objects. It is possible that this may reduce some of the challenges that were experienced when drawing. Grouping the objects together or adding them to layers may make it easier to create more *complete* queries or focus more on the *composition* of the query images. As noted in chapter 6.4, the respondents often started to draw the query participants in the middle of the canvas in order to make sure they had enough space to complete the drawing. Adding additional details to the background, or moving or manipulating the drawn objects after they were finished were difficult. In some cases the respondents just finished the query even if they ideally would like to add more details. Furthermore, if the queries were made using *objects* rather than *bitmaps*, the query images might present the retrieval system with additional information about the *contents* of the query image. Rather than letting the CBIR algorithms try to segment the query images, these images would be segmented based on the objects and object groups created by the user. Additionally, if the query image was created using *layers*, additional spatial information could be extracted from this, such as one object being placed “behind” another object.

It should be noted that a few respondents were very vocal *against* this idea, particularly the two respondents with a master in fine arts and one respondent who stated that she previously had very negative experiences using vector based drawing. They felt that vector based drawing would be too restrictive and take away the freedom and flexibility of freehand drawing. It is possible that while a vector based approach might provide the users with more *flexibility* and *versatility*, it might *reduce* the expressive convenience provided by freehand drawing.

Summarized, the respondents would like the ability to manipulate the objects once they were added to the canvas. This included the option to *modify and deform* the objects by dragging them, *group objects*, *add objects to layers* and *move objects without changing the rest of the image*. This might provide the users with more flexible and versatile tools, and it could be used to assist the retrieval system when segmenting the query images. However, the usability of this approach needs to be evaluated in further user studies, and the feasibility and potential benefits with regards to segmentation needs to be tested.

9.2 More Usable Drawing Tools: Shape Templates

A majority of the respondents stated that they would use drawing tools more if these tools were more suitable for the domain they were working with. A very common request was the possibility of having *shape templates* representing the central concepts they were attempting to retrieve. The suggestions varied from a very high level of detail, e.g. having access to shapes representing detailed elements such as *a dolphin fin* or *the beak of a seagull*, to high level abstractions such as *human*, *fish* or *boat*:

> *I would like to have a set of shapes to aid me, [..] more domain specific such as whales. And it would have helped having fins, some wings, legs, eyes, beaks and similar things.* ***Participant 2***

> *If there was some menu options, where you could find... Let’s say, you’re drawing a whale. Then you go to the menu, select “animals, mammals, whale”, and then it inserts the shape of a whale into the canvas.* ***Respondent 5***

The respondents highlighted two main reasons for requesting such templates. First, some respondents stated that shape templates might reduce the impact of their low drawing skills. By being able to just “drag and drop” images, drawings or sketches that were of “better quality” than what they were able to make themselves, they would be able to create queries without having to worry about the visual modality of the query image participants:

> *[I think that] having the ability to select from a set of predefined shapes which are difficult to draw, would be very nice. Or, I don't know... I guess that since the boats I was able to draw were really poorly drawn I'd get much better results if I just could use a finished shape looking something like the things I was going to search for.* ***Respondent 3***

> *Well, there are things like bird heads, beaks and other things that are very difficult to draw unless you're very skilled. But if you could have some characteristic shapes as a starting point, like a dolphin, with the head and beak in front, then you might be able to draw the rest of the animal in whatever shape you'd like. [..] Such prefabricated shapes would make it much easier for me… Comparing my [drawing of a dolphin] to... Actually, it looks more like a seagull. I definitely would have found using more domain specific shapes much easier. And I'm sure I'd get better results too.* ***Respondent 15***

Several respondents also stated that shape templates might make the query specification process much faster:

> *Using [templates] would at least make it much easier to create the queries and get results. For example, an application at the Bergen Aquarium might typically have a set of predefined shapes of things they know are in the database, such as dolphins and penguins. Rather than letting users spent three quarters of an hour drawing the queries and then get a small dolphin in the end, they could just compose the query... A faster way of getting more meaningful results.* ***Respondent 12***

> *If I had access to a palette of [templates of objects like dolphins and fish] and was querying for a dolphin, then I might just drag-and-drop the dolphin shape to the canvas, resized it to a suitable size, and then maybe added a boat and some things… That would have made it faster, very much faster.* ***Respondent 9***

Some respondents stated that while they would love to have access to more domain specific shapes, they were worried it might cause the interface to become more difficult to use, and that finding and identifying the relevant shapes might be difficult and time consuming:

> *Well, I'd love to have [templates], but I'm thinking… If, for example, I had access to a number of these predefined shapes that I could manipulate… This would mean that the contents of my database would have to be very limited. That is, I could only have three or so different concepts in the database, and not endless. [..] So I guess I would have preferred the system to have these, but not an endless number of them. If I have to browse through 15 pages of predefined, domain specific shapes, then I think it would have been too much. The benefit would be removed.* ***Respondent 4***

> *Yes, [having access to templates] would of course have helped, at least here, since it's a very limited domain, this would have made the process much easier. But it's like… It would only be applicable when you have a small number of shapes. In a global query, you'd have to have an enormous amount of different shapes. I guess you could have different categories of shapes or a tree-structure, making it easy to navigate to the relevant shapes and just drag-and-drop them into the query, but I guess this would be too complex to make.* ***Respondent 10***

Some respondents also had had some reservations towards the *variance* of the different shapes. They noted that the predefined shapes might not have the desired format, e.g. a default *dolphin shape* might not necessarily be created in the desired pose, scale or rotation:

> *While it would have been nice [to use templates] the shape might change from image to image. It might be beneficial if you had a model you could manipulate, or move and change flexible joints and so on, but… I think you'd need a very large number of shapes for each concept.* ***Respondent 20***

> *[Templates] would be nice and OK, but only if I could have flipped, scaled and rotated it. In that case, it would be great. If I could get a predefined dolphin, then could.. If I could turn and twist it in 3D without destroying the shape, but just turn it… Then I could finally create that drawing of a beached whale. Otherwise it would… I don't think I could have been bothered browsing through endless numbers of different shark shapes.* ***Respondent 4***

Finally, a majority of the respondents in the KHIB group were negative towards using predefined shapes. Most of them immediately identified the problems of visual variance in most real-world, living objects. Additionally, they stated that they felt that using templates would take too much time and reduce their freedom and flexibility. They also claimed that they thought that using predefined shapes would primarily be of benefit to users without well-developed drawing skills:

> *I actually think [using templates] would have removed quite a lot of the speed of the query process. I think it's much faster just to draw a dolphin very quickly. Create a sketch and maybe modify it, creating a very abstract representation. Because if you're going to add shapes for dolphins, you'd suddenly end up with 500 different templates, you could draw a dolphin in two seconds, rather than spending 15 seconds looking for a template.* ***Respondent 31***

> *Well, I don't know. You'd need at least 200 different dolphins, or you could just type "dolphin". Because when you're drawing the dolphin on your own, you're recreating the mental image you're trying to retrieve, and that is always in a particular way. You just can't have a predefined shape for that.* ***Respondent 17***

> *Oh no! That [Templates] would have been far too messy! (Laughs) But then, I guess normal people with fear of drawing might believe that they needed it. You'd need an entire library of different shapes and representations of different concepts.* ***Respondent 25***

> *I'm thinking [templates] are not really necessary. It probably wouldn't help much. There are no standard images, so why use standard shapes? Or standard representations of a human or a dolphin. They are all unique.* ***Respondent 28***

Including templates appeared to be something that a large number of the respondents were interested in, particularly among the respondents without a visual background. While two respondents in the KHIB group were positive towards this topic, they also rated themselves with low drawing skills. This indicated that the possibility of using shape templates appealed more to respondents who were unaccustomed to drawing. The respondents most negative towards shape templates were the same respondents who stated that freehand drawing provided the highest freedom, expressive power and expressive convenience.

This is reflected in what the respondents said when discussing the potential of these shapes. As noticed above, they discussed several reasons for why templates would be beneficial. The main reason for their enthusiasm towards shape templates was that it might reduce the problem of low drawing skills. They claimed that using shape templates would allow them to benefit from the compositional nature of the QBD CBIR approach without being limited by their ability to draw realistic objects.

The largest obstacle towards including shape templates in a QBD CBIR system is that creating a library of such shapes, even for a small domain, would represent a major undertaking. Depending on the level of detail of the shape templates, the number of different concepts and potential variations in shape of these concepts might escalate very fast. Creating a large number of these shapes would require a significant effort, and creating interfaces that are easily navigated by the user when selecting a shape will present additional challenges.

Furthermore, given the high degree of visual variance of living objects, using shape templates alone will still present CBIR systems with the challenges of the semantic gap. Even if the user can *select* a shape representing a dolphin, the retrieval system *still* faces the problem that this shape is very similar to that of a banana (e.g. Figure 11 and Figure 12, page 26).

However, it should be noted that shape templates could be used to provide the retrieval system with additional data about the query image. Using independently defined *objects* rather than bitmaps might assist the retrieval in segmenting the images. Predefined shapes could be given descriptive *semantic labels*. The labels could assist the retrieval system in *filtering* results. If a shape labelled as a *dolphin* was included in the query image, a retrieval system could use this information in combination with text-based tools such as a *thesaurus* to identify specific images or collections. Even if the shape templates might not improve the similarity functions, semantically labelled objects could provide the system with information about *which* objects are present in the image and the *spatial placement* of these objects. This may be particularly useful if the images in the collection are pre-segmented in a similar manner.

Summarized, a majority of the respondents request access to a larger set of tools, particularly *domain specific tools*, e.g. the option do *drag and drop* shapes representing the objects they were requesting. This would allow them to quickly compose the queries and give them a method of specifying image requests of a spatial nature without requiring high drawing proficiency and without spending a lot of time. Furthermore, several respondents stated that if *shape templates* could be combined with the possibility of *deforming and manipulating* the drawing this would give them a very high expressive power and expressive convenience. Creating a library of such templates may represent a major undertaking, and using such shapes is still prone to the problems of the semantic gap. However, if an approach based on shape templates could be combined with text based techniques, the approach may potentially provide users with a simple way of expressing image queries where the spatial structure of the request is of high importance.

9.3 Using Icons and Pictograms to Express Queries and Query Contents

Several of the respondents stated that they would prefer using a symbolic approach for expressing the queries rather creating realistic drawings. Two issues related to this were discussed: Expressing the queries entirely using icons or pictograms, and expressing complex content using icons.

First of all, several respondents stated that they didn't like the idea of creating realistic query images. However, they did like the idea of expressing visual structures using visual queries. Several respondents suggested using *pictograms* in order to represent the query objects, by quickly composing a background or colour, and then spatially distributing relevant pictograms. Similarly to using predefined shapes, some respondents felt that this would give them access to the power of QBD CBIR without having to draw realistic query images.

> *If you're going to have a complex query using humans and dolphin, then you could just compose the query using ready-made pictograms, right, as a kind of drag-and-drop manner. Then you could just find the humans, dolphins and all that other stuff. Drag it in along with the trees, and then draw the mountains and the sea and the background. I think the combination would be great. Colour the*

> *background, add pictograms, add some more colour and then you'd have a complete query.* ***Respondent 23***

Other respondents would like the *option* to use this. They would like to draw some objects, and use pictograms to represent other objects. If they were interested in images with dolphins jumping in front of a group of humans, they could create a detailed representation of the dolphin, and add several people or even a "crowd" using pictograms.

Related to this, is another potential use the respondents saw for pictograms. One of the more difficult things to express, according to the respondents, was different types of complex contents, such as *actions*, *interactions* and *conditions*. Several respondents reported that they would have enjoyed using icons or pictograms in order to express this type of content, e.g. by adding a "red cross" or a "heart" in order to signify "aid" or "help", using a "star" in order to indicate "actions" or "violence", or using "lines" or "Puffs of air" in order to signify movement. Several respondents compared this to how similar actions are represented in cartoons.

> *For example, if the system could understand that if I placed a "heart" between the beached whale and those lumps [referring to humans] meant "nursing", that would have made things much easier. If the computer reacted symbolically... Or it reacted to certain symbols in addition to recognizing the shapes… It depends on what you're after, but… I think it would have been far more effective.* ***Respondent 28***

> *[In order to express that a group of humans are aiding a beached whale] I guess I could have… I could have drawn a red cross on the back of one of the humans, or something like that. Just add a symbol indicating help, or assistance, or something like that. A symbol for help would have been nice.* ***Respondent 26***

> *Well, I realize that using symbols and stars and stuff won't work well, but it is a… It's limiting the search tool. You're already able to select from a lot of different symbols as it is, but it would be great if… As some sort of addition… If you know you're after humans running or something like that, then you could have symbols for "running" or "humans running", and place it where you like.* **Respondent 30**

Using pictograms or icons to represent the participants or other structures in an image is more or less an extreme variant of the *shape template* suggestion. However, rather than using shape templates that *resemble* the objects they are representing, *pictograms* or *icons* represent another layer of abstraction in the query images. While the use of pictograms or icons would allow the user to indicate *what* objects are present in a query image and *where* these objects are spatially located, the query system can no longer rely on similarity functions when processing the query. This requires that the system must first *interpret* the query and *requires* that semantic labels are used. Similarly, if *narrative contents* such as actions, interactions and conditions are represented through icons, the retrieval system needs to process the queries and interpret these symbols in addition to using similarity functions. This approach may be feasible when searching through images that have been pre-segmented and semantically labelled, but it is difficult to imagine how this could be used when querying against uncategorized images.

Summarized, some respondents wanted to have the option to use pictograms to compose the query images. Some respondents stated that they would prefer to represent the query objects through pictograms, either by drawing these pictograms or by building the queries using a simple drag-and-drop method based on existing pictograms. Other respondents stated that they would like to use pictograms or icons in order to detail queries with *narrative content*, as this proved very difficult when expressing the queries. This approach may have a very high expressive convenience, e.g. the users can express the contents and spatial distribution in an easy and effective manner. However, it cannot be used directly in CBIR system based on traditional similarity functions: the query processor must be able to interpret the meaning of the pictograms. The feasibility and usability of this approach needs to be determined through further studies.

9.4 A Colour-Neutral Drawing Tool and a Colour Neutral Canvas

As noted earlier, a majority of the queries for generic objects were created as “visual keywords”, i.e. by drawing one or more objects on a white (neutral) canvas. The respondents were generally dissatisfied with the results this returned. When they were told that the system interpreted this as queries for images that were primarily white (“White background”), some respondents were disappointed by this. Several respondents stated that they would like either have the “white” canvas represent a “neutral” background, or have completely “neutral” background, as illustrated in Figure 116, where a yellow object is placed on a colour-neutral background. This would allow them to choose if they were requesting general images of an object, or images of an object in a natural setting:

> *Having the option to include a colour-neutral background would allow me to choose if I’m looking for just general images. I could choose if I’m simply searching for images of dolphins, or for dolphins in their natural habitat… Or dolphins in a specific setting, such as that [..] dolphin in the Norwegian fjords. That would have been very nice.* ***Respondent 16***

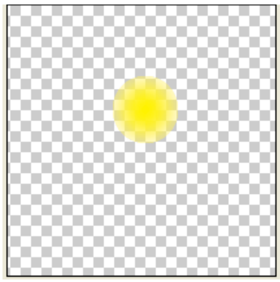


Figure 116 - A yellow circle placed on a neutral background

Similarly, several respondents stated that they did not feel happy about being forced to select a colour in Retrievr, and that they would prefer to draw without having the CBIR system focus too much on the colours they used:

> *You should have the option to select if you want the system to be colour-specific or not. Then you could… Just draw a general thing, and not consider the colours. Some sort of colour-neutral colour, or (laughs)… I guess I mean a colour-neutral tool.* ***Respondent 24***

> *There should be a tool just for shapes. Maybe a particular colour or something, so that when you use that colour, the system understands that it should only read the shape, and not the colour. [..] Now, when I was trying to find a chair… I know what a chair looks like, but colour? I didn't want to decide upon what colour it should have, or rather, I didn't have a clue. I just wanted a chair.* ***Respondent 25***

This was primarily related to the Retrievr system, as VISI allowed the respondents to specify that colours should not be given any weight. But even in VISI, some of the respondents stated that it would have been much easier just to select a tool, a pen or a specific colour, and then the system should disregard colour for the particular object drawn using that tool, pen or colour.

These remarks are also directly related to the desire to ensure *relevant* contents discussed in chapter 6.2: The respondents did not want to retrieve images that are *irrelevant* for their requests or to *exclude images* that contain *relevant* contents.

Summarized, several of the respondents reported frustration over the systems interpreting *white* as a colour and not as a *neutral canvas*, and that *black lines* were treated as *lines coloured black* and not as a basic representation of the *shape* of the objects. They stated that they either would like to have a *colour neutral tool and canvas*, or the option to specify that the retrieval system should only focus on their shapes, not the colour of the objects or the colour of the canvas. Even though the VISI system supported this through defining the query parameters, the respondents did not find this to be a very optimal solution.

9.5 A Fully Integrated and Dynamic Query Process

As noted above, the respondents were generally very pleased with the dynamic query process presented by the Retrievr system, where the image results were updated whenever they paused in their drawing process. All respondents agreed that a similar approach would be very useful for the VISI system. Similar opinions were also voiced in the first experiments by respondents, who were not given an opportunity to use Retrievr. This might allow them to be more efficient when creating the queries and make them better able to understand how their queries were processed by the system:

> *[Real time results] would have been really neat. I'd be able to see what the computer was thinking. And then maybe you'd go "Ok, I'm on the wrong track here, we'll try again", or go back and modify the image. That would have been really great.* ***Respondent 8***

However, the most pressing issue seems to be with the parameter specification phase, as this was both the least used and the least understood element of the VISI query process:

> *If I had the option to use the same query image and try other combinations of the [query parameters], I'd be able to see how the results changed according to my weights. It certainly would have made the process less cumbersome.* ***Respondent 1***

Several respondents stated that adding real-time result presentation based on the query parameters would be one of the most useful additions to the VISI query process. Furthermore, several respondents stated that they would like to choose between simple and advanced settings for the query parameters. Advanced settings would allow the users to precisely define the query parameters, similar to the current query parameter definition in VISI. But they would also very much like to see *simple* or *automatic* query parameter definition:

> *If it could be possible for the system to determine the weights... If it could be possible that the system understood what the weights should be, then it would be much easier to do the queries. Like, if you're using other colours than black and white, then the system would put more weight on colours, increased weight the more colours you used. And if you used the texture tools, then it would add weight to textures and so on. I think that would have been much easier.* ***Respondent 10***

> *What annoys me is that the system keeps searching based on colours, even if I've not used any colours at all. This [Points at the screen showing query in VISI] query here, I didn't use any colours or any textures at all, and yet it insists on using including colour in the queries. It's the same with that Flicker system.. It's like.. It insists on searching on colours all the time, when I didn't use any. It's not that difficult to determine that I've not used any colours at all, is it?* ***Respondent 23***

Summarized, all respondents agreed that they preferred the dynamic query process and result presentation provided by the Retrievr system. This allowed them to focus their drawing process on the elements they found worked best, they were better able to understand how the system interpreted their queries, and they felt they expressed the queries more efficiently. Most respondents also reported that while they liked the option to define the query parameters, they would prefer to have the option to specify the parameters *after* they had created the query, and have the results *dynamically update* based on their query parameters. This would make the query parameters more useful, and they felt they would better be able to understand how the parameters influenced their queries. Some respondents also stated that they would prefer if the query system could automatically specify the query parameters based on their actions, e.g. include colours or texture if they actively included these elements in their queries.

9.6 Other Suggestions

In addition to the suggestions detailed above several other suggestions were made by some of the respondents. The remaining suggestions are briefly described below.

- **Multimodal queries - text:** Almost all respondents stated that they would prefer combining the visual queries with text based queries, using it to filter the image set either prior to or after the query process. This is discussed in detail in chapter 8.3
- **Multimodal queries - sound:** Two respondents stated that they would like to use sound to aid the queries, either by stating keywords (“Shark!”) or singing / whistling a song (e.g. the theme from “Sharks” in order to find images of sharks)
- **Texture tool**: Some respondents suggested that a tool for adding textures should be included, allowing them to use textures rather than colours to fill objects. These textures could either be uploaded to the system by the user, or selected from a set of textures (e.g. “Sea”, “Grass” and “Bricks”).
- **Dynamic pen size:** The ability to freely choose the size of the pen tool. Several respondents claimed that this would render the different pen tools and sizes in the two systems redundant.
- **Colour gradient tool:** The ability to select two different colours and use this to create colour gradients when filling objects or the background with colour gradients.
- **Simpler colour selection:** Reduce the number of available colours in the interface to a few basic colours, but allow the user to select from any colour using a colour palette, simplifying the interface will increasing the expressive power.

- **Dynamic canvas size:** The ability to resize and reformat the canvas in order to support retrieval of different image formats, such as *portrait* and *landscape*.
- **Zoom tool:** The ability to zoom or enlarge parts of the canvas in order to add minor details

9.7 Summary: Respondent-Suggested Improvements

The respondents identified and discussed a large number of potential improvements that they believed would improve their experience of the QBD CBIR process. A summary of the suggestions is presented in Table 59. The main suggestions can be summarized as:

- Include deformable shapes and objects in the QBD interface. This would provide users with the ability to manipulate shapes once they are placed on the drawing canvas. Users could get more flexibility when using the predefined shapes and reduce some of the problems they experienced when trying to use these shapes. The predefined shapes could also be a viable alternative to freehand drawing for users who feel they lack drawing skills. It might also potentially assist the retrieval system when segmenting the query images.
- Include shape templates representing objects important to the query domain, allowing the user to use these rather than drawing them by freehand. This might reduce some of the problems related to creating realistic drawings, and potentially reduce the time required to create the query images.
- Support the use of *icons* or *pictograms* for object representation. This would allow the user to focus on the spatial characteristics of the query without having to spend time creating realistic representations of query objects.
- Support the use of *icons* or *pictograms* for representation of narrative content. This would reduce the problems related to expressing narrative content.
- Include colour-neutral drawing tools. This would allow users to create queries without having to use colours or spend time on defining the query parameters.
- Present the user with a fully integrated and dynamic query process, including automatic definition of query parameters. By presenting results in real-time, users might be better able to understand how the retrieval processed the visual queries. Automatic definition of query parameters based on user actions could potentially increase the speed of the query process, while still giving the user some influence on how the system should interpret the query images.

Table 59 - Respondent suggested improvements

Area	Suggestion	Goal
Tools	Be able to modify existing shapes	Make the drawing tools more useful and usable through deformable shapes and layers
		Give the user with higher expressive power and convenience when using the drawing tools
	Shape templates	Reduce impact of low drawing skills by presenting the user with a pre-defined shape, removing the need to draw difficult objects
		Reduce the impact of the page-zero problem
		Present the user with a more efficient and faster query process
	Colour neutral drawing tool	Let the user express queries based on shape alone
	Query based on icons and pictograms	Give the user higher expressive convenience
		Let the user express complex requests in a simple manner
	Add a tool for texture specification	Give the user the option to quickly add textures to surfaces in the query image
	Dynamic pen size	Reduce interface complexity by having resizable pen tool
		Allow the user to draw with a higher level of detail if desired
Colours	Add a tool for creating colour gradients	Give the user a simple way to use colour gradients in queries
	Free definition of colours	Let the user freely select the colours from a full colour palette rather than choosing from a pre-defined colour palette
Canvas	Zoom	Give the user the option to zoom part of the canvas to add additional details
	Colour neutral canvas	Give the user the option to express queries for generic objects without regard to the background colour
	Dynamic canvas size	Give the user the option to freely define the size and format of the canvas in order to query different types of images
Query process	Combine with text	Allow the user to use text to filter images in the collection (prior to and after the query)
		Allow the user to use text to specify query parameters such as complex content
	Fully interactive query process	Give the user the option to specify query parameters and then immediately see the effects on the query results
		Give the user the option modify the drawing and see the effects on the query results
		Present the user with a higher understanding of how the query system process their visual query images
	Automatic query parameters	Have the retrieval system automatically define the query parameters based on the user's actions, reducing the need for the user to use and understand these settings

10 Conclusion: The Role of Query by Drawing

The previous chapters presented a detailed analysis of the research hypotheses in this study. In this final chapter these results are discussed with regards to how they relate to current CBIR systems, the quality of the results and how the results can be used as a basis for further CBIR research.

Section 10.1 presents an overview of the central research hypotheses. The research questions are discussed and answered.

Section 10.2 discusses the differences between how the respondents in this study *preferred* to draw the visual query images, and how these images *should* be drawn in order to achieve optimal results with current CBIR technology.

Section 10.3 discusses Query by Drawing in light of the different image information needs identified in section 2.4, with a primary focus on identifying request types where current QBD CBIR systems prove useful for real-world users *despite* the current technological limitations

Section 10.4 suggests four steps than be taken in order to elevate current QBD CBIR systems from an interesting research area into a tool that can be of real benefit for users when performing image retrieval tasks.

Section 10.5 summarizes the experience of using the framework for Visual Query Image Classification described in chapter 4.

Section 10.6 discusses the quality of the results made in the study. Two problematic areas are discussed, focusing on which steps have been taken to ensure that these results are both valid and generalizable.

Section 10.7 presents how the results in this study can be used as a basis for further research. Four extensions to this study are presented and discussed.

The final section (10.8) presents some concluding remarks with regards to how this work contributes to research in Content Based Image Retrieval.

10.1 Answering the Research Questions

This study was focused on five research questions relating to different aspects of the QBD CBIR process:

- **RQ1:** How do users utilize the visual query interface when they draw visual queries?
- **RQ2:** How realistic are the query images drawn by QBD CBIR users?
- **RQ3:** What are the major challenges encountered when users draw visual queries?
- **RQ4:** How do users feel about expressing image requests by drawing visual queries?
- **RQ5:** What improvements can be made to CBIR systems in order to better support users when drawing visual queries?

These research questions were examined in detail in by evaluating the research hypotheses detailed in the previous five chapters. Each of these questions are answered and discussed in the following sections. An overview of the evaluation of the research hypotheses are available in Appendix 3 - Research Questions and Hypotheses, and summarized in Table 63 (page 282).

10.1.1 Interface use: Evaluating the QBD process

The first research question addressed *how* users behave when expressing image requests by drawing visual queries. Four aspects of the respondents interface usage were identified:

1. Drawing visual queries is a time consuming process
2. The respondents preferred using freehand drawing
3. The respondents preferred the drawing interface offered by VISI
4. The respondents preferred the dynamic query process offered by Retrievr

10.1.1.1 A time Consuming Process

The first result is that QBD was a time-consuming process, particularly compared to expressing image requests as text. The empirical data suggested three major reasons for this:

- The drawing process itself takes time
- A lack of drawing experience increases the required time
- Query by Drawing represents a new approach to image retrieval

First of all, the physical actions required by users expressing image requests by drawing query images take more time than expressing these using texts. This appears to be particularly true for *generic* requests. For example, requesting images containing “dolphins” in text simply requires that the user types the word “dolphin”. Expressing this by drawing requires that the user spends time considering what a dolphin *looks like*, determining if the query image should contain any *other* elements than the dolphin (e.g. contextual elements or background), considering the spatial characteristics of the image, and finally creating a drawing they feel is representative of this request. However, several respondents noted that as the request’s level of detail increases, the *difference* in *time required* decreases. Expressing more complex requests through text may become increasingly difficult, while the increased level of detail may help the user determine the actual perceptual characteristics of the query image.

Next, some of the respondents were not very used to drawing, and several had not made any attempts at drawing in many years. For these respondents, the process of drawing was unfamiliar, and they had to spend considerable time in order to get comfortable with it. Respondents who either had a high level of drawing competency or were used to work with images on a daily basis were able to draw the query images much faster than the other respondents. This indicates that the time required to draw query images may decrease as users become more accustomed to drawing.

Finally, the QBD process represented a new approach to querying for images, and most of the respondents reported that they were unfamiliar with the process. They were not able to determine *how* they should create the query images in order to get the best possible results, what *level of detail* they should use or how they should compose the query images. This caused an increase in the time spent during the drawing process. However, most respondents stated that *increased familiarity with the QBD process* would reduce the time they needed to draw the query images.

10.1.1.2 Freehand Drawing is the Preferred Drawing Technique

The second result is that the respondents preferred to use *freehand drawing* as their main method of drawing. Three major reasons for this were identified:

- The respondents wanted to keep the process *simple*
- The *drawing tools* provided were not suitable for the retrieval tasks
- Freehand is the most *natural* way to draw

The primary reason the respondents gave for this was that they wanted to keep the drawing process as simple as possible. Several respondents compared freehand drawing to grabbing a pencil and quickly sketching something on a piece of paper, representing a very simple action. While there were indications that some respondents were able to express simple queries quickly using drawing tools, it was significantly faster to draw the query images using freehand than to use a combination of freehand drawing and drawing tools.

All respondents agreed that most of the drawing tools were not very well suited for expressing the image requests defined in the project. They felt that representing real-world objects using basic geometric shapes was difficult and resulted in query images with a very low resemblance to the objects they represented. Accordingly, basic geometric shapes may not be well-suited when working with real-life objects.

Finally, the respondents claimed that using freehand drawing represented the most natural and "fun" way of drawing. This indicates that unless the drawing tools have an affordance that encourages users to use these, freehand drawing may be the preferred approach when new users start using QBD CCBIR systems.

10.1.1.3 The Respondents Prefer the Interface Provided by VISI

The third observation was that the respondents preferred the *drawing tools*, *the canvas* and the *query parameter specification* represented by the VISI system compared to the options represented by Retrievr.

The combination of canvas size and pen tools in VISI provided the respondents with the ability to be very precise and detailed when drawing the visual query images. The relative size of the canvas and pen sizes provided by Retrievr did not give the respondents similar freedom, and often resulted in query images with a lower level of detail than desired by the respondents. Next, while the respondents felt that the tools available in VISI were unsuited for image retrieval purposes, they preferred having access to these tools. This was particularly true for some of the basic geometric shapes (e.g. circles and squares) and the *fill* tool. They felt that these tools represented useful additions in particular situations, such as when creating a background or adding several identical elements to the query image. This indicates that a QBD system should provide users with a combination of canvas size and drawing tools that allow them to express the queries *at a level of detail of their own choosing*.

The respondents also preferred having the option to specify the query parameters, i.e. how much weight the retrieval system should put on the colours, shapes, textures and the spatial arrangement of the query image. While several respondents stated that they had problems *understanding* exactly how these parameters influenced their queries, they also were frustrated with Retrievr's high reliance on colours. This indicates that a QBD system should include a technique for letting the users' specify *which* perceptual structures should be prioritized during the retrieval process.

10.1.1.4 The Respondents Prefer a Dynamic Query Process

The final observation was that the users preferred the dynamic query process provided by Retrievr, which continuously returned results while the query image was drawn. The respondents felt they required *less time* when expressing these queries, as they were less uncertain about the level of detail in the query image and how the query image should be composed. The dynamic result presentation allowed them to add as many details as required in order to obtain good results. This indicates that a dynamic process may help users draw visual queries in a *more efficient manner*.

The respondents also claimed that the process helped them to *understand* how the system processed their query image. The immediate feedback helped them to understand how their actions caused changes in the result set. The respondent claimed that this transparency increased understanding and helped them to create *better queries*.

Finally, the respondents were also happy to be able to modify and adapt the queries based on this feedback. This allowed them to adapt their queries based on this feedback, providing them with access to more images without having to *initiate new queries*.

10.1.2 Evaluation of Query Image Modality

The second research question addressed the *realism* of the visual queries: The *visual modality* of the query images drawn by the respondents. The evaluation showed that the visual modality of the query images was *very low*, particularly with regards to the *representational* modality. Two primary causes for this were found: The *ability* and the *willingness* to draw query images with a high visual modality.

10.1.2.1 The Respondents' Ability to Draw Realistic Images

Two influences on the *ability* to draw realistic images were identified: The respondent's *drawing competencies* and *properties of the two drawing interfaces*. For the respondents without a visual background and those with low drawing skills, the primary reason stated was a lack of drawing competency; the respondents were simply not able to create realistic representations of the query objects. Additionally, some of the respondents who *felt* that they were able to create realistic drawings felt that they were *limited* in this by the drawing interfaces: The choice of colours, the selection of drawing tools and the relative size of the canvas and the drawing tools.

10.1.2.2 The Respondents' Willingness to Draw Realistic Images

Three influences on the respondents' *willingness* to create realistic query images were identified: they wanted to keep the query process *simple,* they adapted the visual modality based on *the nature of the retrieval task,* and some respondents drew images with a low visual modality because they were *inexperienced with the QBD process*.

Some of the respondents who rated themselves as skilled drawers stated that while they felt they were *capable* of creating queries with a high visual modality, they were not willing to spend time on this. They wanted to *keep the query process as simple as possible*.

Next, the *nature of the image retrieval task* influenced the visual modality of some query images. Some respondents stated that they made conscious choices to keep the level of detail low. This was primarily the case when the retrieval tasks were of a *generic* nature, e.g. when querying for generic objects without regard to the context of the objects. In these cases the respondents refrained from adding contextual details and background, they did not use more colours than absolutely necessary to describe the subject matter of the query and they did not consider the spatial composition of the query image. This was done either in order *to not exclude relevant images* or *to not include irrelevant images* based on colours, background, details or contextual elements. For such simple queries, these respondents treated the visual query images in a manner very similar to expressing simple text based queries, e.g. rather than typing the word "dolphin" they expressed the request as a simple drawing of a dolphin in the middle of a white canvas. However, it should be noted that even if the visual modality of the query images were higher for other retrieval tasks, the visual modality of *all* images was very low.

Finally, the respondents were *unfamiliar* with how the retrieval system worked. Several respondents stated that they would have created queries with a higher visual modality, particularly with regards to contextualization and background, if they had a higher understanding of the retrieval reliance on perceptual structures.

These results indicate that one cannot expect that users will create visual query images with a high visual modality. However, it is possible that users can be *encouraged* to increase the visual modality of the query images, particularly with regards to the use of *colours*, *context* and *composition*. Respondents who *understood* how the CBIR system processed their images created images with a somewhat higher visual modality than the other respondents. But based on the response from the respondents who were competent drawers, it is not reasonable to expect that users will create images with a high *representational* modality, e.g. query images where the objects are depicted with a high degree of realism.

10.1.3 Major Challenges Facing the Query by Drawing Process

The third research question addressed the *major challenges* the users experienced when drawing visual query images. Four potential challenges were identified and discussed:

- Problems related to *drawing*
- The *time required* by the QBD process
- The "*Page-Zero*" problem
- Problems expressing *narrative content*

The two respondent groups stated that they experienced the nature and severity of these challenges differently. The problems related to the *drawing process* and the *time required* were primarily experienced by the respondents without a visual background, while the *page-zero* problem and the problems expressing *narrative content* were experienced by both groups. This confirms that users *without* a visual background experience *more* challenges when using QBD CBIR systems than users *with* a visual background.

10.1.3.1 Problems Related to the Drawing Process

The respondents without a visual background reported that the main challenges were difficulties related to creating *realistic representations* of the objects they were drawing. Most of these respondents were very dissatisfied with their own ability to create representations with a high level of visual modality. The respondents generally attributed this to their "low drawing skills", and believed that they would be able to overcome this challenge if they were "better skilled in drawing". However, there were only small differences in visual modality between their query images and the query images created by respondents with a visual background.

The respondents *with* a visual background were aware that their query images held a low visual query modality, but reported that they did not consider this to be a challenge. Creating query images with a low visual modality was often a deliberate choice on their behalf, they were not *willing* to spend the time and effort required to create query images with a high visual modality even though they were *capable* of doing this.

It is possible that the *real* challenge here may be that the respondents without drawing experience and drawing skills *feel* that they are unable to create realistic representations, and not the fact that they are *incapable* of creating realistic representations.

10.1.3.2 Problems Related to Time

Some of the respondents without a visual background stated that the time required to draw query images represented a significant challenge, particularly compared to the time required to express the same queries as text. None of the respondents with a visual background reported this as a challenge. Two explanations for this difference were found: *The drawing competency* of the respondents, and *their real-world relationship with image retrieval*.

The respondents with a visual background used significantly less time on drawing the query images than the other respondents, i.e. the impact of the time spent expressing the queries was lower for this group. The respondents without a visual background used more time in the drawing process. They needed more time to do the actual drawing and they were unfamiliar with the drawing tools.

The respondents with a visual background could identify situations where they potentially could spend *less* time on their retrieval tasks by drawing visual queries than expressing the requests through text. Accordingly, they did not consider *time* as a significant challenge.

10.1.3.3 The "Page-Zero" problem

A majority of the respondents stated that they had some problems to decide *how* they should create the queries: *What* objects should they include, *how* they should draw the objects and *where* the objects should be located in the query image. This was primarily true when initiating a *new* request type, i.e. the *first* time they expressed queries for a particular type of content.

The respondents without a visual background were more influenced by this than the other respondent group, primarily because they experienced more problems related to expressing the actual objects. However, most of the respondents with a visual background also mentioned this as a challenge.

This challenge was primarily experienced when the *nature* of the retrieval task was *low in details*, e.g. when creating queries based on *generic contents*. When the detail of the image request increased, e.g. when requesting images of complete scenes or images with very specific content, the respondents were able to start drawing *sooner* (i.e. they had to spend less time considering *what* to include in the query), and they were able to add *more details* to the query images.

This indicates that expressing queries based on image requests with a *low level of detail* may cause some problems related to *initiating* the drawing process, indicating that QBD may more useful and easier when users have a very clear mental image of the *composition* and *contents* of the images they are attempting to retrieve, and *less* useful when they are trying to retrieve images based on *generic* requests.

10.1.3.4 Problems Expressing Narrative Content

Almost all respondents reported that expressing queries for *narrative contents* represented a significant challenge. This was primarily caused by difficulties relating to *how* this type of content should be described. Images represent a snapshot of time, and while the human mind is capable of *interpreting* these snapshots and *understanding* narrative structures, *creating* such narrative structures and using these in a visual query appeared to be very difficult.

This did not represent a *significant* challenge to the respondents, e.g. it did not seem to have any influence on their *opinions* and *attitudes* towards the QBD process, and most of the respondents stated that they found it unlikely that they would use QBD as a primary tool when expressing these types of image requests.

This indicates that current QBD CBR systems are not well-suited for image retrieval tasks where the *narrative content* of the queries is important, as creating visual queries for this by drawing may prove too difficult for most users. Even users who rated themselves as very competent drawers experienced problems related to this.

10.1.4 Respondents Attitudes and Opinions towards Visual Query by Drawing

The fourth research question addressed user's opinions about drawing visual query images. Three different aspects of this were evaluated: Their *attitudes* towards QBD, their *willingness* to use it and *the potential uses* they saw for it. Five major influences on these elements were identified during this work:

1. The respondents' *image needs* and their relationship to *image retrieval*
2. How the respondents experienced the *challenges* related to the QBD process
3. The *quality of the results* obtained by using QBD
4. The *expressive convenience* of the QBD interfaces
5. The *size of the image collection* accessible through QBD

10.1.4.1 The Respondent's Relationship to Image Retrieval

The respondents' *real-world needs* and their *relationship to image retrieval* seemed to be the most significant influence on their opinions towards query by drawing. Their opinions and attitudes depended on whether the respondents were able to identify real-life situations where they might *benefit* from using visual queries, or situations where they previously had experienced *problems* expressing image requests using traditional techniques. Respondents who were able to identify such situations were *more positive* towards drawing visual queries, they were *more willing to use* this approach, and they were better able to identify relevant situations where they might *benefit* from using visual queries. It should also be noted that respondents *without* clearly defined real-world needs generally agreed that they found the QBD approach fun, and might be willing to use this for fun.

10.1.4.2 The Challenges of the Query by Drawing Process

The *challenges* experienced during the QBD process represented a major influence on the respondents' opinions. The respondents who experienced most challenges were also the least positive towards the QBD process. This was particularly true for the challenges related to the *drawing process*. Several of the respondents stated that these challenges represented the largest obstacle against them using QBD CBIR systems. They did not feel that the potential benefits QBD might give them would outweigh these challenges.

10.1.4.3 Query by Drawing Result Quality

The *quality* of the results returned from the QBD process had a strong influence on the respondents' attitudes towards QBD CBIR systems. Again, this was most prominent for the respondents without a visual background, and may be related to their real-world needs. The respondents *with* a visual background were often able to identify potential uses that went beyond retrieval of images based on their semantic contents, e.g. retrieving images based on the perceptual structures. The other respondents were very concerned that the results they got back held a low *semantic* similarity to the query images they drew. These respondents stated that that the low degree of semantic similarity discouraged them from using the QBD approach in the future.

10.1.4.4 Query by Drawing and Expressive Convenience

The *expressive convenience* provided by the QBD approach influenced the respondents' attitudes towards QBD, particularly when compared to text based queries. This was also particularly true for the respondents without a visual background. Again, a likely explanation for this was that these respondents often were unable to identify situations where text based queries might be difficult to use. The respondents *with* a visual background also reported that they considered the expressive convenience to be a problem *when expressing simple queries* (, but that for *more complex tasks* the expressive convenience of the QBD might be *higher* than that of text based queries.

10.1.4.5 Image Collection Size

Finally, the respondents who were positive towards QBD held a major reservation: They were worried about the *size of the image collections*. They stated that their positive opinions were dependent on the size of the collections. The low number of images in the collections and the domain of the images (in VISI) represented a limitation to the usefulness of the approach. Several stated that if the QBD approach could be used on large scale image collections or search engines such as *Google Images*, they would start using it immediately, and would most likely benefit from using it.

This indicates that *users working with images at a professional level* may be *very positive* towards expressing *some types* of image requests by drawing visual queries, and *very willing* to use the approach. And, based on the results obtained in this study, these users see potential benefits and uses from the QBD approach *in its current state*. Even if these systems are currently unable to identify and retrieve semantic contents, the systems may prove useful *given* that they contain a *sufficiently large collection of images*.

10.1.5 Respondents Suggestions for QBD Improvements

The respondents identified and discussed a large number of potential improvements that they believed would improve their experience of the QBD CBIR process[105]. Most of these suggestions could be classified into five categories:

1. **Deformable shapes**. The option to manipulate objects once they were added to the canvas through modification, grouping, layering, and repositioning.
2. **Shape templates.** The option to select domain-specific shapes directly in the interface, reducing the need to create these shapes through drawing.
3. **Using icons and pictograms**. The option to use icons or pictograms for representing important query objectives and for indicating narrative structures in a query image.
4. **Colour-neutral drawing tools**. The option to specify shape and spatial relationships without the retrieval system focusing directly on the chosen colour.
5. **Dynamic query process**. Dynamic and continuous presentation of query result and automatic definition of query parameters**.**

The common denominators for these suggestions is a desire to *reduce the problems related to drawing skills*, *reducing the complexity* and *the time required* by the query process, and *increasing the user's understanding of the query process*. Section 10.4 discusses how some of these suggestions can be used to improve QBD CBIR systems within the scope of currently available technology. Section 10.7 describes future work and research projects that build upon the other suggestions.

10.2 Visual Query by Drawing and Current CBIR Systems

The results presented in the previous sections present some interesting observations on the way current CBIR systems process drawn query images. As noted in chapter 2.3, most basic CBIR systems work by extracting feature vectors from digital images, comparing these using similarity functions, and returning images that have similar perceptual structures, as defined by one or more similarity requirements (e.g. colours, textures, shapes or the local and global spatial distribution of these). In order for these systems to successfully process drawn query images, they *require* that the query images share at least *some* perceptual structures with relevant query images. However, the analysis of drawn query image modality described in chapter 6 showed that the visual modality of the query images was very low. The query images held a *low degree of contextualization*, they were created using *few colours*, the query objects held a *high degree of abstraction* and the query images held a

[105] A summary of the suggestions are presented in Table 59, page 234.

very low *compositional modality*, e.g. the objects in the query images were *not* given a realistic scaling and the respondents generally placed the objects in the middle of the query image.

Query 72 (Figure 117) represents the “average” query image created in this project: A single object represented monochromatically with relatively high degree of abstraction, placed near the middle of the query image without contextual details or background. The respondent created the query in a similar manner to a textual keyword: A basic representation of the subject matter of the query.

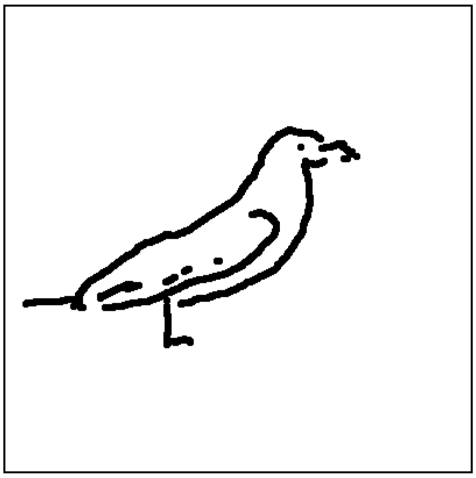


Figure 117 - An illustration of an "average" query image created in this project (Query 72)

While this image might be usable when querying a collection of homogeneous schematic drawings, it is unlikely that the query would return very relevant results when querying a large scale collection of heterogeneous images from a broad domain. Three potential problems are identified: *The low use of colours*, *the low degree of contextualization* and, in some cases, *the size and the placement of the seagull*.

Primarily, similarity comparison based on colours currently represents the best similarity based approach. Extracting colours and comparing images based on their colour distribution is relatively straight forward, and current CBIR systems are capable of achieving good and accurate comparisons based on both local and global colour distributions. The respondents in this study often used “black” and “white” as neutral colours, and expressed generic requests by drawing black lines on a white canvas. Combined with the low use of query parameters this resulted in retrieval of black and white drawings without any semantic resemblance to the query image. Even if users use query parameters to reduce the weight of colours, the retrieval system must rely on other similarity criteria (e.g. shape or texture). This introduces a number of other challenges related to segmenting and identifying the various elements of the image.

Furthermore, the low degree of contextualization represents a similar challenge. While some of the motivations for creating queries without contextualization may be logical to the user (e.g. not wanting to limit the query results or "confusing" the system by introducing elements irrelevant to the subject matter of the query), this is counterproductive to the way current similarity based retrieval systems work. Unless the user is requesting images containing *only* the query object present (e.g. when requesting images similar to the image represented in Figure 118), it is very unlikely that current CBIR systems will be able to retrieve images containing the requested objects.

Figure 118 - An image containing a clearly defined seagull[106].

Finally, the respondents often used a strategy of drawing large objects in the centre of the query image, particularly if they did not have strong opinions on the size and placement of these objects in the desired result set. While this may seem like a good idea to the user, this strategy will, with current systems, result in retrieval of simple images with the objects placed at a similar location to objects in the query image. This can potentially exclude relevant images where the objects are smaller, or are placed in other locations.

It should be noted that the modality aspect the respondents found most problematic, the *representational modality*, is probably the *least* problematic issue for current CBIR systems. Current retrieval systems are generally not very good at segmenting and identifying semantically meaningful objects in an image. Even if the respondents were able to create representations with very high representational modality, it is unlikely that a CBIR system would be able to benefit from this unless comparing the query image to a collection of simple, schematic images where the objects are clearly distinguished from the background.

Consequently, queries similar to query 15 (Figure 119A) may be *more* effective than queries similar to query 21 (Figure 119B), even if query 21 might have a (slightly) higher visual modality, since, in

[106] The image was retrieved from http://www.flickr.com/photos/bigd2112/3649908068/. Some rights reserved.

some regards it contains *more* usable perceptual structures: The colour blue may be used to retrieve images primarily depicting a blue sea and the three distinct shapes which may be used to identify distinct shapes in a similar location.

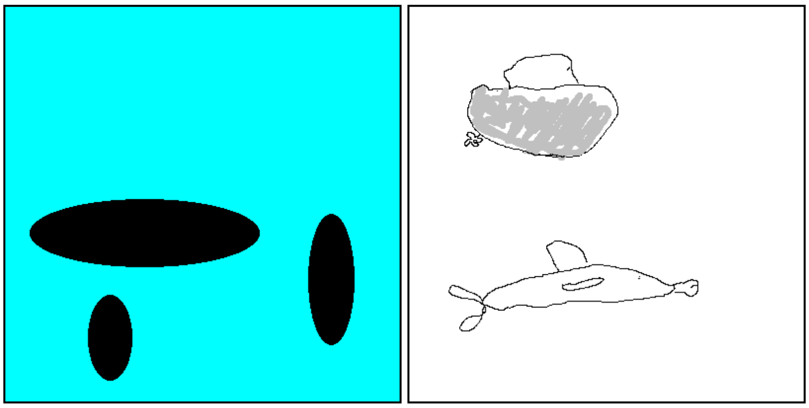


Figure 119 - Two queries representing requests for "dolphins and boats" (Queries 15 and 21)

It should also be noted that the visual modality of the query images increased as the level of detail of the image request increased, particularly requests for *complete scenes*. In these cases, the queries contained *a higher degree of contextualization* and *included more background*, they had a *more active colour use* and there were a more active use of *compositional measures*.

These results confirm that there is a gap between how users *prefer* to draw visual query images, and the way these queries should be expressed in order to obtain *the best possible results* using *current CBIR systems*. This gap is defined as the *query specification gap: The difference between the way users prefer to draw visual query images and the way these images should be drawn in order to be optimal for current CBIR systems (Definition 16)*. The *query specification gap* is *largest* when users create queries based on requests for generic image contents, and becomes *smaller* as the *level of detail* of the image requests increases.

10.3 Query by Drawing and User Retrieval Tasks

Section 2.3 presented an overview of different types of *user image requests*, and the discussion in section 2.5.6 presented a general discussion on the possibility of using *query by drawing* for the different user requests. While the retrieval tasks used in this study were primarily requests for *generic content*, *narrative content* and *complete scenes*, the resulting data can be used to present some general observations on all these levels. The findings are summarized in Table 60 (page252).

Concerning the use of QBD for non-visual content (e.g. metadata), no additional data was collected in this project.

Requests for *specific images* were not directly evaluated in this project. However, it was observed that the *completeness* and *modality* of the visual query images increased with the level of detail in the image request. It is likely that if users have a clear mental image of the specific image they are requesting, particularly with regards to the *colours* and the *colour distribution*, they might be able to create a relatively detailed query image representative of the image they are requesting. Furthermore, some of the respondents in the KHIB group specifically mentioned using QBD as a tool when querying for specific images. While this needs to be confirmed by empirical studies, it is considered as *likely* that users would find the QBD approach very useful when requesting specific images, *given* that they have a clear mental image of the image they are requesting.

Two observations were made concerning r*equests for perceptual structures*. First of all, contrary to the findings reported in section 2.3, the respondents in the KHIB group stated that this is a form of image request that has a high priority for them. While the size of the KHIB group makes it difficult to make generalizations, this observation indicates that requests for perceptual structures may be more important for *certain user groups* than what is suggested in the literature reviewed in chapter 2.4. This supports the assertion made by Enser and Sandom (2003) that this type of queries may be very relevant for users working professionally with images. Furthermore, a majority of the respondents in the KHIB group and some of the respondents in the IFIM group reported that they considered the QBD approach ideal for requests for perceptual structures. They had previously looked for an alternative to text based queries when expressing this type of request. Furthermore, current CBIR techniques may achieve good results for this type of request, particularly if the request involves colours.

Requests for *generic content* represented two of the three types of retrieval tasks used in the project. For *generic objects*, the empirical data indicates that the respondents' preferred way of expressing these queries presents several challenges for QBD. The preferred strategy was to draw the queries as *visual keywords*. This approach that is *not* optimal for current CBIR systems. The respondents' tendency to keep the query images as simple as possible is in direct conflict with the needs of current similarity functions. Combined with the desires to avoid *confusing the system* and to ensure *relevant results*, the feasibility of using QBD for this request type seems to be low. Furthermore, most of the respondents stated that for requests for generic objects and similar very simple searches, the expressive convenience of text based queries surpass the expressive convenience of the QBD approach. They found it unlikely that they would prefer using QBD over text based queries. Successfully using QBD for this type of query requires that users are encouraged to draw query images with a higher visual modality, but this directly conflicts with the aforementioned desires. Consequently, it is difficult to imagine that using QBD in combination with current CBIR systems will represent a usable way of expressing these image requests.

However, for more complex generic requests, e.g. requests for *complete scenes* or other generic requests where the *spatial structure* is important, it appears as if the QBD approach may be more useful. First, the increased visual modality of these query images indicates that users are willing create more detailed visual query images that can be more useful for current CBIR systems. Next, attitudes reported by the respondents, and particularly the respondents in the KHIB group, indicates that they may be positive towards using QBD for this type of request. As the complexity of the image request increases, the expressive convenience of QBD increases and may even surpass the expressive convenience of text based queries. Several respondents also noted that if they had the option to *combine* QBD with text based queries, they might have a very powerful tool for expressing this form of request. Consequently, the findings in this project indicate that QBD may be very useful when expressing queries based on such detailed requests, *particularly* for users working visually. Further studies combining QBD with a text based approach on a large scale data collection is required in order to verify these findings.

Next, while *specific* requests were not directly evaluated in this project, the results present some general observations. First of all, it is unlikely that QBD alone may be feasible when expressing requests for *specific individuals* or *specific classes of generic contents* (e.g. a particular species of fish or a particular type of boat). The very low representational modality found in the queries makes it very unlikely that current CBIR systems would be able to identify and process this type of request: If the respondents find it difficult to indicate that a certain object is a *human*, it is unlikely that they would be able to indicate that it is a *specific, named individual.* However, as noted in section 2.5.6, it is possible that the approach may be successfully used if requesting images containing specific landmarks or objects with very distinct perceptual structures. Further studies explicitly focusing on this type of retrieval tasks are required in order to verify this.

Requests for narrative content represent the third request type evaluated in this study. Based on the empirical data, it seems unlikely that the QBD approach alone may represent a viable way of expressing these queries. All respondents reported problems related to expressing this type of content, and in some cases the modality of these queries were *lower* than the modality of other queries. The respondents used different strategies to express this type of content, and some of these strategies were in direct opposition to current CBIR systems' need for high modality, e.g. using *value scaling* or introducing icons, pictograms or structures from cartoons in the query images. Consequently, it is unlikely that QBD alone may represent a very convenient and usable method for expressing this type of request. However, it should be noted that several respondents indicated that if QBD could be *combined* with text based queries, they might be able to achieve a very high level of recall when expressing this type of request.

Finally, requests for *abstract* content were not directly evaluated in this study. However, several of the respondents in the KHIB group, particularly the two respondents with a background in fine arts, stated that they thought that the QBD approach might be very useful, particularly if the requests were based on perceptual structures. In these cases, these requests are more or less identical to requests based on perceptual structures. Further studies are required in order to evaluate the usefulness of QBD with regards to abstract content.

These results indicate that QBD based on current CBIR systems *may* represent a very powerful tool for *some* users expressing *some types* of image requests. It is unlikely that current systems and solutions will prove very useful for most generic retrieval tasks, but as the complexity of the request increases, the expressive convenience of the QBD approach increases and may surpass the expressive convenience of text based queries. These results indicate that this may be particularly true for requests where the *perceptual* and *spatial* structure of the requests is important. Furthermore, if users can combine the QBD approach with text based queries, it is possible that this may provide the users with a very powerful tool for expressing queries, even with current CBIR limitations.

Table 60 presents a summary of the above results with regards to the 7 levels of image requests. *Evaluation* describes whether the request type was directly evaluated using retrieval tasks or indirectly through discussions with the respondents. *QBD applicability* describes how applicable QBD has been found to be for the request type. *Comments* summarize the main findings related to the request level.

Table 60 - Summary of QBD and Image Retrieval tasks

Request level		Evaluation	QBD applicability	Comments
1	Non-visual	Not evaluated	Not evaluated	Not evaluated
2	Specific image	Indirect	Potentially high	The completeness and the visual modality in the query images indicate that these queries may achieve good CBIR results
3	Perceptual structures	Direct	Very high	The importance of this request type may be underestimated in literature The professional respondents believed that using QBD for these requests may be very useful

4	Generic requests	Direct	Dependent on request complexity	For simple requests the preferred expressing method for these queries (i.e. "Visual keywords) is not optimal for CBIR retrieval Most respondents were unwilling to use QBD for these Willingness to use and CBIR applicability increases as the complexity of the request increases (e.g. when requesting complete scenes) Willingness to use and CBIR applicability may increase if combined with text based queries
5	Specific requests	Indirect	Low	Low visual modality in the query images indicate that QBD applicability for these requests are low Might be useful for requests after objects with distinct perceptual characteristics
6	Requests for narrative content	Direct	Low	Low visual modality in the query images indicate that QBD applicability for these requests are low Respondents experienced major difficulties when expressing these queries
7	Requests for abstract content	Indirect	Potentially high	May be very useful when the request is based on perceptual structures, e.g. colours or shapes

10.4 Improving QBD CBIR Systems

The results obtained in this study indicate that there *is* a need for image retrieval systems that can allow users to express *certain types* of image requests using the QBD approach, as the QBD approach has the potential to provide a higher degree of expressive convenience than the text based approach. The final research question was directed at identifying how QBD SYSTEMS can be improved with respect to the users' *experience* when expressing image requests by drawing visual queries, as well as these systems' *ability to process* these images.

Several different suggestions related to this were identified in the analysis described in chapters 5 through 9. These can be classified in two different categories: Suggestions that can be directly utilized with current CBIR technology, and suggestions that require fundamental changes in the way CBIR systems index digital images and process visual queries expressed through drawing. Using the former suggestions to improve current QBD CBIR systems are discussed here. The latter suggestions are used section 10.7 to discuss future research projects based on this work.

This section presents four steps that can be taken in order to elevate current QBD CBIR systems from an interesting research area to tools that can be used by real users to solve real image retrieval tasks. It is believed that this need may be met by solutions based on *current CBIR technology*. We need to ensure that current CBIR systems allow users with *relevant retrieval tasks* to express these tasks by drawing visual queries as efficiently as possible, and we need to ensure that users who feel *inadequately skilled* to understand that their drawing skills might not represent a real challenge. Four steps are required in order to let current CBIR systems become as useful as possible:

1. The drawing interfaces must be made as usable as possible with regards to *drawing tools* and *dynamic query processing*
2. The user must be *encouraged* to express the query images in a manner that the retrieval system is capable of processing
3. CBIR technology *must* be combined with text based retrieval techniques
4. The systems must be able to work on large scale image collections

The first step is required in order to provide the user with the highest possible expressive convenience. Several suggestions have been identified during this study, as described in chapter 9. Three important suggestions are highlighted here: The users should be provided with a *sufficiently large canvas* and *a suitable set of drawing tools*. Users, and particularly users who work professionally with image applications, are used to have powerful drawing tools with a high usability. The drawing tools provided by a QBD interface should provide these users with enough power and flexibility to let them draw the query images as efficiently as possible, while at the same time keeping the query process as *simple as possible*. Some of the suggestions presented in chapter 9 present some general observations towards this, e.g. including deformable shapes, shape templates, using a dynamic query process and using icons or pictograms for expressing narrative requests. Some further research is required in order to determine *which* of these requests represent the most useful changes, and determine *how* these changes can be implemented in a QBD interface. This is further detailed in section 10.7.1.

The second step is required in order to encourage the users to express the visual query images in a manner that current CBIR systems are capable of processing. This includes educating the user with regards to *how* these systems work most efficiently, e.g. provide the system with sufficiently detailed perceptual structures. The observations made in this project suggest that this can be partially solved by having a dynamic result presentation: Users are capable of adapting their approach based on the feedback they receive, and draw in a way that the system is capable of processing as long as this does not directly conflict with the user's need to keep the process as simple as possible. This step might also reduce the major challenge experienced by users who have low drawing competencies. However, as illustrated in section 10.2, there is little reason to believe that users with a low level of drawing competency may be unable to express query images that CBIR systems can process. QBD CBIR interfaces must be designed in a way that *encourages* these users to *look past* their low drawing competencies and create images that the CBIR system is capable of processing.

The third step is combining the QBD approach with a text based approach. Current CBIR systems suffer from problems related to the *semantic gap:* The systems are not capable of processing queries at a semantic level. Combining QBD with text based queries will allow the user to express multimodal queries that have a significantly higher expressive power than the two query techniques alone, without any reduction of the expressive convenience of QBD. It is unlikely that giving the users the option to specify a simple linguistic query in addition to drawing would conflict with the users' desire to keep the query process simple. And, as noted by the respondents, adding textual keywords to their queries could reduce the *size* of the result sets, allowing the CBIR systems to focus on the perceptual structures while the text based queries would remove images with a low *semantic* similarity to their query tasks.

While combining QBD with text based queries would not solve the problems of *volume* or *subjectivity*, the respondents in this study stated that unless they had *very* specific image requests, they were not particularly bothered by these problems. When querying a sufficiently large image collection (e.g. using Google Images), there would almost *always* be images relevant to their query task in the result set; the problem was primarily that the result set was so large that actually *finding* these images represented the largest challenge. Introducing QBD based CBIR searches may provide a very powerful tool for increasing the *precision* of their query results.

The fourth step is making QBD CBIR systems available for large scale image collections and image web search engines. Both of the systems evaluated in this project as well as most of the systems described in section 2.5 operate on a limited collection of images. As noted by the respondents, this was one of the most important obstacles for the *usefulness* of these systems. While the respondents agreed that the QBD approach might be very useful for small scale collections (e.g. using variants of QBD to access images of information from a specialized collection), they were primarily interested in using it on large scale collections or search engines such as Google Images. Previously, the biggest challenge towards this has been the cost in storage and computational power to index and compare large scale image collections. However, with the advent of *cloud computing* this may not longer represent a significant problem. One example of this is the TinEye search index (Idée 2009), which reports that it has indexed over 1.1 billion images (Idée 2009), and is capable of performing similarity searches against these images in less than one second. Furthermore, by using text as a filtering mechanism, the number of images being processed by the similarity functions may be scaled down to a manageable level.

By following these four steps, QBD CBIR systems may be promoted from their current position as experimental prototypes to powerful tools that may be used by real world, professional users to solve retrieval tasks currently difficult to solve.

10.5 Evaluation of the Visual Query Classification Framework

The framework described in chapter 4 was developed with two primary goals in mind:

1. It should provide a set of precisely defined concepts which can be used to evaluate the hypotheses
2. It should provide a set of tools capable of evaluating these concepts

How well did the suggested framework achieve these two goals? Chapter 6 described an analysis of the query images based on the suggested framework. The results obtained through this analysis allowed for a detailed description of query images created by the respondents, particularly with regards to the degree of realism in these images. This description also provided important support and aid when performing the other evaluations described in this project. Consequently, the framework was very important for the project and achieved the two major goals. However, there were some issues with the different modality markers. The following sections briefly discuss the different modality markers as well as an overall evaluation of the usefulness of the framework.

10.5.1 *Contextualization and Query Image Completeness*

The *contextualisation* marker proved to be very useful with regards to analyzing how *complete* the query images were. The only challenge related to this marker was the two different types of *contextual elements* (*Symbolic* and *minor*). As noted in section 6.1, it was difficult to distinguish between these two modality criteria. Accordingly, the two elements were combined into a single criterion. While this might have caused some loss in detail with regards to discriminating between different types of contextual elements, the overall use of such elements were so low that it is not believed that this has had a significant impact on the results.

Counting the number of individual objects in the query image also proved to be a useful tool for evaluation of query image completeness. All three evaluators experienced some minor problems relating to determining if some elements should be counted as one or two objects, but this was only the case in a low number of query images. Accordingly, it is not believed that these problems had a significant impact on the results: The mean number of objects in a query image was 2.10 (3.79 for level 3 queries), and the inclusion of one extra object in a few images would not have caused a large change in this.

10.5.2 *Use of Colours*

The *colour* marked also represented a useful tool when evaluating the query images, particularly with regards *colour neutrality* and the fact that the respondents felt that using colours complicated the query process.

However, there were some issues with the modality criteria. As noted in chapter 6.2, the *monochromatic* marker was replaced by *lack of colours*. In retrospective, it is difficult to see the usefulness of measuring whether an image was drawn in a monochromatic manner. Changing this criterion to measuring *a lack of colour* assisted the identification of the use of black lines on a white canvas as a *colour neutral* manner of expressing queries. It is possible that some of the queries classified as “Simple colour use” should have been classified as “Lack of Colours”. However, the number of potentially problematic query images was low (10 query images), and the overall effect of this would not have any significant impact on the overall results.

The two last modality criteria (*use of colour gradients* and *use of illumination*) were not found in any of the query images, indicating that these criteria might not be very useful as tools for discriminating between different query images. However, including these helped indicate that the respondents in this study did not use colour gradients or illumination effects when drawing the query images.

10.5.3 Representation and Degree of Abstraction

The *representational* modality criteria proved useful when determining the degree of realism of the query images. It was particularly helpful in identifying the increased abstraction seen in the more complex query images. However, it is believed that some changes to this marker would have allowed for better analysis of representational modality.

First of all, as noted in section 6.3, it was sometimes difficult to determine whether some objects should be classified as *geometric primitives* or *outlines*. Use of a combination of geometric primitives to represent objects (e.g. “Straw figures”) was classified as “Geometric primitives”. In retrospective, this was probably not a very good idea. A higher accuracy in the results may have been achieved if these elements could have been classified as “Icons” or “Pictograms”. By doing this, “Geometric primitives” would have been a more “pure” category, allowing for more distinct categorization of the query images.

Similarly, the “Outline” criterion may have been too inclusive. There were large differences between different objects classified as *outlines*, illustrated in Figure 120. When classifying these images, all evaluators classified these sharks as represented using “Outlines”. However, there is clearly a difference in the techniques and the way the two respondents have drawn these outlines. Figure 120a was drawn in a single stroke using a thick line, while Figure 120a was drawn using multiple short lines, more focused on creating a "realistic" outline. Including a method for discriminating between different types of outlines would have allowed for a better discrimination of such images. However, it should be noted that in these cases there was a difference in *personal evaluation* score of such images. This at least provided an opportunity to discriminate between the representational modality of the two images, even if the modality criteria did not allow this.

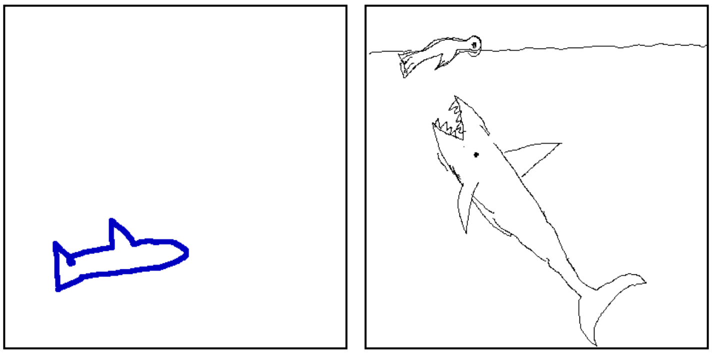


Figure 120a and b - Two different representations a shark (Queries 167 and 145. Both were classified as "Outlines".

Next, similarly to the two *contextual criteria*, the *symbolic* and *detailed representational* criteria were combined into a single criterion: *visual cues*. While this also may have caused some loss of details, the overall effect of this was low. Few representational elements were classified as *detailed*, and it is not believed that much additional detail would have been included in the results if the two criteria had been kept.

10.5.4 Composition and Compositional Structures

The compositional modality criteria proved to be very difficult to use in the context of drawn visual queries. All three evaluators experienced difficulties using these criteria, primarily related to determining if the three criteria were met. Furthermore, the low degree of completeness and contextualization of the query images resulted in a low usefulness of composition and compositional structures as modality marker for drawn query images, other than identifying that the respondents used these structures to a very low degree. The analysis of the query object placement proved returned more interesting results, indicating that this was a more useful tool for analyzing compositional elements of drawn query images.

10.5.5 Summary and Further Use of the Framework

The framework provided an invaluable tool during this project, and it allowed for qualified and justified statements concerning the query images. While the issues described above might have provided *better* results by giving *more* data about *how* the query images were created, it is not believed that this additional data would have provided very much additional insight into how QBD CBIR systems are capable of interpreting and processing these images.

Based on the experiences from using the framework in this study, it is believed that the framework can be useful for further, similar studies of drawn visual query images. However, the framework should be refined and developed further, particularly with regards to *compositional structures*.

10.6 Validity and Data Quality

The validity of these results and the quality of the data are dependent on the reliability and the validity of the methodological framework. Two elements in the study present some threats to the reliability, validity and generalizability of this study: Problems related to the *time measurements*, and problems related to the *image retrieval tasks*. These problems and the steps taken to reduce their impact are discussed in the following sections.

10.6.1 Problems Related to Time Measurements

The time spent drawing the query images was probably influenced by the methods used in the experiments. The use of a think-aloud protocol most likely caused an increase in time spent on creating queries, and the participants might have been interrupted by the questions asked by the researcher. Consequently, measurements of time are not likely to be very accurate or representative of the time these respondents would have spent in a real situation. However, while this might have influenced the exactness of these measurements, the inter-experiment measurements are most likely still valid. All queries were executed under similar conditions, and the observed differences are still likely to be valid.

10.6.2 Problems Related to the Image Retrieval Tasks

There were 5 potential problems related to the *image retrieval tasks* presented to the respondents in the 3 experiments:

1. Different retrieval tasks were used in the different experiments
2. The requests may have been less detailed for the Retrievr tasks than the VISI tasks
3. The order of complexity of tasks was the same for all respondents
4. There was an uneven number of respondents in the experiments
5. There was an uneven number of tasks performed in the two retrieval systems

The retrieval tasks given to the respondents differed between the three experiments. Some tasks were used in all experiments, while others were unique to the individual studies. Some of the retrieval tasks used in the experiments were of a very narrow domain, e.g. limited to tasks based on a maritime scenario. Several of the respondents remarked on this, particularly the respondents in the KHIB group. The maritime scenario formed the basis for the two first experiments, and accordingly most of the queries created by the respondents in the IFIM group were created based on these tasks. This difference in retrieval tasks may have introduced a structural bias between the three experiments, which may have caused some of the observed differences between the two groups.

Similarly, *all* the queries made in VISI were based on tasks from a maritime scenario, while *all* queries created in Retrievr were based on tasks from a more generic domain. The retrieval tasks used in Retrievr may have been less detailed and less specific then the requests used in VISI. For example, one of the requests for *narrative content* in VISI was "Find images of humans nursing a beached and injured whale", while a similar request in Retrievr was "Find images of humans practicing sports". It is possible that this may have introduced a structural bias in the results, e.g. that the differences in query modality observed between *respondent groups* or *interface types* may have been influenced by the differences in retrieval tasks. In order to have completely comparable results, some respondents should have expressed some of tasks based on the maritime scenario in Retrievr, and some of the more generic tasks in VISI. However, most of the observed differences were very small, and the most notable differences were between the *query types*, and this difference was observed for both sets of retrieval tasks. Consequently, while the retrieval tasks should have been more equally distributed between the interfaces with regards to their domain, the actual impact of these differences was determined to be low.

The *order* in which the image retrieval tasks were presented to the respondents may have introduced a structural bias in the data material. Most respondents were given the image requests in the same order: Type 1 (Basic semantic content), Type 2 (Narrative content) and type 3 (Complete scenes), followed by the scenario based tasks. This represents an increase in complexity in the image retrieval tasks. Ideally, the order of the tasks should have been randomized or a latin-square distribution should have been used. It is quite possible that a learning effect happened to the respondents, and that this may have influenced how the queries were created and the time spent on the more complex queries. It is possible that the respondents used relatively *more* time on the generic queries than queries for *scenes,* as they became more familiar with both the interface and the drawing tools the more they worked with these. If the order of the tasks had been randomized, the learning effect would have been less.

The *number of respondents* in the two groups was not equal. There were 17 respondents in the IFIM group, and 13 respondents in the KHIB group. However, while the two groups are not of equal size, they produced a reasonable amount of data and queries. It might not be possible to draw valid statistics based on the two groups, but the group sizes might provide indications to any differences between them.

Similarly, there were 256 queries performed in the VISI interface and 157 queries performed in Retrievr. While this might allow for statistically sound comparisons between the two interfaces, there may have been a structural bias within this data as a majority (79%) of the queries created in Retrievr were made by the KHIB group. Ideally, a larger number of respondents with an information science background should have tried the Retrievr interface.

10.6.3 Quality Check for the Primary Data

In an attempt control the potential structural biases introduced by some of these issues, it was decided to perform a quality check by comparing the query images created in this study with other drawn visual query images. These query images were collected from two different sources.

In (Hove 2004), 6 respondents created 36 query images using different drawing applications they chose themselves. These requests were created based on the same maritime scenario used in this work, and may represent a similar structural basis with regards to the *nature of the retrieval tasks*. However, as the images were created by a different group of respondents, at a different time and using different drawing tools, these images are not directly comparable to the query images created in this work. A sample of these query images are shown in Figure 121.

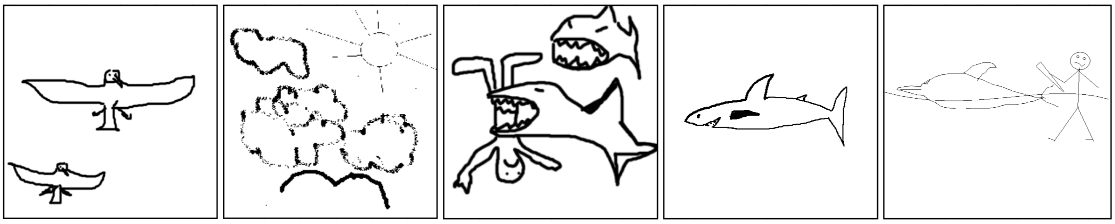


Figure 121 - Some sample images from (Hove 2004)

Next, a set of 200 query images selected from Retrievr's "Rate this image" page (SystemOneLabs 2009). The images were collected over a period of 8 weeks. While no information was available regarding the persons who created the images, what requests the query images were based on, or the context of the query process, it is likely that the queries were created by different people at different times. Consequently, these images may represent a cross-view of how different respondents chose to create their query images in Retrievr. Some of these images are represented in Figure 122.

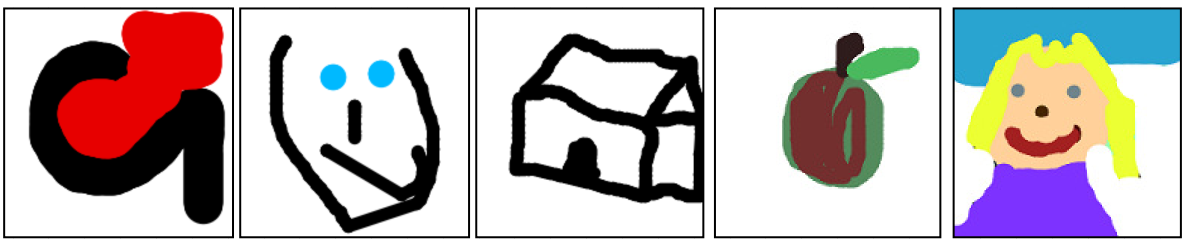


Figure 122 - Some sample images created by random Retrievr users

The query images collected from these two sources could not be put through a similarly rigorous classification as the queries during the main data collection phase. For the images created in Retrievr, no data about the query tasks were available, and as the number colours used in Retrievr could not be automatically counted this data was not available. Additionally, no information about the query time or the drawing sequence was available. Finally, the resources for doing this analysis were limited, and the same amount of time could not be spent on evaluating each of the 236 query images as was spent on the 414 query images in the main study. Consequently, this test does not present a fail-safe method of verifying these results. Nevertheless, it represents an indicator towards the generalizability of these results with regards to how different people behave. Table 61 presents a comparison of the results from the main study the results of the quality test.

Table 61 - Comparison of results from the main study and the quality test

	Overall		**VISI**		**Retrievr**	
Study	Main study	Quality test	Main study	Quality test	Main study	Quality test
N	414	236	256	36	158	200
Completeness						
Mean number of objects	2,1	2,14	2,14	2,06	2,02	2,16
Participants	98,3 %	95,30 %	99,6 %	100 %	96,2 %	94,5 %
Background	48,3 %	29,70 %	61,7 %	58,3 %	32,5 %	24,5 %
Contextual elements	8,3 %	5,90 %	18,00 %	22,2 %	3,8 %	3,0 %
Colours						
Lack of colours	24,6 %	24,2 %	26,6 %	97,2 %	20,9 %	22,0 %
Basic colour use	61,1 %	51,70 %	64,8 %	2,8 %	55,1 %	46,5 %
Varied colours	25,1 %	39,40 %	19,50 %	5,6 %	34,2 %	45,5 %
Representation						
Geometric Primitives	16,4 %	20,30 %	11,7 %	16,7 %	24,1 %	21,0 %
Outlines	89,6 %	85,20 %	95,7 %	91,7 %	80,4 %	84,0 %
Visual Cues	47,6 %	30,50 %	59,4 %	41,7 %	27,2 %	28,5 %
Texture	3,9 %	0,40 %	6,3 %	2,8 %	0,0 %	0,0 %
Composition						
Scaling	9,7 %	4,70 %	12,9 %	13,9 %	2,4 %	3,0 %
Overlap	16,4 %	8,50 %	23,0 %	30,6 %	7,8 %	4,5 %
Perspective	7,0 %	1,70 %	8,6 %	5,6 %	5,4 %	1,0 %

There are some differences in the two studies, primarily related to the use of colours. All but one of the query images created in VISI were without colours. This particular result is interesting with regards to the discussion concerning colours, as the respondents in (Hove 2004) created these images outside the context of a CBIR system. They never got to see the results of feeding the query images to a CBIR system. As noted in the discussion on colour use, respondents who *understood* that the CBIR process worked very well when colours were included tended to use colours more. For Retrievr, the use of colours is quite similar between the main study and the quality test. With the exception of colour, the overall results from the two studies are quite similar, indicating that the query images created by the respondents in this study are similar to the query images created by other people.

Summarized, some structural biases may have been introduced by the design of the experiment, the research methods and the way the image retrieval tasks were defined. Furthermore, the low number of respondents included in the study and the narrow domain may reduce the overall generality of the results. It is also likely that users would behave differently if they had more experience using QBD, as they would learn and understand how they could express their queries in a more optimal manner. Despite this, it is believed that the results provide important insights into the way new QBD users behave when drawing visual query images. The results may also be applicable to users working professionally with images. Some support for this was found based on the analysis of the 236 query images created outside the scope of this project. In addition, the use of structured interviews and an approach based on grounded theory provided a thick description of the experiences, opinions, attitudes and behaviours of the 30 respondents in the project, which would not be possible with a larger number of respondents.

10.7 Further Work and Future Research

The results obtained in this study identified several new questions and areas that should be given further studies. Five extensions of this study are suggested and discussed in the following sections:

1. Improving the drawing interfaces in QBD CBIR systems based on the results of this study
2. Introducing *shape templates* as a tool when drawing visual query images
3. Using *icons and pictograms* drawing for visual query images
4. Using *narrative structures* as an aid when interpreting drawn visual query images
5. Using *community based techniques* for semi-automatic image segmentation

10.7.1 Developing and Evaluating a Better Interface for Visual Query Specification

The most immediate follow up to this project is to develop an improved interface for drawing visual query images based on some of the suggestions presented in chapter 9. Some of these suggestions represent features that already exist in other software applications, and it is believed that adding these features to a the interface of a QBD CBIR system will present users with more flexibility without increasing the complexity of the drawing process. Sorted by the respondents' order of importance, these features are:

1. Combining QBD CBIR queries with text based queries
2. Having a result presentation interface that directly interacts with the query specification process (e.g. similar to retriever)
3. Having a result presentation interface that directly interacts with a set of query parameters (e.g. changing the weight of the colour, texture and shape parameters)
4. A more usable method of selecting colours
5. Providing the users with a dynamic and resizable pen tool
6. Allowing the users to resize and reformat the canvas
7. A better method for specifying textures

In addition to these suggestions, a majority of the respondents suggested three additional features that need to be further evaluated with regards to usability, expressive convenience and feasibility before they can be added to a retrieval system:

1. Vector based drawing and deformable objects
2. Automatic definition of query parameters
3. A colour neutral drawing process

Vector based drawing and its related feature requests (e.g. being able to modify existing objects and manipulate the shapes of the drawing tools) was a feature requested by several of the respondents in the IFIM group. While some of the respondents were very positive towards this, a minority were very vocal against it. It is possible that introducing vector based drawing may oppose the respondents' desire to keep the drawing process as simple as possible. A further investigation into this is required in order to determine if vector based drawing would improve or complicate the query drawing process.

Next, several respondents stated that they would like to have an *automatic definition of query parameters* based on their actions in the interface. This might improve the efficiency of the query process while keeping it simple, e.g. by increasing the weight of colours if colours were used much when drawing the query. The feasibility, usability and potential benefits of this approach needs to be evaluated further.

Finally, several respondents stated that they would prefer having access to a *colour neutral tool* and *a neutral canvas*. This would allow them to tell the query system that it should focus on the *shapes* in the query image and the *spatial arrangement of these*, without forcing the user to consider which colours they wanted to use. While it may be relatively straightforward to implement, the implications of this should be determined further. While this is by default possible in VISI and other systems that allow the users to specify query parameters (e.g. setting the weight of "colours" to zero), this may be in direct conflict with the objective of encouraging users to create the query images in a manner that the CBIR systems may successfully process. It is possible that other solutions should be chosen for this. Further studies are required in order to determine the implications of reducing the impact of colour.

10.7.2 Visual Query Specification using Shape Templates

As noted in section 9.2, a significant portion of the respondents in the project would like to have the option to use *predefined shapes* as drawing tools when expressing visual queries. This might help some users overcome the challenges related to their low drawing skills. A related approach was evaluated in (Hove 2004), where a thesaurus of shape templates was developed in order to assist a CBIR system to identify objects in a visual query and retrieve images. Based on the experiences made in that project, the feasability of introducing shape templates in the visual query process needs to be evaluated further with regards to the costs and and benefits of introducing it. Three steps must be taken to determine this:

1. The feasibility of building a usable set of shape templates
2. The usability of these shape templates when introduced into the drawing interface
3. The actual benefits of using shape templates with regards to CBIR similarity feature comparisons

Significant time and effort was required to create the small set of shape templates used in the shape thesaurus described in (Hove 2004), indicating that building a usable set of shape templates may require considerable work. Identifying the concepts that should have their own shape templates, and determining the number of different templates for each object (e.g. different visual representations of these objects) must be evaluated. Further studies are required in order to determine if it is possible to develop a set of shape templates useable for broad image retrieval systems, and to determine different ways of creating this. It is likely that creating an all-purpose set of shape templates represents a significant effort that is not possible to manage in a single project. However, it may be possible to compartementalize the development by creating shape templates for smaller domains and combining these over time. Furthermore, it is possible that by combining a basic set of shape templates with a *vector based approach* to the drawing process may reduce the need for a large and varied set of shapes. Determining different ways creating a set of shape templates and the overall feasability of this needs to be evalauted further.

Next, as noted by several of the respondents, it is possible that while introducing shape templates may reduce the problem of creating realistic representations, it may introduce *new* challenges, e.g.

determining how the user can have easy access to the potentially very large number of different shape templates. The feasibility and usability of this need to be evaluated further in order to determine how shape templates can be introduced to the drawing interface without complicating.

Finally, the benefits of introducing shape templates with regards to CBIR similarity feature comparisons needs to be determined. While shape templates may reduce the impact of low user drawing skills, it is uncertain if similarity functions based on perceptual structures may benefit directly from these structures. The results in (Hove 2004) indicated that a CBIR system could achieve higher levels of recall and precising by introducing a thesaurus based on shape templates, but this was primarily true when querying for *simple images*, e.g. schematic drawings or images with low contextual details, few colours and a few objects. Given the high level of visual variance in most real-world objects, the number of shape templates required in order to identify a large number of different objects may be very high, and the computational costs may be very high. However, it is possible that by adding *semantic labels* to such shape templates may *assist* the retrieval system if combined with techniques from text based information retrieval. Further studies with a larger scale than presented in (Hove 2004), combined with the use of semantic labels is required in order to determine if the benefits of this approach may outweigh the costs.

10.7.3 Visual Query Specification using Icons and Pictograms

Another feature highly requested by the respondents were the ability to compose visual query images using *pictograms* or *icons*. Introducing this requires that the query interpreted is modified and improved with the capability to identify and understand icons and pictograms. While icons and pictograms are both visual structures, their relationship to the real world objects they represent is primarily *semantic*, not *perceptual*. Consequently, in order to process these queries the retrieval system needs either to be able to *translate* the semantic structures to *perceptual* structures, or work on images that have been *pre-segmented* and *indexed* with semantic labels describing their major participants.

The major difference between using *shape templates* and using *icons and pictograms* is that when using icons, the user no longer has to consider the visual appearance of the objects they are retrieving, only the spatial relationship between them. The icon or pictogram may contain enough data to give a semantic description to the object (e.g. “Shark”), while the user can specify the spatial structure of the image[107] request, and combine this with freehand drawing or colours in order to specify other characteristics of the image request.

Several steps are required in order to analyse this further:

1. Determine the level of abstraction for the library of icons and pictograms and determine the feasibility of building this library
2. Determine how an icon library could be included in a QBD interface in a usable manner and determine the expressive convenience of using this library
3. Develop a query interpreter capable of processing the query images
4. Resolve the problem of segmenting the images in the image collection and classifying the image contents

Determining the contents and scope an icon library needs to be determined in a similar manner to defining a shape template library. However, unlike an approach based on shape templates, it would not be necessary to include different icons for each object, as the retrieval system would no longer use perceptual similarity as the primary method of identifying the query participants. Nevertheless, further research is required in order to determine *how* this library could be built and the *feasibility* of building the library.

Next, the usability of including an icon library in the QBD interface must be evaluated in a similar manner as described for the shape templates in the previous section.

Third, a query interpreter must be developed. This interpreter must be able to identify and interpret the icons placed in the image (e.g. using textual labels), determine the spatial characteristics and relationships of these icons, and retrieve images with *similar* objects placed with *similar spatial properties*, and combine this with similarity functions based on perceptual structures, if these have been used to add additional detail to the query image.

Finally, as the query interpreter and the query process can no longer use perceptual similarity as the only criteria for comparisons, the query system needs both query images and an image collection where the images been *pre-segmented*; objects in the images must be *identified* and their *spatial properties* must be defined. This is currently an unsolved problem within CBIR. One potential step towards solving this is suggested in section 10.7.5.

[107] Note that a similar approach might also be used when using shape templates. In this case, the use of shape templates would be very similar to the use of icons, the primary difference being that the user might have the possibility of describing specific visual characteristics of the participant described by the template.

10.7.4 Interpreting Visual Queries based on Narrative Structures

A major challenge for the respondents in this work was to express requests for *narrative contents* through drawing visual query images. One possible step towards reducing this challenge is to introduce the use of *narrative structures* in the visual query process. Some of the visual structures suggested by Kress and van Leeuwen (2006) (e.g. *participants, actors, goals and interaction vectors*) were briefly presented in section 2.2.5. A possible approach for using these as an aid when processing QBD CBIR queries is suggested in (Hove 2007):

1. An indexing scheme for describing narrative structures must be developed
2. A *set of rules* based on these concepts must be developed, and these rules must be translated into a set of formal rules which can be executed by a software application
3. The feasibility of the approach must be evaluated on a sufficiently large and varied image collection.

10.7.5 Community Based Image Segmentation

One of the major challenges facing CBIR is a lack of good methods for automatic image segmentation based on semantic image contents. And as noted above, having access to pre-segmented images is a requirement for some of the suggestions presented in this work.

One possible step towards semantic segmentation could be to use an approach based on *community-based* segmentation. Some recent initiatives involving community-based indexing methods such as The ESP Game (von Ahn and Dabbish 2004) and Peekaboom (von Ahn, Ruoran et al. 2006) has shown that it is possible to semi-automatically detect and index objects present in an image, as well as the spatial distribution of these objects. While these approaches are primarily based on *identifying the presence* of objects in an image, it is possible that this also can be used to describe the *spatial properties* of these objects, e.g. their placement in the image. Similarly, social web applications such as Flicker and Facebook currently allow users to add semantic labels to areas of an image.

Further studies should be made in order to determine if such approaches could be used to add meaningful semantic labels to areas of images, and if these labels could be used as a tool to improve the QBD CBIR process. The following steps are required:

1. Define a standard for describing the presence and position of participants in a digital image
2. Create a support system for community based segmentation of images and determine if it is feasible to use this to describe the presence and spatial properties of image participants
3. Analyse the feasibility of using this information as a tool for a QBD CBIR system

10.8 Concluding Remarks

This work presents three major contributions to the field of Content Based Image Retrieval:

1. It presents an empirical evaluation of how users behave when drawing image requests through a visual query interface. While the study does not cover a large number of respondents or a varied application domain, the empirical data collected in this study presents the CBIR community with important data on *how* some users draw visual queries and *what* their opinions are about the QBD approach. Further, this study *identifies real situations* where these users may *benefit* from using the QBD *despite* the current limitations of CBIR.
2. It identifies the importance of perceptual image requests for *specific users*. As noted in section 2.4, the importance of this type of request for some user groups may have been underestimated in existing literature. Consequently, even though these users might benefit from using current QBD CBIR systems, CBIR systems have not developed much past a prototype stage.
3. Finally, it identifies four steps that should be taken in order to elevate current QBD CBIR systems from a prototype stage to applications that may present *real users* with *significant improvements* in the way they can search for images.

It is hoped that the results presented in this work may represent a small step towards one of the major goals in the field of Content Based Image Retrieval: presenting end users with usable applications providing easy access to the enormous amounts of images that have been made available to the public.

References

[Reference list will be moved her for the final version. It is presently at the end of the document for formatting reasons]

Appendix 1 - Definitions

Actor	An actor represents the active part in a narrative process (Definition 11)
Contextual element	A contextual element is defined as a visual element providing situational description to the narrative structures (Definition 9)
Expressive convenience	The expressive convenience of a visual query interface is defined as the ease a user experiences when expressing a given image information request using the interface (Definition 6)
Expressive power	The expressive power of a an image query interface is defined as the type of image information requests that can be expressed using the interface (Definition 5)
Goal	A goal represents the receiving part in a narrative process (Definition 12)
Image	Images are all representations of objects, concepts, scenes, persons or abstraction, produced or stored on some medium (Definition 7)
Modality marker	A modality marker is defined as an indicator used to determine the visual modality of a query image (Definition 14)
Narrative process	A narrative process is defined as an interaction between two participants (Definition 10)
Participant	A participant is defined as an important visual element in an image (Definition 8)
Pictogram	A pictogram is a pictorial representation, an iconic sign which represents complex facts, not through words or sounds but through visual carriers of meaning (Definition 15)
Query by Drawing	Query by Drawing is defined as expressing an image need by creating visual structures through drawing using either freehand sketching or using one or more of drawing tools (Definition 4)

Query specification gap	The query specification gap is defined as the difference between the way users prefer to draw visual query images and the way these images should be drawn in order to be optimal for current CBIR systems (Definition 16)
Visual modality	Visual modality is defined as the degree to which an image represents a naturalistic rendition of the concepts depicted in the image (Definition 13)
Visual query	A visual query is defined as a request for images based on submitting, manipulating or creating visual structures, expressed in a visual query interface (Definition 2)
Visual Query Interface	A visual query interface is an interface for expressing visual queries (Definition 3)
Visual Structures	Visual structures are the basic syntactical structures present in an image, such as shapes, colours, textures and the spatial relationships between these structures (Definition 1)

Appendix 2 - Systems surveyed

59 different systems were surveyed for the evaluation described in chapter 2.5. Table 62 presents an overview of these systems, along with the query formulation techniques supported by these systems.

Table 62 - Query Formulation Techniques

System	Reference	Query Method						#
		QBT	QBF	QBIE	QBEE	QBA	QBD	
AltaVista Photofinder	(Veltkamp and Tanase 2000)	x		x				2
Amore	(Mukherjea, Hirata et al. 1997)	x		x	x			3
Berkeley DLP	(Carson and Ogle 1996)		x					1
BLOBWORLD	(Carson, Thomas et al. 1999)	x		x		x		3
CANDID	(Kelly, Cannon et al. 1995)				x			1
C-Bird	(Li, Zaïane et al. 1998)		x	x				2
CBVQ	(Smith and Chang 1995)		x	x	x			3
ChaBot	(Ogle and Stonebraker 1995)	x	x					2
CHROMA	(McDonald, Tait et al. 2001)			x	x		x	3
CIRES	(Iqbal and Aggarwal 2002)	x		x				2
CORTINA	(Manjunath, Moenich et al. 2004)	x		x				2
DrawSearch	(Kherfi, Ziou et al. 2004)					x	x	2
Excalibur VR Ware	(Veltkamp and Tanase 2000)			x				1
FIDS	(Berman and Shapiro 1999)			x				1
FIR	(Volmer 1997)				x			1
Fire	(Deselaers, Keysers et al. 2008)			x	x			2
Focus	(Das, Riseman et al. 1997)			x		x		2
Google Similar Images	(Rosenberg 2009)	x	x	x				3

Hermitage Museum[108]	(Hermitage 2003)		x				x	2
Image Management Environment (IME)	(Petraglia, Sebillo et al. 2001)							0
ImageFinder	(Veltkamp and Tanase 2000)				x			1
ImageMiner	(Kreyss, Röper et al. 1997)	x						1
ImageRETRO	(Vendrig, Worring et al. 1999)			x				1
ImageRover	(Kherfi, Ziou et al. 2004)	x		x				2
ImageScape	(Lew 2000)						x	1
Imatch	(Venters and Cooper 2000)			x				1
Jacob	(Ardizzone and La Cascia 1996)		x		x			2
LCPD	(Lew, Huijsmans et al. 1996)			x				1
MARS	(Ortega, Rui et al. 1997)		x	x				2
MetaSeek	(Benitez, Beigi et al. 1998)	x		x	x			3
MFIRS	(Pilevar 2008)			x				1
MIR	(Sirihari, Zhang et al. 2000)	x		x				2
NETRA	(Manjunath and Ma 1999)			x		x	x	3
Octagon	(Octagon 2007)			x				1
Photobook	(Pentland, Picard et al. 1996)			x				1
Picasso	(Del Bimbo, Mugnaini et al. 1997)				x		x	2
PicHunter	(Cox, Miller et al. 2000)			x				1
PICSOM	(Iivarinen, Rautkorpi et al. 2004)			x				1
PicToSeek	(Gevers and Smeulders 2000)			x	x			2
PIXIMILAR	(Idée 2008)		x	x	x			3
QBIC	(Niblack, Barber et al. 1993)		x				x	2

[108] An implementation of QBIC at the State Hermitage Museum

QuickLook[2]	(Ciocca, Gagliardi et al. 2001)			x				1
Query by Visual Keywords	(Lim 2000)						x	1
Retrievr	(Langreiter 2006)				x		x	2
Simba	(Siggelkow 2001)			x	x			2
SQUID	(Veltkamp and Tanase 2000)			x				1
SurfIMAGE	(Nastar, Mitschke et al. 1998)		x					1
Synapse	(Veltkamp and Tanase 2000)			x		x		2
Tiltomo	(Tiltomo 2006)	x		x				2
TinEYE	(Idée 2009)				x			1
TODAI	(Veltkamp and Tanase 2000)	x			x			2
Video Google	(Sivic and Zisserman 2006)			x		x		2
VisualSeek	(Smith and Chang 1997)	x					x	2
VP IR system	(Veltkamp and Tanase 2000)						x	1
WebSeek	(Veltkamp and Tanase 2000)	x		x				2
WebSeer	(Kherfi, Ziou et al. 2004)	x						1
WISE	(Wang, Wiederhold et al. 1997)			x			x	2
Xcavator	(CogniSearch 2007)	x		x		x		3

Appendix 3 - Research Questions and Hypotheses

Research question 1

How do users utilize the visual query interface when they draw visual queries?

- **RH1.1:** Respondents make frequent use of graphical drawing tools rather than drawing by freehand.
- **RH1.2**: Respondents prefer the query interface provided by VISI to the query interface by Retrievr.
- **RH1.3**: Respondents draw query images more quickly in the interface provided by VISI than in the query interface provided by Retrievr.
- **RH1.4:** Respondents with a visual background express queries faster than respondents without this background.

Research question 2

How realistic are the query images drawn by QBD CBIR users?

- **RH2.1:** Respondents will create query images with a low degree of visual modality.
  - **RH2.1.1:** Respondents will create query images that are *simple*
  - **RH2.1.2:** Respondents do not make much use of colours when creating visual query images
  - **RH2.1.3:** Respondents will depict query image participants as geometric primitives without using representational components
  - **RH2.1.4:** Respondents do not use compositional structures when creating query images

- **RH2.2**: Respondents with a visual background will create query images with a higher degree of visual modality than respondents without this background
  - **RH2.2.1:** Respondents with a visual background will create more complete query images than user without this background'
  - **RH2.2.2:** Respondents with a visual background make more use of colours than respondents without this background
  - **RH2.2.3:** Respondents with a visual background will depict image participants more realistically than respondents without this background
  - **RH2.2.4:** Respondents with a visual background use more compositional structures than respondents without this background

- **RH2.3:** Query images created in the VISI interface will have a higher degree of visual modality than queries created in the Retrievr interface
  - **RH2.3.1:** Respondents will create more complete query images in the VISI interface than in the Retrievr interface
  - **RH2.3.2:** Respondents make more use of colours in the VISI interface than in the Retrievr interface
  - **RH2.3.3:** Respondents will depict image participants more realistically in the VISI system than in the Retrievr system
  - **RH2.3.4:** Respondents use more compositional structures in the VISI system than in the Retrievr system

- **RH2.4**: The visual modality of the query images increases with the complexity of the image requests
    - **RH2.4.1:** The completeness of the query images increases with the complexity of the image requests
    - **RH2.4.2:** The use of colours in the query images increases with the complexity of the image requests
    - **RH2.4.3**: The degree of abstraction decreases with the complexity of the image requests
    - **RH2.4.4**: The compositional modality of the query images increases with the complexity of the image request

Research question 3

What are the major challenges encountered when users draw visual queries?

- **RH3.1**: Lack of drawing skills is a major challenge when respondents draw visual query images
- **RH3.2:** Drawing visual queries is too time-consuming to be an efficient tool for image retrieval
- **RH3.3:** Lack of usable interface tools is a major challenge when drawing visual query images

Research question 4

How do users feel about expressing image requests by drawing visual queries?

- **RH4.1:** Respondents do not like to express image retrieval tasks by drawing visual query images
- **RH4.2:** Respondents with a 'visual background' are more positive towards expressing image retrieval tasks by drawing visual query images than respondents without this background
- **RH4.3:** Respondents do not prefer to use drawn visual queries over text based image queries

Research question 5

What improvements can be made to CBIR systems in order to better support users when drawing visual query images?

- **No research hypotheses suggested**

Overview of Hypothesis Answers

Only the main hypotheses are included. Sub-hypotheses are not included in this summary. As there were no hypotheses for RQ 5, it is omitted from the table. For more details, see chapters 5 through 9.

Table 63 - Overview of hypothesis evaluation

RQ	Hypothesis	Status	Comments
1	1.1	Rejected	The respondents used *freehand drawing* as the primary drawing technique. However, there was an increased use of drawing tools in queries for *complete scenes*
	1.2	Rejected	The respondents reported that there were elements they enjoyed in both interfaces, and would have preferred a combination of the two interfaces
	1.3	Rejected	The respondents spent significantly shorter time creating queries in Retrievr than in VISI. This was valid for both respondent groups and all query categories
	1.4	Accepted	The respondents in the KHIB group spent significantly less time creating the queries than the respondents in the IFIM group
2	2.1	Accepted	The respondents created queries with a low degree of visual modality
	2.2	Rejected	There were no major significant differences in the visual modality of the query images drawn by the two respondent groups
	2.3	Accepted	The queries created in VISI held a higher visual modality than the queries created in Retrievr. This was true for all modality markers
	2.4	Partially accepted	There were significant differences in the visual modality of the query images created for the different image requests. This was particularly true when querying for *complete scenes* There was generally an *increase* in *completeness, use of colours* and *use of compositional structures* as the query complexity increased, but there was a *decrease* in the *representational modality*
3	3.1	Partially accepted	The hypothesis is *accepted* for the respondents in the IFIM group, while it is *rejected* for the respondents in the KHIB group The respondents in the IFIM group found their perceived lack of drawing skills to be a major challenge when drawing visual query images
	3.2	Rejected	The hypothesis is *rejected*. Most of the respondents did not find the time required to express visual queries to be a major challenge It should be noted that a minority of the respondents stated that they were most likely not willing to spend this much time creating queries. The respondents who request images regularly on a professional basis were positive towards spending time using QBD CBIR systems
	3.3	Rejected	The tools available in the two interfaces did not present a major challenge when drawing visual query images
4	4.1	Rejected	Most of the respondents stated that they had a pleasurable experience expressing image queries using QBD.
	4.2	Accepted	The hypothesis must be *accepted*. There was a definitive difference in the way the two groups felt towards using QBD. Both groups were positive, but the KHIB group was more positive than the IFIM group, had fewer reservations and saw fewer challenges with QBD than the IFIM group
	4.3	Accepted	While most of the respondents saw some potential uses for QBD, very few of the respondents were willing to use QBD instead of text based queries. Most respondents saw a number of areas in which they felt QBD could be a *complement* to text based queries. In addition, some of the respondents identified some particular areas where they claimed QBD might be *better than* text.

Appendix 4 - Data Collection Tools

Introductory Letter (In Norwegian)

Introduksjon

Takk for at du deltar i dette forsøket!

Hensikten med dette forsøket er å undersøke en ”ny” måte å søke etter bilder og informasjon. De fleste er kjent med den tradisjonelle tilnærmingen ved å søke ved hjelp av enkle søkeord, for eksempel gjennom Google.

Istedenfor å søke ved hjelp av tekst, skal du nå få forsøke å søke ved hjelp av å tegne. Vi har laget et system som forsøker å finne bilder basert på likhet mellom en tegning og bilder lagret i en database. Bildene du skal tegne skal lages i et enkelt søkeprogram med enkle tegneverktøy. Vi er interessert i å finne ut mer om hva du synes om hvor lett det er å bruke disse verktøyene til å uttrykke søk gjennom å tegne fremfor å bruke tekstlige søk.

Bildesamlingen består av bilder av maritimt dyreliv og aktiviteter knyttet opp mot dette: Delfiner, hvaler, måker, dykkere og så videre. Søkeoppgavene vil være knyttet opp mot dette.

Du vil først få en kort introduksjon til søkeverktøyet og få anledning til å gjøre deg kjent med dette før vi går i gang med selve forsøket. Selve søkesystemet er ikke så veldig avansert, så du bør ikke forvente alt for gode resultater av søket ditt. Vi er først og fremst interessert i å vite hva du tenker om de verktøyene du kan benytte for å uttrykke søket.

Når du er klar for å gå i gang, vil du få utlevert et lite sett med søk du skal forsøke å utføre i programmet. Jeg kommer til å være tilstede mens du arbeider med oppgavene, og det vil være full anledning til å spørre om ting du lurer på underveis. Jeg kommer til å bruke en lydopptaker, så det vil være veldig fint om du ”tenker høyt” når du jobber, slik at vi kan få mest mulig informasjon ut av dette.

Etter du er ferdig med oppgavene, kommer jeg til å spørre noen spørsmål om hvordan du syntes det var å bruke systemet. Jeg kommer til å stille en del spørsmål, og det vil bli anledning for deg til å komme med synspunkter og kommentarer til søkeverktøyet.

Resultatene fra forsøket vil gå inn som en del av en doktorgradsavhandling, og vil bli antageligvis bli publisert i artikkelform i løpet av 2007.

Tid

Jeg regner med at det kommer til å gå i overkant av 1 time for dette forsøket, men det vil antageligvis variere noe fra person til person.

Datainnsamling

I tillegg til intervjuet, kommer jeg til å bruke lydopptager til å ta opp det vi snakker om mens du gjennomfører søket. Denne kommer jeg også til å bruke under intervjuet. I tillegg kommer jeg til å lagre de ferdige søkebildene dine, samt et opptak av de handlingene du gjør når du bruker systemet. I tillegg til intervjuet vil dette danne grunnlag for analysen vi kommer til å gjøre i etterkant.

Du vil også bli bedt om å fylle ut et skjema med bakgrunnsinformasjon: Kjønn, alder, erfaring med søkemotorer og så videre.

Anonymitet / avidentifisering

Etter at du er ferdig med oppgavene og intervjuet, kommer informasjonen vi har samlet inn til å bli registrert i en database. Lydopptaket vil bli transkribert og råmaterialet vil bli slettet etter transkriberingen er fullført. Informasjonen vil bli lagret slik at din identitet ikke vil kunne kobles til det som blir registrert. Et kallenavn vil bli brukt for å identifisere informasjon som fremkommer fra din deltakelse, men det vil ikke være mulig å koble dette tilbake til deg. Du kan derfor være sikker på at det ikke vil bli mulig å knytte din identitet opp mot resultatene av prosjektet.

Kontakt

Dersom det er noe du lurer på før eller etter forsøket, må du bare ta kontakt med meg. Jeg kan nås på telefon 55 58 91 05 på dagtid, eller pr e-post lars-jacob.hove@infomedia.uib.no.

Form of Consent (Handout in Norwegian)

Samtykkeskjema

Undertegnede bekrefter herved å ha mottatt skriftlig og muntlig informasjon om prosjektet og er villig til å delta i undersøkelsen. Jeg gjør oppmerksom på at det ikke vil bli gjort koblinger mellom dette skjema og de data som blir samlet inn i undersøkelsen.

____

Navn (Blokkbokstaver)

____

Dato, Signatur

Tasks Used in Experiment 1(Handout in Norwegian)

Før du går i gang med søkeoppgavene kan du gjøre noen øvelsesoppgaver for å bli kjent med søkeverktøyet og hvordan det skal brukes. Bruk så lang tid du vil på å gjøre deg kjent med tegnebrettet og søkegrensesnittet.

Når du er ferdig med øvingsoppgavene, skal du utføre søkene som er beskrevet nedenfor.

1. Finn bilder av en måke
2. Finn bilder av en dykker
3. Finn bilder som inneholder én eller flere haier
4. Finn bilder som inneholder én eller flere fugler
5. Finn bilder av én eller flere måker som spiser
6. Finn bilder av haier som angriper

Tasks Used in Experiment 2(Handout in Norwegian)

I denne delen av forsøket skal du gjennomføre et 8 søkeoppgaver jeg har definert på forhånd. Du skal gjennomføre minst ett søk for hver av oppgavene, men du står fritt til å gjøre flere søk dersom du ikke er fornøyd med et søk.

1. Finn bilder av en dykker
2. Finn bilder som inneholder én eller flere hvaler
3. Finn bilder av en eller flere måker som spiser
4. Finn bilder av en hai som angriper
5. Finn bilder av en måke som spiser en fisk
6. Finn bilder av en person som mater en delfin
7. Finn bilder av en fugl som er skadet
8. Finn bilder av en lykkelig delfin

Det er ikke noe tidspress på å få gjennomført disse spørringene, men jeg ønsker at du gjennomfører dem som om du var i en reell situasjon.

Tasks Used in Experiment 3(Handout in Norwegian)

VISI tasks

I denne delen av forsøket skal du gjennomføre 6 forhåndsdefinerte søkeoppgaver, og 2 søk du definerer selv ut fra en tekst. Du skal gjennomføre minst ett søk for hver av oppgavene, men du står fritt til å gjøre flere søk dersom du ikke er fornøyd med et søk.

Bildesamlingen inneholder primært bilder knyttet opp mot et maritimt miljø: Hval, delfiner, fisk, måker, og mennesker og redskaper knyttet opp mot dette.

1. Finn bilder av et havdyr
2. Finn bilder av et skip
3. Finn bilder av et rovdyr som angriper et byttedyr
4. Finn bilder av mennesker som pleier en strandet og skadet hval
5. Finn bilder av to delfiner som underholder mennesker i en delfinpark
6. Finn bilder av et eller flere havdyr som svømmer i et arktisk landskap

Det er ikke noe tidspress på å få gjennomført disse søkene, men jeg ønsker at du gjennomfører dem som om du var i en reell søkesituasjon.

Retrievr tasks

I denne delen av forsøket skal du gjennomføre 6 forhåndsdefinerte søkeoppgaver, og 2 søk du definerer selv ut fra et dikt. Du skal gjennomføre minst ett søk for hver av oppgavene, men du står fritt til å gjøre flere søk dersom du ikke er fornøyd med et søk.

Bildesamlingen inneholder utvalg bilder hentet tilfeldig fra Flickr, og svarsettet kan derfor inneholde alle mulige motiver og bilder.

1. Finn bilder som inneholder en blomst, et tre eller en annen form for vekst
2. Finn bilder som inneholder et møbel eller et interiørobjekt
3. Finn bilder av mennesker som utøver en sport
4. Finn bilder av en lykkelig jente
5. Finn bilder av et bylandskap (”Skyline”)
6. Finn bilder av flere mennesker og/eller dyr samlet i et naturlandskap

Det er ikke noe tidspress på å få gjennomført disse søkene, men jeg ønsker at du gjennomfører dem som om du var i en reell søkesituasjon.

Scenario Texts (Handout in Norwegian)

VISI Scenario (Newspaper article)

”Flipper” - Delfin på besøk i norske farvann

Flipper er en delfin, nærmere bestemt en tumler. Alderen anslås til 20-25 år (i 2001), så han er en voksen "mann". Flipper er antatt å veie mellom 300 og 400 kilo. Lengden antas å være 3,5 meter. I en årrekke har Flipper kommet til Rogaland for å søke kontakt med mennesker. Men hvorfor?

- Fra andre flokkdyr kjenner vi til at tidligere dominante hanner trekker seg ut av flokken. Dette har vi sett eksempler på, både blant spermasetthvalene og hos ulver. Jeg antar at Flipper sitt alternative liv, i menneskers selskap, kan ha bakgrunn i at han ikke taklet å bli fratatt sin status som den seksuelt dominerende hannen i flokken, sier Arne Bjørge ved Havforskningsinstituttet i Bergen. Egentlig er det ikke vanlig at tumlere besøker Vestlandet.

- Flipper, som er en tumler, opererer utenfor sitt egentlige utbredelsesområde. De fleste bestander av denne delfinarten holder til i sørligere farvann, og vi har ingen tumlerbestand ved norskekysten, forteller Bjørge.

Gledet mange - en internasjonal stjerne!

Både store og små har hatt glede av Flippers besøk. Den kontaktsøkende delfinen liker å bli klappet og kost med, og den svømmer gjerne om kapp med fiskebåter og fritidsbåter. Reisende mellom Stavanger og Skudeneshavn har ofte opplevd å se Flipper lekende i kjølvannet fra ferga. Også på Kvitsøy og i Åkrehamn har Flipper hatt faste besøk. Når han viser seg i havna, strømmer barn og voksne til kaikanten. Ellers har mange båtfolk fått lettere sjokk når Flipper plutselig dukker opp. Man vet ikke at han er der før man hører hans hvesende utblåsninger idet han puster ut.

Flipper har også vist seg å være en leken delfin. Det har vært rapportert om flere tilfeller der Flipper har begynt å leke med baller, eller ball-lignende objekter i sjøen. Han har bokset ballen med snuten, for så å jage etter ballen, til stor jubel og entusiastiske bifall fra tilskuerne.

Mange turister har hatt en uforglemmelig tur til Karmøy på grunn av Flipper. Noen av disse har også laget hjemmeside der de forteller om Flipper og deres møte med ham. Den nederlandske hjemmesiden er skrevet på engelsk og inneholder også en del bilder. Der finner du også linker til en del avisomtale fra flippers opptreden de to siste årene.

Skadet av propell

Flippers mange besøk langs Rogalandskysten har ikke gått helt smertefritt for seg. Den lekne delfinen liker å svømme om kapp med båter, og elsker strømningene fra propellvannet. I fjor (2001) sommer var Flipper døden nær på grunn av sin kjærlighet til motoriserte fartøy. I Vågen i Stavanger kom Flipper i nærkontakt med propellen til brannvesenets båt; "Nøkk". Sammenstøtet resulterte i tre dype kutt på hode og rygg. Heldigvis har delfiner et tykt spekklag, og sårene var ikke dødelige. Episoden kunne imidlertid ha medført døden dersom Steinar Bastesen hadde fått gjennomslag for sitt forslag om avliving. Flipper overlevde både kuttskader og Bastesen, og året etter kom Flipper tilbake. Sårene hadde grodd, men skaden hadde etterlatt store arr.

Retrievr Scenario (Poem)

"I skogen" Av Marit Irene Jensen

Grønn og fin i all sin prakt
Skogen står der i sin drakt

Stubbe, stein og mosedott
Bær og sopp vi plukker opp
Lyng og løv det tar vi med
Pynter hus og gjør det pent

Vinterkledd og ren
Snøen dekker den så pen

Jegern kommer med sin bue
Skyter haren midt i hue
Hagla tok han også frem
Gaupa står der nå i spenn

Skogen fanger, skogen gir
Mangfold rundt om hver en sti

Rype rev og harespor
Leder Jegern dit dem bor
Og når Jegern kommer frem
Fest for venner i hans hjem

Questionnaire 1 - Background

Only the questionnaire used in the third experiment is included. Questions marked with * were not asked in all three experiments.

Deltakernummer:

Alder: ______ **Kjønn**: Mann / Kvinne

1. Hvor stor erfaring har du med søkemotorer som Google, MSN og Kvasir?

Ingen Mye

1 2 3 4 5

2. Hvor stor erfaring har du fra søkemotorer tilpasset bilder, som Google Images eller Kvasir Bildesøk?

Ingen Mye

1 2 3 4 5

3. Hvor stor erfaring har du fra bildelagringssystemer som *Flickr*

Ingen Mye

1 2 3 4 5

4. Hvor stor erfaring har du fra visuelle søk - søk der du tegner for å søke?

Ingen Mye

1 2 3 4 5

5. Hvor ofte vil du anslå at du søker etter bilder via ulike søkemotorer?

Daglig Ukentlig Månedlig Sjeldnere Aldri

6. Hvis du søker etter bilder, søker du da privat eller i forbindelse med jobb (studier)?

Kun privat Mest privat Likt Mest Jobb Kun Jobb (Søker ikke)

7. Hvordan vil du vurdere dine egne tegneferdigheter?

Lite flink Svært flink

1 2 3 4 5

8. Hvor ofte vil du anslå at du tegner for hånd? *

Daglig Ukentlig Månedlig Sjeldnere Aldri

9. Hvor stor erfaring har du med å tegne på en datamaskin ved hjelp av mus?

Ingen				Mye
1	2	3	4	5

10. Hvor ofte vil du anslå at du tegner på en datamaskin ved hjelp av en mus? *

Daglig Ukentlig Månedlig Sjeldnere Aldri

11. Hvor stor erfaring har du med å tegne på en datamaskin ved hjelp av et tegnebrett?

Ingen				Mye
1	2	3	4	5

12. Hvor ofte vil du anslå at du tegner på en datamaskin ved hjelp av et tegnebrett? *

Daglig Ukentlig Månedlig Sjeldnere Aldri

13. Hvis du har erfaring fra bruk av visuelle søk - vennligst list opp de søkemotorene / søkesystemene du har erfaring med: *

14. Hvis du har noen formell utdanning innen tegning, visuell kommunikasjon eller andre visuelle områder, vennligst marker her: *

Videregående	JA	Type:______________________________
Bachelorstudier	JA	Type:______________________________
Masterstudier	JA	Type:______________________________
Annen	JA	Type:______________________________

Questionnaire 2 - After the Query Session

The questionnaires from experiment 2 and 3 are included. The questionnaire from experiment 1 is similar to the questionnaire used in experiment 2, but some new questions were added in experiment 2. These questions are indicated by *.

Experiments 1 and 2

1. Hvor godt likte du å søke etter bilder på denne måten

Ikke godt Svært godt

1 2 3 4 5

2. Hvor lett synes du det var å uttrykke søk ved hjelp av bilder?

Svært vanskelig Svært lett

1 2 3 4 5

3. Hvor lett synes du det var å *tegne* søkebildene i dette verktøyet

Svært vanskelig Svært lett

1 2 3 4 5

4. Hvor lett synes du det var å bruke penn og brett til å tegne

Svært vanskelig Svært lett

1 2 3 4 5

5. Hvor fornøyd er du med utvalget av tegneverktøy / hjelpemidler i grensesnittet

Ikke fornøyd Svært godt fornøyd

1 2 3 4 5

Hvis du tegnet for frihånd

6. Hvor lett synes du det var å tegne for frihånd (tegne fritt)

Svært vanskelig Svært lett

1 2 3 4 5

Hvis du brukte noen av de ferdige formene

7. Hvor lett synes du det var å bygge/tegne søkebildet ved hjelp av de ferdige formene?

Svært vanskelig Svært lett

1 2 3 4 5

8. Hvor lett var det å forstå bruken av vektene (farge, form, tekstur, spatial)? * (Only exp 1)

Svært vanskelig Svært lett

1 2 3 4 5

9. Hvor lett var det å forstå bruken av terskelverdien? * (Only exp 1)

Svært vanskelig Svært lett

1 2 3 4 5

10. Dersom et slikt system for bildesøk var tilgjengelig i dag, hvor sannsynlig er det at du ville brukt det istedenfor tekstlig bildesøk?

Svært usannsynlig Svært sannsynlig

1 2 3 4 5

11. Dersom et slikt system var tilgjengelig i dag, hvor sannsynlig er det at du ville brukt det i *tillegg til* tekstlige bildesøk?

Svært usannsynlig Svært sannsynlig

1 2 3 4 5

12. Hvor tidkrevende opplevde du at denne søkemetoden er? * (Exp2 only)

Svært lite tidkrevende Svært tidkrevende

1 2 3 4 5

13. Hvor problematisk opplevde du tidsaspektet ved søkemetoden? * (Exp 2 only)

Svært lite problematisk Svært problematisk

1 2 3 4 5

14. I hvor stor grad følte du at dine egne tegneferdigheter hadde innvirkning på din evne til å lage gode søk? *(Exp 2 only)

Svært liten grad Svært stor grad

1 2 3 4 5

15. I hvor stor grad følte du at tilgangen på verktøy i grensesnittet hadde innvirkning på din evne til å lage gode søk? * (Exp 2 only)

Svært liten grad Svært stor grad

1 2 3 4 5

16. Nedenfor finner du en rekke ord. Marker de ordene du mener passer godt for å beskrive denne søkemåten

Tidkrevende Hurtig Morsom Tungvindt Nyttig Leketøy

Unyttig Komplisert Brukbar Enkelt Effektivt

Arbeidssparende Kreativt Arbeidskrevende Kjedelig

Mangelfull

Experiment 3

DEL A: Generelt om visuelle søk

I denne delen er spørsmålene relatert til det å søke etter bilder på denne måten, og ikke direkte til de to systemene du har prøvd. Du vil bli spurt om de to ulike systemene i del B og del C.

1. Hvor godt likte du å søke etter bilder på denne måten

Ikke godt — Svært godt

1 2 3 4 5

2. Hvor lett synes du det var å uttrykke søk ved hjelp av bilder?

Svært vanskelig — Svært lett

1 2 3 4 5

3. Hvor lett synes du det var å bruke penn og brett til å tegne

Svært vanskelig — Svært lett

1 2 3 4 5

4. Dersom et slikt system for bildesøk var tilgjengelig i dag, hvor sannsynlig er det at du ville brukt det istedenfor tekstlig bildesøk?

Svært usannsynlig — Svært sannsynlig

1 2 3 4 5

5. Dersom et slikt system var tilgjengelig i dag, hvor sannsynlig er det at du ville brukt det i *tillegg til* tekstlige bildesøk?

Svært usannsynlig — Svært sannsynlig

1 2 3 4 5

6. Hvor tidkrevende opplevde du at denne søkemetoden er?

Svært lite tidkrevende — Svært tidkrevende

1 2 3 4 5

7. Hvor problematisk opplevde du tidsaspektet ved søkemetoden?

Svært lite problematisk — Svært problematisk

1 2 3 4 5

8. I hvor stor grad følte du at dine egne tegneferdigheter hadde innvirkning på din evne til å lage gode søk?

Svært liten grad — Svært stor grad

1 2 3 4 5

9. I hvor stor grad følte du at tilgangen på verktøy i grensesnittet hadde innvirkning på din evne til å lage gode søk?

Svært liten grad — Svært stor grad

1 2 3 4 5

10. Nedenfor finner du en rekke ord. Marker de ordene du mener passer godt for å beskrive denne søkemåten

Tidkrevende	Hurtig	Morsom	Tungvindt	Nyttig	Leketøy
Unyttig	Komplisert	Brukbar	Enkelt	Effektivt	
Arbeidssparende	Kreativt	Arbeidskrevende		Kjedelig	
Mangelfull					

DEL B: Oppfatninger om VISI

Denne delen omfatter søkemotoren knyttet opp mot de maritime bildene og det maritime scenarioet - VISI.

11. Hvor lett synes du det var å lage søkebildene dine i dette verktøyet?

Svært vanskelig Svært lett

1 2 3 4 5

Hvis du tegnet for frihånd

12. Hvor lett synes du det var å tegne med frihåndsverktøyet?

Svært vanskelig Svært lett Ikke frihånd

1 2 3 4 5

Hvis du brukte noen av de ferdige formene

13. Hvor lett synes du det var å bygge/tegne søkebildet ved hjelp av former?

Svært vanskelig Svært lett Ikke former

1 2 3 4 5

14. Hvor fornøyd er du med utvalget av tegneverktøy / hjelpemidler i grensesnittet

Ikke fornøyd Svært godt fornøyd

1 2 3 4 5

15. Hvor fornøyd er du med utvalget av farger i grensesnittet

Ikke fornøyd Svært godt fornøyd

1 2 3 4 5

DEL C: Oppfatninger om Retrievr

Denne delen omfatter søkemotoren knyttet opp mot de maritime bildene og det maritime scenarioet - Retrievr.

16. Hvor lett synes du det var å lage søkebildene i dette verktøyet

Svært vanskelig Svært lett

1 2 3 4 5

17. Hvor lett synes du det var å tegne med frihåndsverktøyet?

Svært vanskelig — Svært lett

1 2 3 4 5

18. Hvor fornøyd er du med utvalget av tegneverktøy / hjelpemidler i grensesnittet

Ikke fornøyd — Svært godt fornøyd

1 2 3 4 5

19. Hvor fornøyd er du med utvalget av farger i grensesnittet

Ikke fornøyd — Svært godt fornøyd

1 2 3 4 5

20. Hvor godt likte du at resultatene ble vist fortløpende?

Likte det ikke — Likte det svært godt

1 2 3 4 5

Interview Guide (In Norwegian)

The final interview guide used in the experiments is shown. The interview guide evolved continuously throughout the experiment sessions.

A. Generell oppfatning og førsteinntrykk av visuelle søk

1. Hva synes du om å søke på denne måten?
2. Er dette noe du kunne tenke deg å bruke til vanlig?
3. Kan du ha nytte av denne type søk? Og i hvilke situasjoner?
4. Er det noen spesielle situasjoner, eller spesielle yrker, du tror vil ha nytte av denne type søk?

B. Utfordringer

5. Hva var de største utfordringene dine ved å søke på denne måten?
6. Hvor lett var det å formulere et søk visuelt fremfor å gjøre det skriftlig?
7. Hvilken rolle mener du dine egne tegneferdigheter spilte i denne sammenhengen?
8. Hva synes du om tiden det tok å lage denne type søk?
9. Hadde utvalget av verktøy i grensesnittene noen innvirkning på dette?
10. Var det noen type innhold som var vanskelig å få uttrykt?
    - Handling, samhandling, tilstand?
11. Når du tenker tilbake på søkeoppgavene, var det noen av disse som opplevdes som vanskeligere eller lettere enn andre?
    - Hvorfor / hvorfor ikke?

C. Utforming av visuelle søk

12. Hvordan gikk du frem for å utforme de visuelle søkene dine?
    - Eksisterte det et mentalt bilde?
13. Var du i så fall fornøyd med bildet du fikk tegnet?
    - Hvorfor / hvorfor ikke?
14. Var du bevisst på ditt valg av abstraksjonsnivå?
    - Diskuter deltakerens valg av realistisk/ikonisk fremstilling
    - Realistisk eller generalisere / abstrahere tegningene?
    - Hvorfor gjorde du dette?
15. Hvor lett var det for deg å få uttrykt ulike former for komplekst innhold?
    - Handling, tilstand og samhandling?
    - Ta utgangspunkt i deltakerens bilder
    - Hvordan gikk du frem for å uttrykke dette komplekse innholdet?
16. Var du bevisst på bruk av komposisjon i bildene?
    - Hvordan gikk du frem for å komponere bildene?
    - Hvorfor valgte du å plassere objektene bildene der du gjorde?

- Diskuter bruken av kontekstuelle elementer
- Diskuter sekvensen i tegningen - hva tenker deltakere om dette?

17. Kan du fortelle meg litt om de verktøyene du valgte å bruke?
    - Hvorfor brukte du [frihånd, former]
    - Hvorfor brukte du / brukte du ikke farge?
    - Oppfatter du hvitfargen på lerretet som en "nøytral" farge, eller oppfattet du den som fargen hvit?

D. Oppfatninger om søkeverktøyene og grensesnittene

18. Hva synes du om de to grensesnittene?

19. Hva synes du var den største forskjellen mellom grensesnittene?

20. Hvilket grensesnitt synes du var enklest å bruke, og hvorfor?

21. I hvor stor grad egnet søkeverktøyene seg for de søkene du skulle gjennomføre? (Begge grensesnitt)

22. Hva synes du om utvalget av verktøy (begge grensesnitt)

23. Hva synes du om utvalget av farger (Begge grensesnitt)?

24. Hva synes du om å tegne for frihånd (Begge grensesnitt)?

25. Hva synes du om størrelsen på lerretet (Begge grensesnitt)?

26. Hvorfor valg mellom frihånd / andre verktøy (VISI)

27. Hvor lett var det å forstå vektprinsippene (VISI)

28. Bruk av mus / tegnebrett

E. Forbedringspotensial

29. Hva kunne gjort denne type søk lettere?

30. Hvordan kunne verktøyet ha støttet bedre opp i forhold til domenet man søkte etter?

31. Hvordan kunne verktøyet ha avhjulpet noen av problemene og utfordringene du opplevde?
    - Tegneferdigheter
    - Tidsbruk
    - Domeneproblemer
    - Verktøymangel
    - Andre problemer deltakeren selv har presisert

32. Mot slutten av intervjuet, dersom deltakeren selv ikke har tatt opp disse tingene: Hva tror du om følgende tillegg?
    - "Vektorbaset" tegning
    - Mulighet for å tilpasse figurene
    - Mulighet for å se endringer i sanntid
    - Hva tror du om å erstatte hvitfargen på lerretet med en "ikke-farge" eller nøytral bakgrunn?

Information Memo to the Evaluators

Innledning

Dette dokumentet er et arbeidsdokument for å foreta en analyse av bildematerialet som er samlet inn. Hensikten er å forsøke å analysere hvordan brukerne mine har laget søkebildene.

Jeg har satt opp et rammeverk for vurdering av disse bildene basert på en utvikling og tilpasning av teori fra boken ”Reading Images: The Grammar of Visual Design” av Gunther Kress og Theo van Leeuwen. Rammeverket skal kunne brukes for å undersøke to ting:

- Hvor realistiske og komplette bildene er - bildenes modalitet
- Hvilke verktøy som er blitt benyttet for å tegne bildene - verktøybruk

For å vurdere bildenes modalitet, er det definert fire modalitetsområder som skal vurderes:

- Bruk av farge
- Bruk av kontekstualisering
- Abstraksjonsnivå
- Bruk av dybde og perspektiv

For hvert av disse kriteriene har jeg satt opp ett sett med kriterier som kan være oppfylt eller ikke oppfylt for et gitt bilde, altså ”JA” eller ”NEI”.

I tillegg er det satt opp et mål der du kan gi en subjektiv vurdering av hvor ”realistisk” et gitt bilde er for et gitt modalitetsområde, på en skala fra 1 til 5, der 1 representerer den laveste verdien, og 5 representerer den høyeste verdien. Dette skal settes ut fra den som evaluerer bildenes egen subjektive oppfatning av det gitte modalitetsområdett. De ulike begrepene er presentert og definert under dette. Denne vurderingen er minst like viktig som de mer ”objektive” kriteriene.

Hovedområder / modalitetsbegrep

Fargebruk

Dette beskriver i hvor stor grad, og på hvilken måte, farger er blitt benyttet når et bilde er laget. Her har jeg satt opp følgende kriterier:

- **Monokromatisk**: Dette beskriver bilder som er laget utelukkende ved hjelp av én farge på et hvitt lerret.
- **Enkel fargebruk:** Dette beskriver bruk av ”enkle” farger. Med dette mener jeg at et eller flere objekter er i enkeltfarger. For eksempel én grønnfarge for å fargelegge et tre, én blåfarge for å fargelegge en sjø eller én rødfarge for å fargelegge et objekt.
- **Nyansert fargebruk:** Dette beskriver ”enkel” fargebruk der ett eller flere objekter er gitt flere farger. For eksempel ulik farge på trestamme og treblader, ulike farger i en persons genser osv.
- **Fargegradering**: Dette beskriver situasjoner der et område av et bilde har flere graderinger av samme farge, for eksempel at en himmel varierer fra lyst blått fra mørkt blått eller at fargen til et objekt endres som følge av Lyssettingr.

- **Lyssetting**. Dette representerer bilder der tegneren har forsøkt å representere lyssetting, der én eller flere Lyssettingr har innvirkning på bildet i form av fargemodulering, fargespill, skygger eller lignende.

Subjektiv vurdering

Dette er din subjektive vurdering av i hvor stor grad farge er blitt brukt til å lage et realistisk bilde. Settes på en skala fra 1 - 5, der 1 representerer et (tilnærmet) fravær av fargebruk, mens 5 representerer en (tilnærmet) realistisk bruk av farger.

Kontekstualisering

Dette beskriver i hvor stor grad *kontekstuelle detaljer* og *bakgrunn* er brukt for å komponere et bilde. *Kontekstuelle detaljer* representerer ting man vanligvis vil finne i et "ekte" bilde, men som ikke er direkte relevant for det man søker etter. For eksempel vil et bilde av en delfin gjerne inneholde skyer, bølger, et korallrev eller en fiskestim. Altså detaljer som er med på gå i en kontekst til det objektet eller objektene man er interessert i å finne. Her har jeg satt opp følgende kriterier:

- **Bruk av interesseobjekter**. Et *interesseobjekt* er det eller de sentrale objektene som er målet for et søk. For eksempel vil en tegning av en delfin være et interesseobjekt i et søk av typen "finn bilder av delfiner". De aller fleste bilder vil ha med slike interesseobjekter, men det kan tenkes at de ikke er inkludert.
- **Bruk av bakgrunn**. Dette beskriver bruk av annen bakgrunn enn hvit / nøytral bakgrunn, for eksempel en farget bakgrunn for å representere sjø eller himmel.
- **Symbolske kontekstelementer:** Et *symbolsk kontekstelement* er et kontekstelement som har en høy symbolsk verdi for bildene man forsøker å finne. Eksempel på dette kan være bruken av en "sol" eller en "sky" for å representere at bildet er "utendørs", eller at det er "fint vær", eller bruken av en bølget eller rett linje for å representere havoverflaten.
- **Detaljerte kontekstelementer:** Et *detaljert kontekstelement* er et kontekstelement som ikke har høy symbolsk verdi for selve søket, men som det vil være naturlig å finne representert i et reelt bilde. Eksempler på dette kan være en fiskestim, trær eller andre "ikke-relevante" objekter.

Subjektiv vurdering

Dette er din subjektive vurdering av i hvor stor grad kontekstualisering er brukt for å lage et realistisk bilde. Settes på en skala fra 1 - 5, der 1 representerer et fravær av kontekst (for eksempel interesseobjekter direkte representert på en nøytral bakgrunn), mens 5 representerer full kontekstualisering (interesseobjekter er plassert i sin naturlige kontekst).

Representasjon

Dette beskriver i hvor stor grad et bildes innhold er abstrahert. Med abstraksjon menes her prosessen ved, eller resultatet av, å generalisere eller forenkle et objekt (interesseobjekt eller kontekstuelt objekt) fra en fullstendig realistisk avbildning til en enklere avbildning, samtidig som man forsøker å beholde en forståelse (visuell fremstilling) av hva objektet er. For eksempel vil man kunne representere et menneske på en tilnærmet realistisk måte, forenkle det ned til omriss, tegne fyrstikkmennesker eller representere mennesker ved bruk av enkle geometriske former. Her har jeg satt opp følgende kriterier:

- **Geometriske primitiver:** Et *geometrisk primitiv* er en helt enkel geometrisk form (sirkler, ovaler, firkanter). I denne sammenhengen mener jeg bruk av slike primitiver til å representere virkelige objekter - for eksempel å representere en båt ved hjelp av en firkant.
- **Omriss:** Et *omriss* beskriver objekter som er representer ved hjelp av et omriss, enten fylt eller hult.
- **Symbolske visuelle elementer**: Et *symbolsk visuelt element* er en visuell detalj som er svært viktig for representasjonen av et objekt, for eksempel øyne, munn eller lemmer på et menneske; seil, påhengsmotor eller vinduer på en båt; eller greiner på et tre.
- **Detaljerte visuelle elementer:** Et *detaljert visuelt element* er en visuell detalj som ikke har høy symbolsk verdi, men som bidrar til at et objekt får en mer realistisk avbildning. Eksempler på dette kan være individuelle hårstrå, fingrer på en hånd, blader på et tre og tilsvarende.
- **Tekstur**: Er *tekstur* ("mønster") benyttet for å representere overflateegenskapene til et visuelt objekt, for eksempel sjøsprøyt eller varierende farge på fjærene til en måke.

Subjektiv vurdering

Dette er din subjektive vurdering av i hvor stor grad abstraksjon er brukt når et bilde er laget. Settes på en skala fra 1 - 5, der 1 representerer full abstraksjon (For eksempel utelukkende bruk av geometriske primitiver) mens 5 representerer tilnærmet ingen abstraksjon (Objektene i bildet er forsøkt tegnet tilnærmet helt realistiske).

Dybde

Dette representerer i hvor stor grad dybde er brukt for å gi bildet perspektiv og komposisjon. Her har jeg brukt følgende kriterier:

- **Skalering**. Dette representerer hvorvidt de visuelle objektene i bildet er forsøkt skalert korrekt i forhold til hverandre.
- **Overlapp**. Dette representerer hvorvidt objektenes rekkefølge og avstand i bildet er representert ved hjelp av overlapping.
- **Sentralperspektiv**. Dette representerer hvorvidt det er gjort forsøk på bruk av sentralperspektivet for å gi bildet dybde.

Subjektiv vurdering

Dette er din subjektive vurdering av i hvor stor grad dybde og perspektiv er brukt for å oppnå et realistisk bilde. Settes på en skala fra 1 - 5, der 1 representerer ingen bruk av dybde (Ingen grep er gjort for å representere dybde) mens 5 representerer tilnærmet full representasjon av dybde.

Antall unike objekter

I tillegg trenger jeg å få vite hvor mange *selvstendige* objekter det er tegnet inn i hvert bilde. Et selvstendig objekt er et objekt som ikke er del av et annet objekt. For eksempel vil et ansikt/hode være et unikt objekt dersom det er alene, men dersom det er satt sammen med resten av en person, vil personen utgjøre det unike objektet. Det kan være noen grensetilfeller, for eksempel et bilde av en måke med en fisk i nebbet. Selv om disse objektene klart henger sammen, vil jeg at dette regnes som to unike objekter.

Vanskelighetsgrad ved vurdering

Til slutt ønsker jeg at den som evaluerer bildene oppgir hvor vanskelig det var å vurdere hvert enkelt bilde, på en skala fra 1 - 5, der 1 representerer ”svært lett å vurdere” mens 5 representer ”svært vanskelig” å vurdere.

I tillegg er det mulig å føre opp notater dersom det er noe spesielt ved et gitt bilde.

Skjema for vurdering

Jeg har satt opp et skjema for vurdering av bildene. Skjemaet er delt tre hovedområder. Den øverste delen av skjemaet består av selve bildet, verktøy som er blitt brukt for å opprette bildet, og noen detaljer omkring søket. Det viktigste for deg som skal vurdere dette, er selve bildet (til venstre) og ”søketekst” (helt til høyre). Resten av detaljene på den øverste delen av arket er ikke vesentlig for deg - dette skal jeg bruke for å se på en del andre ting. De ikke-aktuelle områdene er ”grået ut”, slik at kun det du skal jobbe med vil være trykket i svart.

Den andre delen av arket består av vurderingen av de ulike kriteriene, samt et felt for å notere hvor mange unike objekter det er på bildet. Først er de ulike kriteriene satt opp etter hovedområder. Her ønsker jeg at du enten markerer de kriteriene du mener er oppfylt - enten ved en fargepenn eller ved å sette ring rundt de aktuelle kriteriene. Videre er det en seksjon der du kan fylle ut din subjektive oppfatning av de ulike modalitetsbegrepene.

Området ”Bruk av former” skal du ikke gjøre noe med - dette er også til intern bruk.

Siste delen av arket består av en skala for ”kompleksitet”. Her fører du opp hvor lett (1) eller vanskelig (5) det var å vurdere bildet. Til slutt er et felt for notater hvor du kan notere eventuelle stikkord eller spesielle ting ved vurderingen av bildet.

Du vil få alle bildene vedlagt på en CD, nummerert på samme måte som i dette heftet. Dersom noen bilder er for små slik at detaljene ikke vises, kan du eventuelt se på bildene i full størrelse ved hjelp av denne.

Eksempelskjema

Bildedetaljer
Bildenummer: 1
Deltakernummer: 1
Deltakersøk: 1
Tegnetid: 30

Søkedetaljer
Søketype: Nivå 1

Søketekst:
Finn bilder av en måke

Verktøybruk

Frihånd:	**JA**	NEI
Former:	JA	NEI
Farger:	JA	NEI
Tekstur:	JA	NEI

Unike objekter __1__

Klassifisering av modalitet - kriterier

Fargebruk:	**Monokrom**	Enkel	Nyansert	Gradering	Lyssetting
Kontekst:	**Objekter**	Bakgrunn	Symbolsk	Detaljert	
Representasjon:	Primitiver	**Omriss**	Symbolsk	Detaljert	Tekstur
Dybde:	Skalering	Overlapp	Sentralperspektiv		

Klassifisering av modalitet – Subjektiv evaluering

Fargebruk:	**1**	2	3	4	5
Kontekst:	**1**	2	3	4	5
Representasjon:	1	**2**	3	4	5
Dybde:	**1**	2	3	4	5

Bruk av former:

Markering av større områder:	JA	NEI
Tegning av geometriske former:	JA	NEI
Representasjon av andre objekter:	JA	NEI

Kompleksitet	1	2	**3**	4	5

Notater:

Litt vanskelig å avgjøre hvorvidt avbildningen av en måke på denne måten er en abstraksjon eller realistisk. Det er jo på mange måter veldig stilistisk, samtidig vil man ofte se at måker har dette utseendet i luften, sett fra avstand.

References (Temp. placement)

Abdullah, R. and R. Hübner (2006). Pictograms, Icons & Signs. London, Thans and Hudson.
Ardizzone, E. and M. La Cascia (1996). Jacob: Just a Content-Based Query System for Video Databases. IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, Georgia, IEEE.
Baeza-Yates, R. and B. Ribeiro-Neto (1999). Modern Information Retrieval. New York, Addison Wesley.
Benitez, A. B., M. Beigi, et al. (1998). "Using relevance feedback in content-based image metasearch." IEEE Internet Computing **2**(4): 10.
Berman, A. and L. Shapiro (1999). "A flexible image database system for content-based retrieval." Computer Vision and Image Understanding **75**(1): 175-195.
Bonhomme, C., C. Trépied, et al. (1999). A visual language for querying spatio-temporal databases. 7th ACM international symposium on Advances in geographic information systems, Kansas City, Missouri, United States, ACM Press New York, NY, USA.
Bryne, S. (2008). "Ny Posten-logo vekker assosiasjoner." Retrieved Sep 8th, 2008, from http://www.dagbladet.no/kultur/2008/09/08/546177.html.
Carson, C. (1997). Region Based Image Query. Proceedings of IEEEE, CVPR '97, Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, Calefornia.
Carson, C. and V. E. Ogle (1996). "Storage and Retrieval of Feature Data for a Very Large Online Image Collection." IEEE Computer Society Bulletin of the Technical Committee on Data Engineering **19**.
Carson, C., M. Thomas, et al. (1999). Blobworld: A system for region-based image indexing and retrieval. VISUAL '99, Amsterdam, The Netherlands, Springer.
Choi, Y. and E. M. Rasmussen (2003). "Searching for Images: The analysis of Users' Queries for Image Retrieval in American History." Journal of the American Society of Information Science and Technology **54**(6): 498-511.
Ciocca, G., I. Gagliardi, et al. (2001). "Quicklook2: An Integrated Multimedia System." Journal of Visual Languages & Computing **12**(1): 22.
CogniSearch. (2007, March 23rd, 2008). "Xcavator - Photo Search for Professionals." Retrieved March 9th, 2008, from http://www.xcavator.net/.
Cox, I. J., M. L. Miller, et al. (2000). "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments." IEEE transactions on image processing **9**: 17.
Cunningham, S. J., D. Bainbridge, et al. (2004). How people describe their image information needs: a grounded theory analysis of visual arts queries. 4th ACM/IEEE-CS joint conference on Digital libraries, Tuscon, AZ, USA, ACM Press New York, NY, USA.
Cunningham, S. J. and M. Masoodian (2006). Looking for a Picture: An Analysis of Everyday Image Information Searching. JCDL'06, Chapel Hill, North Carolina, ACM.
Das, M., E. M. Riseman, et al. (1997). Focus: Searching for multi-colored objects in a diverse image database. IEEE Conference on Computer Vision and Pattern Recognition.
Datta, R., D. Joshi, et al. (2008). "Image Retrieval: Ideas, Influences, and Trends of the New Age." ACM Computing Surveys **40**(2): 60.
Del Bimbo, A., M. Mugnaini, et al. (1997). Picasso: Visual Querying by Color Perceptive Regions. 2nc International Conference on Visual Information Systems, San Diego.
Deselaers, T., D. Keysers, et al. (2008). "Features for Image Retrieval: An Experimental Comperison." Information Retrieval **11**(2): 77-107.
Dix, A., J. Finlay, et al. (2004). Human-Computer Interaction. Harlow, Pearson Education Limited.

Eakins, J. P., B. Burford, et al. (2003). "A taxonomy of the image: on the classification of content for image retrieval." Visual Communication **2**(2): 123-161.
Eakins, J. P. and M. E. Graham (1999). Content Based Image Retrieval: A report to the JISC Technology Applications Program. Newcastle, Inst. for Image Data Research, University of Northumbria.
Edwards, B. (1999). The New Drawing on the Right Side of the Brain. New York, Penguin Putnam.
Egeland, E. (2007). Brukergrensesnitt for søk i bildedatabaser, Telenor R&I.
Egenhofer, M. J. (1997). "Query Processing in Spatial-Query-By-Sketch." Journal of Visual Languages and Computing **8**(4): 403-424.
Engan, K. and K. F. Fretheim (2004). Compression of Digital Mammographs Without Interering With A System for Automatic Detection of Microcalcifications - Mammoscan μCaD. Norwegian Conference on Image Processing and Pattern Recognition. Stavanger, Norway, NOBIM.
Enser, P. G. B. (1993). "Query Analysis in a visual information retrieval context." Journal of Document and Text Management **1**: 25-52.
Enser, P. G. B. and C. Sandom (2003). "Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval." Lecture Notes in Computer Science **2728/2003**: 5.
Faloutsos, C., R. Barber, et al. (1994). "Efficient and effective querying by image content." Journal of Intelligent Information Systems **3**: 231-262.
Flickner, Sawhney, et al. (1995). "Query by Image and Video Content: The QBIC System." IEEE Computer **28**(9): 23-32.
Foskett, D. J. (1997). Thesaurus. Readings in information retrieval. K. S. Jones and P. Willet, Morgan Kaufmann Publishers**:** 111-134.
Geertz, C. (1972). Thick description: Toward an interpretive theory of culture. The interpretation of cultures. Selected essays. C. Geertz, Harper Collins Publisher**:** 3-30.
Gentikow, B. (2005). Hvordan utforsker man medieerfaringer - Kvalitativ metode. Kristiansand, ij-forlaget.
Gevers, T. and A. Smeulders (2000). "Pictoseek: Combining color and shape invariant features for image retrieval." IEEE Transactions on Image Processing **9**(1): 102-119.
Hartvedt, C. (2008). Smashing Reality (in)to Bits? Views on representation and interpretation in Image Retreival. http://caim.uib.no/publications/CH_Vit-teori_essay.pdf, Dept. of Information Science and Media Studies, University of Bergen.
Hermitage. (2003, 2003). "QBIC Colour and Layout Searches." Retrieved April 2nd, 2008, from http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English.
Hove, L.-J. (2004). Extending Image Retrieval Systems with a Thesaurus for Shapes. The Norwegian Information Technology Conference, Stavanger, Norway, NIK.
Hove, L.-J. (2004). Improving Image Retrieval with a Thesaurus for Shapes. Department of Information Science and Media Studies. Bergen, University of Bergen**:** 105.
Hove, L.-J. (2007). Evaluating Use of Interfaces for Visual Query Specification. NOKOBIT 2007, Oslo, Tapir Akademisk Forlag.
Hove, L.-J. (2007). Understanding and Interpreting Visual Content - Visual Grammar and Content Based Image Retrieval. Bergen, University of Bergen, Department of Information Science and Media Studies.
Huang, T. S. and Y. Rui (1999). "Image Retrieval: Current Techniques, Promising Directions And Open Issues." Journal of Visual Communication and Image Representation **10**(4): 39-62.
Idée. (2008). "Piximilar." Retrieved May 24th 2009, 2009.
Idée. (2009). "TinEye - Image Search Engine." 2009.
Idée. (2009). "TinEYE FAQ." Retrieved August 28th, 2009, from http://tineye.com/faq#count.
Iivarinen, J., R. Rautkorpi, et al. (2004). "Content-Based Retrieval of Surface Defect Images with PicSOM." International Journal of Fuzzy Systems **6**(3): 160-166.
Iqbal, Q. and J. K. Aggarwal (2002). CIRES: A System for Content-based Retrieval in Digital Image Libraries. International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore.

Jacobs, C., A. Finkelstein, et al. (1995). "Fast Multiresolution Image Querying." Proceedings of SIGGRAPH 95: 9.

Jaimes, A. and S.-F. Chang (2002). Concepts and Techniques for Indexing Visual Semantics. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and L. D. Bergman, John Wiley & Sons, Inc.**:** 497-565.

Jørgensen, C. and P. Jørgensen (2005). "Image Querying by Image Professionals." Journal of the American Society of Information Science and Technology **56**: 1346-1359.

Jose, J. M., J. Furner, et al. (1998). Spatial querying for image retrieval: a user-oriented evaluation. SIGIR'98, Melbourne, Australia, ACM.

Joystiq.com. (2008). "Norway Post's new logo: Pokéball or Marathon symbol variation?" Retrieved Sep 5th, 2008, from http://www.joystiq.com/2008/09/05/norway-posts-new-logo-pokeball-or-marathon-symbol-variation/.

Karlsen, R. and J. Nordbotten (2008). Utilizing Information Collection Information to Capture Image Semantics. ICIW'08 - 3rd Int'l Conf. on Internet and Web Applications and Services. Athens, Greece.

Keister, L. H. (1994). User Types and Queries: Impact on Image Access Systems. Challenges in Indexing Electronic Text and Images. R. Fidel, T. Bellardo Hahn, E. M. Rasmussen and P. J. Smith. Medford, NJ, Information Today**:** 7-22.

Kelly, P. M., M. Cannon, et al. (1995). Query by image example: the CANDID approach. Storage and Retrieval for Image and Video Databases, SPIE.

Kherfi, M. L., D. Ziou, et al. (2004). "Image Retrieval from the World Wide Web: Issues, Techniques and Systems." ACM Computing Surveys **36**(1): 35-67.

Kress, G. and T. van Leeuwen (2006). Reading Images - The Grammar of Visual Design. Oxon, Routledge.

Kreyss, J., M. Röper, et al. (1997). Video retrieval by still image analysis with ImageMiner. IS&T/SPIE's Symposium on Electronic Imaging: Science & Technologies, San Jose, CA.

Lai, T.-S., S. McDonald, et al. (1999). Image Browsing and Navigation using Hierchical Classification. The Challenge of Image Retrieval 1999 conference, CIR'99, Newcastle upon Tyne, England, Springer-Verlag.

Langreiter, C. (2006). retrievr, System One Labs.

Lee, J., S. Y. Jeong, et al. (2004). "Image Navigation: A Massively Interactive Model for Similarity Retrieval of Images." International Journal of Computer Vision **56**(1/2): 131-154.

Lew, M. (2000). "Next Generation Web searches for visual content." IEEE Computer **33**(11): 46-53.

Lew, M., D. P. Huijsmans, et al. (1996). Content based image retrieval: KLT, projections, or templates. IDB-MMS'96, Amsterdam, The Netherlands, Amsterdam University Press.

Lew, M., N. Sebe, et al. (2006). "Content-Based Multimedia Information Retrieval: State of the Art and Challenges." ACM Transactions on Multimedia Computing, Communications and Applications **2**(1): 1-19.

Li, Y. and C. C. J. Kuo (2002). Introduction to Content-Based Image Retrieval - Key Techniques. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and L. D. Bergman. New York, John Wiley & Sons**:** 261-284.

Li, Z. N., O. R. Zaïane, et al. (1998). C-bird: Content-based image retrieval from image repositories using chromaticity and recognition kernel, School of Computing Science, Simon Fraser University, Canada.

Like.com. (2009). "Like.com - Shop Visually." Retrieved May 24th, 2009, from http://www.like.com/.

Lim, J.-H. (2000). Explicit Query Formulation with Visual Keywords. ACM Multimedia 2000, Los Angeles, CA, UAS, ACM Multimedia.

Manjunath, B. S. and W.-Y. Ma (1999). "Netra: A toolbox for Navigating large image databases." Multimedia Systems **7**(3): 184-198.

Manjunath, B. S. and W.-Y. Ma (2002). Texture Features for Image Retrieval. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and R. Baeza-Yates, John Wiley & Sons, Inc**:** 313-344.

Manjunath, B. S., U. Moenich, et al. (2004). Cortina: A System for Largescale, Content based Web Image Retrieval. ACM Multimedia, New York, ACM.
McDonald, S. and J. Tait (2003). Search Strategies in Content-Based Image Retrieval. SIGIR '03, Toronto, Canada, ACM.
McDonald, S., J. Tait, et al. (2001). Evaluating a Content Based Image Retrieval System. SIGIR '01, New Orleans, Louisiana, USA, ACM.
Moe, R. (2006). Expressive Models and Species of Language. The XVI international conference on computer and information science and engineering. **16**.
Mukherjea, S., K. Hirata, et al. (1997). Towards a multimedia World-Wide Web information retrieval engine. Selected papers from the sixth international conference on World Wide Web, Santa Clara, California, United States, Elsevier Science Publishers Ltd.
Næss, B. (2007). The VISI Prototype for Image Retrieval. Bergen, University of Bergen, Dept. of Information Science and Media Studies.
Nastar, C., M. Mitschke, et al. (1998). Surfimage: A flexible content-based image retrieval system. ACM International Multimedia Conference, Bristol, England.
Niblack, W. R., R. Barber, et al. (1993). The QBIC Project: Querying Images by Content Using Color, Texture and Shape. Storage and Retrieval for Image and Video Databases, San Jose, California, USA, SPIE.
Nordbotten, J. (2005). Virtual Exhibits - Theory, Methods and Tools for Development of Virtual Exhibits on Demand. Bergen, Department of Information Science and Media Studies, University of Bergen.
Octagon. (2007, February 2nd 2007). "Octagon - Content Based Image Retrieval Software." Retrieved March 28th, 2008, from http://octagon.viitala.eu/.
Ogle, V. E. and M. Stonebraker (1995). "Chabot: Retrieval from a relational database of images." IEEE Computer **28**(9): 9.
Ortega, M., Y. Rui, et al. (1997). Supporting similarity queries in MARS. 5th International Multimedia Conference, Seattle, Washington, USA, ACM.
Pentland, A., R. W. Picard, et al. (1996). "Photobook: Content-based manipulation of image databases." International Journal of Computer Vision **18**(3): 233-254.
Petraglia, G., M. Sebillo, et al. (2001). "Virtual Images for Similarity Retrieval in Image Databases." IEEE Transactions on Knowledge and Data Engineering **13**(6).
Pilevar, A. H. (2008). "Multi-Feature Image Retrieval System." Retrieved April 2nd, 2008, from http://www.pilevar.com/mfirs/.
Rodden, K. (1999). How Do People Organise Their Photographs? 21st BCS IRSG Colloquium on IR, Glasgow.
Rosenberg, C. (2009, 4/20/2009). "Hard at play in Google Labs with Similar Images and Google News Timeline." The Google Blog Retrieved May 24th, 2009, from http://googleblog.blogspot.com/2009/04/hard-at-play-in-google-labs-with.html.
Rowley, H. and S. Baluja (1998). "Neural Network Based Face Recognition." Pattern Analysis and Machine Intelligence **20**: 23-38.
Shatford, S. (1986). "Analyzing the Subject of a Picture: A Theoretical Approach." Cataloging and Classification Quarterly **6**: 39-62.
Siggelkow, S. (2001). SIMBA - Search IMages By Appearance. DAGM, LNCS 2191, München.
Sirihari, R., Z. Zhang, et al. (2000). "Intelligent indexing and semantic retrieval of multimodal documents." Information Retrieval **2**(2): 30.
Sivic, J. and A. Zisserman (2006). Video Google: Efficient Visual Search of Videos. Toward Category-Level Object Recognition. J. Ponce, M. Hebert, C. Schmid and A. Zisserman, Springer. **4170:** 127-144.
Smith, J. R. and S.-F. Chang (1995). Automated image retrieval using color and texture, Columbia University.
Smith, J. R. and S.-F. Chang (1997). Querying by color regions using the VisualSEEk content-based visual query system. Intelligent Multimedia Information Retrieval. M. T. Maybury, AAAI Press.

Sormunen , E. and M. Markkula (2000). "End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive." Information Retrieval **4**: 259-285.
Strauss, A. L. and J. Corbin (2008). Basics of Qualitative Research. London, Sage Publications, Inc.
SystemOneLabs. (2009). "Rate Sketches - Retrievr." 2009, from http://labs.systemone.at/retrievr/rate.
Tiltomo. (2006, 2006-03-08). "Tiltomo." Retrieved March 5th, 2008, from http://www.tiltomo.com/.
Trovåg, A. (2004). Beyond the Divide: Investigating the Expressive Convenience of a Generalised Division Operator in Relational Algebra. Department of Information Science and Media Studies. Bergen, University of Bergen. **Candidatus rerum politicarum:** 173.
Veltkamp, R. C. and M. Tanase (2000). Content-Based Image Retrieval Systems: A Survey. Utrecht, Universiteit Utrecht, Center for Geometry, Imaging and Virtual Environments**:** 49.
Vendrig, J., M. Worring, et al. (1999). Filter image browsing: Exploiting interaction in image retrieval. VISUAL '99, Amsterdam, The Netherlands, Springer.
Venters, C. C. and M. D. Cooper (2000). A Review of Content-Based Image Retrieval Systems. JISC Technology Applications Programme. Manchester, University of Manchester.
Venters, C. C., R. Hartley, et al. (2001). Query by Visual Example: Assessing the Usability of Content-Based Image Retrieval System User Interfaces. Second IEEE Pacific-Rim Conference on Multimedia, Beijing, Republic of China, Springer-Verlag Berlin.
Volmer, S. (1997). Tracing images in large databases by comparison of wavelet fingerprints. 2nd International Conference on Visual Information Systems, San Diego.
von Ahn, L. and L. Dabbish (2004). Labelling Images with a Computer Game. ACM Conference on Human Factors in Computing Systems, Vienna, Austria, ACM.
von Ahn, L., L. Ruoran, et al. (2006). Peekaboom: A Game for Locating Objects in Images. ACM Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, ACM.
Wang, J. Z., G. Wiederhold, et al. (1997). Wavelet-based image indexing techniques with partial sketch retrieval capability. The fourth forum on Research and Technology Advances in Digital Libraries, Washington D.C.
Ward, R. (2001). "Oracle interMedia User's Guide and Reference." Retrieved August, 2004, from http://download-west.oracle.com/docs/cd/B10501_01/appdev.920/a88786/title.htm.
Westman, S. and P. Oittinen (2006). Image retrieval by end-users and intermediaries in a journalistic work context. 1st international conference on Information interaction in context, Copenhagen, Denmark, ACM.