Automatic Facial Anonymisation Using Average Face Morphing

An explorative attempt at designing a tool for investigating the normative views of journalism practices in dealing with manipulation and usage of facial press photographies.

Author

Joar Midtun

Supervisor

Bjørnar Tessem

Department of Information Science and Media Studies

University of Bergen

Norway

11.05.2017

Abstract

This thesis describes the process of developing and evaluating a design science research artefact for anonymising faces in images. The artefact is developed for use in RRI-research in the ViSmedia research group at the University of Bergen. A facial anonymisation algorithm is defined as a system which can provide facial de-identification, while retaining face realism and preserving image quality. Research from psychology and artificial intelligence serve as the theoretical background for the development, where human face perception, computer vision and machine learning are topics of interest. Face morphing techniques are used to manipulate faces according to the facial anonymisation definition.

The thesis aims to answer three questions. First, it is asked if it possible to to create an artefact capable of automatically anonymising faces using an average face morphing approach. Secondly, if the artefact can be used to investigate the views on normative practices in manipulation and usage of face photographies in the news media. Finally, it is investigated if there are other possible use cases for the artefact or its technology.

Two prototypes are proposed for solving the anonymisation problem. Based on findings from a preliminary development prototype, both a featural and a holistic anonymisation prototype is developed and evaluated. The evaluations are a combination of performance The featural prototype performs poorly on de-identification, while the holistic prototype shows more promise. Both are decent at providing realistic faces and in image quality preservation. The facial anonymisation algorithm is argued to have potential as an investigative tool, and several potential use cases are proposed.

The artefact in its current state is but a proof of concept and its limitations and challenges are described. Several proposals for future development are also provided in order to produce a more complete solution.

Acknowledgements

A sincere gratitude is due to the people who have assisted me during the past twelve months. I would like to offer a specific thanks to Bjørnar Tessemfor both providing me with the thesis idea, as well as for providing me with valuable feedback and discussion. Thanks are also due to the rest of the Prosopo work group, where both Lars Nyre and Simen Skaret Karlsen have been of great help. Finally, I would like to offer my gratitude to the ViSmedia research group, and Astrid Gynnild in particular, for the opportunities I have been offered as part of an academic community.

I would also like to thank my father, Øyvind Heggen Eriksen, and my brother, Torbjørn Midtun, for proofreading and scrutinising my writing in the latter stages of the project. Their assistance was very welcome in completing the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter will briefly introduce some key concepts of this thesis, while providing a short introduction to facial anonymisation. The problem scope will be defined and the background and origin of the thesis will be presented. Additionally, an outline of the project will be given.

1.1 Facial Anonymisation

The term ”facial anonymisation” does not stand out as a clear-cut academic problem, and neither is it a specific and established problem in any given discipline. Facial anonymisation can look at anonymising faces for both humans and computers. Previous work has perhaps been more prevalent in the latter case, concerned with privacy issues like automatic tagging and identification services in online social media databases. What are the criteria for a successfully anonymised face? This all depends on the aim of a facial anonymisation algorithm, and in which domain it operates. As a baseline, such an algorithm has to meet some sort of criteria to obfuscate or de-identify faces in images.

1.1.1 Example Applications

The abundance of social media images and traditional media content have already drawn interest from de-identification projects, which share some of the same goals as this project. Driessen and Dürmuth (2013) developed k-anonymisation algorithms in order to keep face recognition algorithms from completing automatic tasks which can threaten social media users, while still allowing for human recognition. Attempting to achieve k-anonymity, where a person is k-anonymous if face recognition algorithms fails to reduce the possible set of people to k persons, has been one way of combating privacy issues in social media content (Newton, Sweeney and Malin 2005). This project will not attempt k-anonymisation in the same regard, but rather attempt to

allow for recognition of the manipulated faces, but without the possibility to identify the original owner.

A similar approach can be found in Korshunov and Ebrahimi (2013). By using a morphing-based visual privacy protection algorithm combined with recognition, they worked towards providing a reversible, flexible and robust morphing algorithm (section 2.2.4 will introduce the concept of morphing). The algorithm relies on interpolation and triangulation, which they claim to be more suitable for achieving privacy protection, opposed to blurring, pixelation and masking filters. However they had not been able to evaluate their results with human participants, but relied on their obtained objective results. Their approach is similar to the one of this thesis, but where manipulating the facial appearance without altering its compositional structure is the key objective.

1.1.2 Definition

In order to create a foundation for this thesis, it will be important to work with a clear and unambiguous problem definition. This definition will serve as the groundwork for the development and evaluation of the artefacts involved in reaching a solution. The thesis title, Automatic Facial Anonymisation Using Average Face Morphing, can be broken down into three essential components.

Automatic To fulfil the goal of automation, the algorithms have to work without correction or additional help after its input has been provided. Given an image and a level of anonymisation, the system has to create an output that successfully meets the requirements of facial anonymisation. One could also talk of completeness, meaning that for all input images containing faces, this automatic algorithm must produce an output which meets the defined requirements introduced below.

Facial Anonymisation There are two important goals for the anonymisation artefact. Where facial anonymisation previously has been described strictly as a de-identification algorithm for computer perception purposes, this thesis will aim to deal with human face perception. While the goal of de-identification remains as one criteria, another goal will be to maintain the perception of a realistic face after anonymisation has taken place. This final point also includes preserving the overall quality of the image itself, limiting the visible effects of manipulation to a minimum. Successful facial anonymisation is thereby defined by the two following criteria.

- De-identification of the face
- Preserving face realism and image quality

Using Average Face Morphing The technology that this thesis will rely on, is a series of morphing technologies, and particularly the concept of face averaging. The thesis will use this approach to attempt reach a solution in facial manipulation which can address the requirements as stated by the facial anonymisation definition. Average faces and face morphing, and the different concepts involved, will be discussed in depth in section 2.2.

1.2 Background

This master thesis was developed from an idea by my supervisor Bjørnar Tessem into the facial anonymisation problem which has now been defined. This idea was presented to me as a problem which had relevance to the research group ViSmedia at the University of Bergen. The solution was exploratively and iteratively developed based on existing research within information science, computer science and psychology, aiming to become an artefact facilitating research carried out in relation to ViSmedia, specifically directed at news media. However, the thesis will also aim at exploring the utility of the artefact outside the news media scope.

1.2.1 ViSmedia

ViSmedia is an international interdisciplinary collaboration with its origin from The Department of Information Science and Media Studies at the University of Bergen. The project is connected to NCE Media (National Centre of Excellence) and the media cluster Media City Bergen. ViSmedia has members from Norway, Finland and USA from the University of Bergen, the University of Virginia, the University of Austin in Texas, the University of Maryland and the University of Jyväskylä; the disciplines involved being information science, media studies, journalism, geography, communications and philosophy. ViSmedia uses the RRI-framework (Responsible Research and Innovation) developed by the EU, which aims to examine opportunities and dilemmas of adoption, innovation, and use of new visual surveillance technologies in the news media (ViSmedia 2017).

1.2.2 Responsible Research and Innovation

"Responsible Research and Innovation (RRI) implies that societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society." (The EU Framework Programme for Research and Innovation 2017a)

RRI is implemented as a package involving multi-actors and public engagement in research and innovation. The aim is to enable easy access to scientific results, admission of gender and ethics in research and innovation content and processes, and formal and informal science education. Inter- and transdisciplinary solutions are to be developed, addressing the different objectives of Horizon 2020. Horizon 2020 the biggest EU Research and Innovation programme in history with €80 billion of funding available over 7 years from 2014 to 2020 (The EU Framework Programme for Research and Innovation 2017b). The programme objectives encourages work to be dedicated to thematic elements of RRI, as well approaches to promote RRI uptake (The EU Framework Programme for Research and Innovation 2017a).

1.2.3 Prosopo

One of the ViSmedia work packages is responsible for producing prototypes and concepts for research. It is in this work package that the Automatic Facial Anonymisation Using Average Face Morphingartefact was to be employed into Prosopo, a sub-group within ViSmedia working on creating RRI-prototypes for research purposes. One of the solutions which Prosopo has developed is a web service, as an application programming interface (API), offering advanced computer vision functionality. These services include face manipulation, facial analysis functionality and custom face databases. The API is available to projects, within the ViSmedia group, which could benefit from utilising such technology.

1.2.4 Facial Anonymisation in Journalism

The process of face anonymisation as a de-identification problem was briefly discussed in section 1.1. Providing ways of anonymisation or identity protection in news media content production is crucial in order to conserve civil privacy rights. How to do this is, however, not always straight forward, and might depend on the severity of a story or subject to the status or desires of those involved in a story.

Claiming that the photography should be a realistic representation of the world, is perhaps not a controversial statement to members of the journalistic domain. The National Press Photographers Association provides a code of ethics (see figure 1.1) which they describe as follows: "This code is intended to promote the highest quality in all forms of visual journalism and to strengthen public confidence in the profession. It is also meant to serve as an educational tool both for those who practice and for those who appreciate photojournalism." (NPPA 2017)

Points 6 and 7 from figure 1.1 are highly relevant for this thesis. Both points refer to editing and/or manipulation of an image. Journalists also have a responsibility to protect both their sources and individuals which for some reason should not

Visual journalists and those who manage visual news productions are accountable for upholding the following standards in their daily work:

1. Be accurate and comprehensive in the representation of subjects.
2. Resist being manipulated by staged photo opportunities.
3. Be complete and provide context when photographing or recording subjects. Avoid stereotyping individuals and groups. Recognize and work to avoid presenting one's own biases in the work.
4. Treat all subjects with respect and dignity. Give special consideration to vulnerable subjects and compassion to victims of crime or tragedy. Intrude on private moments of grief only when the public has an overriding and justifiable need to see.
5. While photographing subjects do not intentionally contribute to, alter, or seek to alter or influence events.
6. Editing should maintain the integrity of the photographic images' content and context.
7. Do not manipulate images or add or alter sound in any way that can mislead viewers or misrepresent subjects.
8. Do not accept gifts, favors, or compensation from those who might seek to influence coverage.
9. Do not intentionally sabotage the efforts of other journalists.

Figure 1.1: The National Press Photographers Association Code of Ethics (NPPA 2017).

be publicly identified. This has implications for the photojournalist. Either the person in question cannot be depicted, or it has to be de-identified in some manner. This de-identification might require manipulation or editing of the image. Some common techniques for de-identification are pixelation and blurring. See figure 1.2 for examples.

1.3 Problem Scope

This section will discuss the problem domain, artefacts and goals of the research.

1.3.1 Domain

Both the goal of automation and the current state of the chosen technology has implications for the scope of a possible solution. The goal will be to create a proof of concept artefact which can perform fairly well within a limited predefined problem domain. Certain challenges will be time consuming and complicated to create robust and generalisable solutions for, and in these cases there will be provided argumentation for any workaround or exclusion required to maintain feasibility. Ways

Figure 1.2: Example of de-identification techniques. From the left: blurring, pixelation, solid.

of dealing with these challenges, as future work, will be investigated and proposed.

Another aspect of the problem domain, is the field in which the technology can be of academic value or possibilities as specific business problem solutions. The research will follow a set of guidelines, called design science research, which will ensure its validity and utility as a problem solving artefact. This essentially means the creation of a novel artefact followed by the study of its usage in a given domain. As a part of the ViSmedia work group the thesis will investigate facial anonymisation in the news media, but will also attempt to identify other domains where there could be potential for implementing such technology.

1.3.2 Artefacts

As a design science research project, there will be created several artefacts as a result of the development phase of this project. These artefacts will be listed and briefly explained, and will serve as references in the coming chapters. The artefacts will be presented further in chapter 4 in terms of their technical constituents.

Artefact 1: Facial Anonymisation Algorithm The main artefact is the facial anonymisation algorithm, and whenever ”the artefact” is mentioned in this thesis, assume that it is this artefact that is referred to. The facial anonymisation algorithm has also gone under the name of ”Prosopo 1”.

Artefact 2: Prosopo API The API, which is as mentioned a ViSmedia and Prosopo artefact, provides the functionality of the main artefact as a web-service. The Prosopo API is not directly a part of this master thesis, but is developed as foundation for all future projects for Prosopo and will continue to be a service

provider for projects beyond the facial anonymisation artefact.

Artefact 3: Prosopo Android The Android-application simply aims to demonstrate the main artefact. It is an implementation that allows the user to provide the parameters for the facial anonymisation artefact and then visualise the output. Prosopo Android communicates with the Prosopo API and reaches the functionality in the same way other projects would in order to take advantage of the possibilities provided by the API.

1.3.3 Research Questions

There are three questions that this thesis will aim to answer. The three questions are categorised as: the artefact as an anonymisation technology, the artefact as an investigative tool, and identifying use cases for the artefact. More specifically the questions are formulated as follows:

Q1: Is it possible to create an artefact capable of automatically anonymising faces using an average face morphing approach?
Q2: Can the artefact be used to investigate the views on normative practices in manipulation and usage of face photographies in the news media?
Q3: Are there other possible use cases for the artefact or its technology?

It has to be emphasised that this project is highly explorative and as such should be considered to successfully answer these questions as a proof of concept. Further work can expand the limited problem domain, building the artefact into a more complete solution.

Chapter 2

Literature Review

This chapter will present and discuss the theory and concepts which will be relevant to the design and development of the artefact. The chapter is divided into two sections. The first section will discuss the research background, i.e. the academic disciplines which are key to understanding the problem, and key to solving it. The second section will go into depth on specific concepts and technologies which are to be implemented in the solution.

2.1 Research Background

There are three main academic fields which are immediately interesting. In order to understand the requirements of facial anonymisation it is interesting to look into the nature of human face perception. Furthermore, computer vision is essential to work with faces as data from images, and machine learning will be key in order to provide automatic and precise processing in the utilisation of this data.

2.1.1 Face Perception

The perception of human faces has a unique position in stimuli perception. Most humans have the remarkable ability to instantly recognize people which are familiar to them, in countless variations like differing orientation, illumination and with obscured vision. Where humans use the combination of individual features in object recognition, it is an holistic experience that allows us to recognize faces. This is illustrated, for instance, in the face-inversion effect, which shows the ease of recognition in upright faces versus upside-down faces. This effect can be seen in figure 2.1. Additionally, facial features are easier to recognize in the context of the whole face, rather than as an isolated feature alone. Neuroscience has also shown how these seemingly similar recognition tasks are at least somewhat distinct. Humans suffering from prosopagnosia retain the ability to recognize objects, but are not able

Figure 2.1: Face inversion effect (Little, Jones and DeBruine 2011).

to produce the holistic recognitions of a face. They can identify the facial features like a nose, ear or eyes, but do not have the ability to combine them into a whole face (Matlin 2014).

Through applied research, psychologists have been able to show that people are less accurate in recognition when dealing with unfamiliar faces in real life settings. For instance, in a situation of identification verification, determining if a photo contains the same face as one on an ID-card presents challenges. Similar experiments when matching whether photos of unfamiliar people were present in a previously shown video, showed that matching accuracy was poor. This has opened the question whether humans are able to accurately match two images of a single face (Matlin 2014).

Experiments by Goffaux and Rossion (2006) have argued that holistic recognition is preliminary based on coarse visual information, represented by low spatial frequencies, contrasted by high spatial frequency, which represent fine detail visual information. The differences in these spatial frequencies are illustrated in figure 2.2. This highlights the need for an approach which not only looks to the features isolated, but one that is conscious of the complexities involved in human face perception.

Another well known psychological phenomena in face perception is the cross-race effect. Humans are better at recognizing faces from their own race, but experience with unfamiliar faces improve facial recognition over time. Holistic processing is dominant in same-race recognition, while featural processing are more common when dealing with unfamiliarity (Tanaka, Kiefer and Bukach 2004). This further illustrates the importance of holistic processing as a cornerstone in understanding how human face perception works.

An interesting curiosity comes from the field of aesthetics within robotics called the uncanny valley. The concept is derived from psychology, where the uncanny means strangely familiar. The worry is that by achieving a high level of realism in

Figure 2.2: Spacial frequencies. From the left: normal spatial frequency, low spatial frequency, high spatial frequency (Feusner et al. 2007).

an artificial human face, but still maintaining a look of incoherence, this will result in a feeling of repulsion (Tinwell 2014).

2.1.2 Computer Vision

Computer vision or, computer perception in general, provides computational agents with information about its world through some sort of sensoring system. This is done by measuring relevant aspects which can be used as an input in some system (Russell and Norvig 2014).

Computer vision is key in order to recognize faces computationally, and creating faces as manipulable data. The Handbook of Face Recognition (Li and Jain 2011) provides background on these specific methods and also contains step-by-step algorithms for facial recognition. This book is a resource for research and professional work within computer vision, biometrics and image processing. This will provide the project with a resource on the concepts which are required to produce reliable working data sets and when extracting facial features or image attributes. The open source project OpenCV, a cross-platform library for real-time image processing, has support for Eigenfaces, Fisherfaces and local binary pattern histograms, which are all well documented algorithms for face recognition (*Face Recognition with OpenCV* 2016). OpenCV is the foundation for this project in terms of access to the application of computer vision theory and will be discussed further in section 3.3.2.

O'Toole (2011) describes some of the connections between human and computer perception, containing a comparative overview, comparisons in face recognition algorithms and evaluation of machine skill in context of human skill. The use of computational models for understanding human face perception and recognition is in many ways the same utilised to develop algorithms for computer-based face recognition systems. Certain challenges that still remain in the field are discussed. These include challenges to achieve robustness in terms of changes in the viewing

conditions, and particularly how representations change as faces become familiar. This highlights the importance of considering human face perception theory in order to find the angle of approach to the facial anonymisation artefact.

2.1.3 Machine Learning

Machine learning is the topic of having computational agents do learning. If an agent improves its performance on future task execution based on observations made in its world, then the agent is learning. By allowing for agents to use learning algorithms, we are able to achieve complex tasks which are trivial for humans, but hard to implement for a programmer (Russell and Norvig 2014).

Computer vision and machine learning are often combined to increase the possibilities of pure computer perception. Where computer vision is required to observe what is going on, machine learning is responsible for understanding it, or putting it into a semantic context. Machine learning algorithms embedded in computer vision allows for teaching the computer how to recognize faces as well as being able to predict facial landmarks coordinates (Takahashi et al. 2006). This might be transferable to other challenges which could arise during development, and it is this combination that is the foundation for working with, and processing, faces as data.

Where computer vision will allow for creating data, specialised machine learning algorithms will account for the analysis and processing of facial attributes. An example of usage is that of Surynek and Lukšová (2011), which involves extraction of suitable attributes from a bitmap image such as contrast, histogram, the occurrence of straight lines in order to provide classification. A decision tree classifier would in this case separate the images by content, based on natural language descriptions. Substituting the extraction attributes which are important for understanding entire bitmaps, extracting specific facial features, or possibly holistic information, would be necessary in order to produce meaningful data for this thesis.

Another interesting technology is one used in Karungaru et al. (2006). By using genetic algorithms to guide a control five-point extraction method, it is possible to perform automatic morphing between two faces. The five points are extracted based on facial features, followed by a triangulation method for matching the warp process of the two images and a colour interpolation. The genetic algorithm aids the process to overcome variations in size and orientation. Zanella and Fuentes' (2004) work involves morphing a source face into a generic predetermined face. This method allows for automatic morphing of face images, using evolutionary strategies (another topic in artificial intelligence), to a simple parameterised face model. However, all images of faces were in frontal view and with the same illumination conditions.

These morphing approaches hold potential when concerned with the coarse scale

manipulation which was mentioned as a desirable approach in section 2.1.1. Being able to work with the entire face, and by combining existing faces into new faces, the manipulation will hopefully be better suited to address human specific challenges, compared to computer-based approaches, such as those mentioned in section 1.1.1.

2.2 Concepts

This section will present and explain the computational concepts and technologies which are used as part of the facial anonymisation process. These concepts will for ease of reference be referred to as facial morphing technologies.

2.2.1 Face Detection/Recognition

"The recognition of human faces is not so much about face recognition at all – it is much more about face detection! It has been proven that the first step in automatic facial recognition – the accurate detection of human faces in arbitrary scenes, is the most important process involved. When faces can be located exactly in any scene, the recognition step afterwards is not so complicated anymore." (Frischholz 2017)

The above quote is taken from Dr. Robert Frischholz' preface on his face detection and recognition information website. It accurately describes the face recognition problem. Li and Jain (2011) further describes the face recognition as a pattern recognition problem. A face can be represented as a three-dimensional object with characterised by varying attributes, e.g. illumination, pose, expression, or similar. A face recognition system uses a four-step approach to a recognition process, as can be seen in figure 2.3. These steps, or modules, are: detection, alignment, feature extraction and matching. Face detection and alignment can be considered as preprocessing requirements before recognition can take place, where recognition consists of feature extraction and matching.

Face detection is responsible for segmenting the face, or more specifically a face area, from the background. By aligning the faces, the aim is to more precisely pinpoint face location and also to normalise the faces as data for the next stages. As such, detection and alignment work in tandem to provide estimates of the location, and scale, of faces detected in the input data. Based on this it is possible to extract facial components, including the mouth, nose, eyes, and the jawline. This is done by performing morphing or geographical transforms. The next step is to further normalise with regards to photometrical properties, such as illumination and grey scale.

After having been geometrically and photometrically normalised the face object

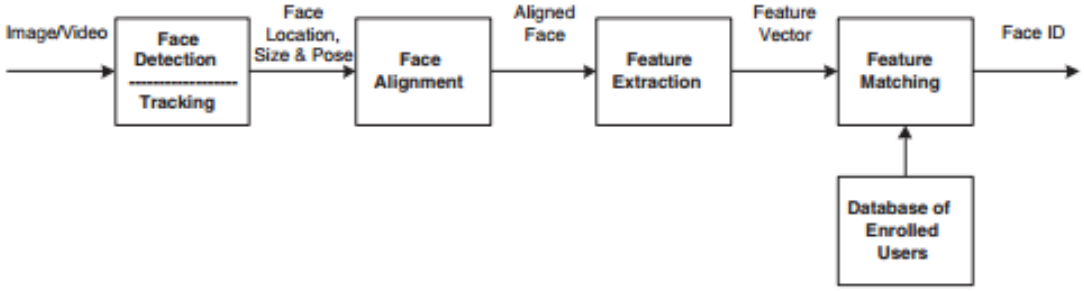


Figure 2.3: Face recognition processing flow (Li and Jain 2011, p. 3).

is ready to undergo feature extraction. In the case of face recognition, the interesting features to extract are those that are useful, and consistent in regards to the geometrical and photometrical variation, for differentiating between faces. The final module of matching is then to compare the extracted feature vector against some applicable database of similarly processed faces. It will then either output a match with a certain degree of confidence, or suggest that the input face is unaccounted for.

The success of face recognition greatly relies upon the features that are selected to embody the face, and also the classification methods used to distinguish between faces. Underlying this is the localisation and normalisation pre-processing which facilitates the extraction of useful and effective features.

2.2.2 Facial Landmark Detection

Facial landmark detection has already been mentioned as part of the process in recognising faces. The facial landmark points make up the identifying points of a face, for a computer, allowing for the use of faces as data. Zhang et al. (2014) describe the importance and challenges of facial landmark detection, an integral part in facial analysis tasks, face verification, and face recognition as mentioned. Even though there has been performed considerable amounts of work in facial landmark detection, Zhang et al. argues that a robust solution still remains to surface. Some of the challenges include partial face occlusion and considerable head pose variations. Historically, there has been two main categories for detection, this being regression-based and template fitting methods. Where regression-based methods rely solely on landmark estimation by regression using image features, template fitting methods builds face templates to fit the input images into.

Another approach is by using cascaded Convolutional Neural Networks (CNNs). The cascaded CNNs requires faces to be divided into separate parts, where each part is handled in turn by its own deep CNN. Outputs are averaged and passed on to cascaded layers where every facial landmark is estimated individually.

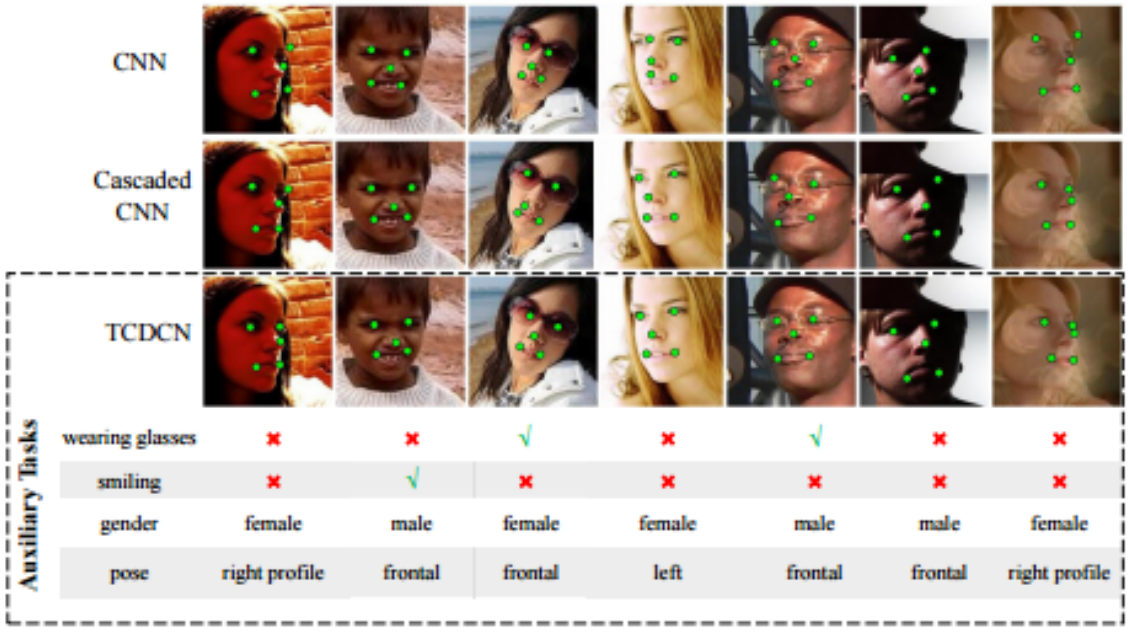


Figure 2.4: Example of landmark detection by a single conventional CNN, a cascaded CNN and a TCDCN (Zhang et al. 2014, p. 95).

Traditionally, it has been treated as an isolated and independent problem, something which Zhang et al. claims to be a shortcoming. They have instead proposed a new approach which combines the use of conventional CNNs with auxiliary tasks, "... which include head pose estimation, gender classification, age estimation, facial expression recognition, or facial attribute inference." This concept is named a Tasks-Constrained Deep Convolutional Network (TCDCN).

2.2.3 Face Averaging

The average face is a concept that has been of interest in several disciplines. It has been subject to much debate within psychology, where several studies have shown that computationally averaged faces are generally regarded as more aesthetically pleasing (Halberstadt and Rhodes 2000; Halberstadt and Rhodes 2003). This phenomena is often credited to the fact that through averaging, individual imperfections and asymmetry are watered down. Koinophilia, an evolutionary hypothesis claims that an average looking individual is more often preferred as a mate as it is less likely to be subject to undesirable mutations within a species (Koeslag 1990). The first average face dates back to 1878, when Francis Galton created a new technique for compositing faces in the development of photographies. By aligning the eyes of several face images and exposing them on the same photography plate, Galton managed to create a new face, the composite face, which combined all the original faces (Benson and Perrett 1991; Galton 1878). The composite technique had its resurgence in the 1990s when computers could take over these operations, and it is

now often referred to as face averaging (Mallick 2016).

The concepts of computationally averaging a set of faces are fairly similar to those of the composite face. All face images to be averaged must go through the same process, starting off by localisation in the form of face landmark detection (see figure 2.4). Additionally, all faces must be normalised. Considering that images come in different sizes, the first step is to create a common reference frame. In this frame, the coordinates of the eye corners are defined, or some other points of reference, and the original image is warped and the landmarks are transformed using a similarity transform (see fig. 2.5). This means that the output coordinates are aligned in such a way that all faces have their eyes at roughly the same location in the frame. However, this only really aligns the eyes, it is also required to align the rest of the facial features. This is done by calculating the mean landmark coordinates of each reference frame and then calculating a Delaunay Triangulation. This means that given the landmarks as coordinates and the face as a plane, triangulation returns a subdivision of this plane into triangles with the landmarks as triangle corners. In other words, the entire face is now represented as triangles between the points of all facial features. It is worth noting that there are many triangulations for a set of points, but the Delaunay Triangulation favours a distribution of triangles with evenly sized angles. It does this by ensuring that no point is within the circumcircle of any triangle in the subdivision, as demonstrated in figure 2.6. Given this triangulation it is possible to warp the face triangles to match the mean average face landmark points using an affine transform. Given a source plane, and a destination plane, this transformation preserves collinearity, which means that all points lying on a line in the source plane still lies on a line in the destination plane. Ratios of distances are also preserved from source to the destination plane, for instance, the midpoint of a line still remains the midpoint post-transformation (Weinstein 2017). This means that all faces will have their facial features aligned to the mean coordinates for the entire face within the landmark points, i.e. all pixels within the triangular subdivision. Ultimately the pixel intensities (e.g. the value of each colour-channel for images using the RGB colour space) of all the warped faces are added onto an output image, and then divided by the number of faces (Mallick 2016).

2.2.4 Face Morphing

Face morphing is the process of creating a fluid transition between two faces. This transition is actually a series of images, comparable to the frames of a video, of differing alpha blending. This blend determines the relationship of pixel intensity between the two images, and by parameterising this alpha value it is possible to decide which face is to be more dominant in the end result. The process of face

What is a similarity transform ? A similarity transform is a 2×3 matrix that can be used to transform the location of points (x, y) or an entire image. The first two columns of this matrix encodes rotation and scale, and the last column encodes translation (i.e. shift). Let's say you want to transform (move) the four corners of a square so that the square is scaled in the x and y direction by s_x and s_y respectively. At the same time it is rotated by an angle θ, and translated (moved) by t_x and t_y in the x and y directions. The similarity transform for this can be written as

$$S = \begin{bmatrix} s_x \cos(\theta) & \sin(\theta) & t_x \\ -\sin(\theta) & s_y \cos(\theta) & t_y \end{bmatrix}$$

Given, a point (x, y), the above similarity transform, moves it to point (x_t, y_t) using the equation given below

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} s_x \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & s_y \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Figure 2.5: Similarity transform (Mallick 2016).

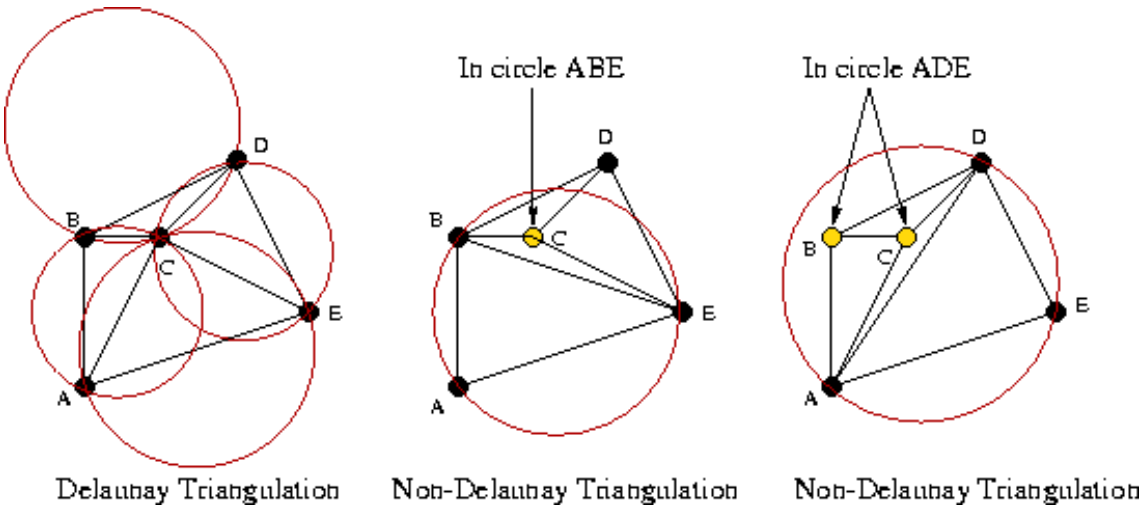


Figure 2.6: All possible triangulations for point A through E (Peterson 2017).

morphing is very similar to the process of creating average faces and uses several of the same operations. By using Delaunay Triangulation it is possible to create corresponding triangles which can be transformed and warped from one face onto the other using the concept of affine transformation as mentioned previously. Finally, the warped faces can be alpha blended using the alpha blend parameter. The result will then be a morphed face which is a combination of the two faces, where the given alpha value decides which face is more dominant (Mallick 2016).

2.2.5 Face Swapping

The concept of face swapping also uses facial landmark detection, face alignment, Delaunay Triangulation and affine warping as described in the previous sections. Given the detected landmarks, the convex hull (the smallest convex set of points that contains all other points) of one face is aligned on top of the other, and potentially vice versa. By using Delaunay Triangulation and affine transform the triangles of the faces are warped to match their destination face. However, the process is not finished here, as an essential operation remains. Seamless cloning is an implementation based on the findings from the paper 'Poisson Image Editing' by Pérez, Gangnet and Blake from 2003. The paper argues that it is beneficial to work with< image gradients as opposed to image intensities as a means to achieve more realistic results when performing cloning. Seamless cloning makes the warped face blend with the destination face by altering aspects of the face like texture, illumination, and colour. This entire process will result in the destination face now having a different facial appearance, but approximately the same photometrical qualities as before the swap (Mallick 2016).

Chapter 3

Methodology

This chapter will explain, discuss and justify the selection of methodology for the research design. First, the design science research methodology framework will be introduced, and the implications it has for the research design. All tools and services used as part of development and evaluation will also be described.

3.1 Design Science Research

The goal of this development period is to produce an information science artefact, in addition to the two auxiliary artefacts mentioned in section 1.3.2. In order to ensure that it adheres to the expectations of such an artefact, it could be helpful to use a relevant research framework. One such framework is the design science research approach, which offers guidelines for evaluation and iteration in the research project. The design science model is shown in figure 3.1. Hevner names seven guidelines which are important for research of this kind (Hevner, March and Park 2004).

3.1.1 Design as an Artefact

"Design science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation" (Table 3.1).

Artefacts need not be complete information systems, but do need to define the requirements of effective and efficient accomplishment. Design science artefacts are innovations that define the ideas, practices, technical capabilities, and products which lead to this accomplishment. This is done through the analysis, design, implementation, and use of the information systems, addressing both the process of design and the product itself (Hevner, March and Park 2004).

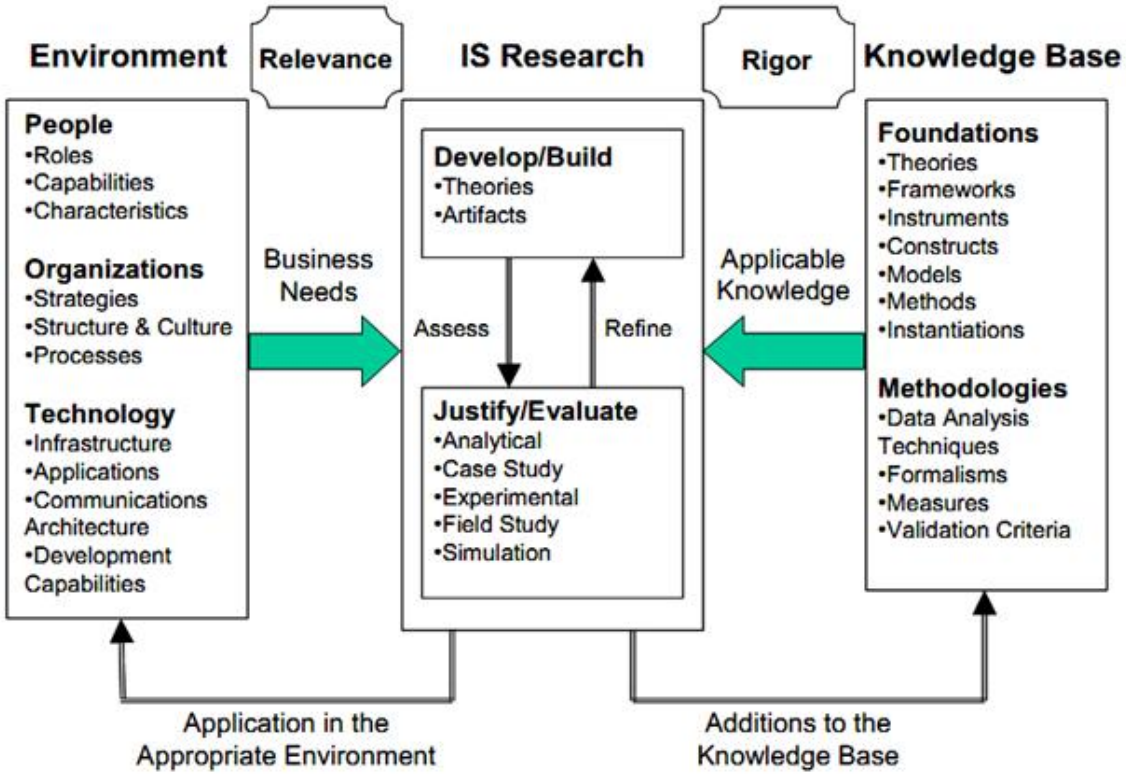


Figure 3.1: Design Science Research Model (Hevner, March and Park 2004, p. 80).

Table 1. Design-Science Research Guidelines	
Guideline	**Description**
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 3.1: Design Science Research Guidelines (Hevner, March and Park 2004, p. 83).

3.1.2 Problem Relevance

“The objective of design science research is to develop technology based solutions to important and relevant business problems” (Table 3.1).

A problem can be defined as the discrepancy between a goal state and the current state of a system. Problem solving can be defined as a search process using actions to reduce or eliminate the differences. How to define the problem relevance is subject to the community in which the problem is present. ”The relevance of any design-science research effort is with respect to a constituent community. For IS researchers, that constituent community is the practitioners who plan, manage, design, implement, operate, and evaluate information systems and those who plan, manage, design, implement, operate, and evaluate the technologies that enable their development and implementation. To be relevant to this community, research must address the problems faced and the opportunities afforded by the interaction of people, organizations, and information technology” (Hevner, March and Park 2004, p. 85). This way, the communities are left with valuable artefacts which are able to address their relevant problems. This means allowing for ways of reflection, representation, exploration, analysis, and optimisation of these problems, and allowing for ways to affect them (Hevner, March and Park 2004).

3.1.3 Design Evaluation

”The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well executed evaluation methods” (Table 3.1).

In design science research the importance of evaluation in the research process is accentuated. Well executed evaluation methods are to rigorously demonstrate the utility, quality and efficacy of the design artefact. Defining problem specific appropriate evaluation criteria are key to this process, where “...evaluation of a designed IT artifact requires the definition of appropriate metrics and possibly the gathering and analysis of appropriate data. IT artifacts can be evaluated in terms of functionality, completeness, consistency, accuracy, performance, reliability, usability, fit with the organization, and other relevant quality attributes” (Hevner, March and Park 2004, p. 85).

Design is described as an iterative process in which the evaluation phase can provide useful feedback to the construction of an artefact. Typically established evaluation methods from the relevant knowledge base are favoured, as seen in figure 3.2 (Hevner, March and Park 2004).

Table 2. Design Evaluation Methods	
1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation – Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Table 3.2: Design Evaluation Methods (Hevner, March and Park 2004, p. 86).

3.1.4 Research Contributions

"Effective design science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies" (Table 3.1).

Hevner et al. (2004) lists three types of research contributions where a minimum of one must be present in a design science research work. One is the artefact itself, in terms of solving one specific previously unsolved problem. Another is the foundation, meaning "the creative development of novel, appropriately evaluated constructs, models, methods, or instantiations that extend and improve the existing foundations in the design science knowledge base" (Hevner, March and Park 2004, p. 87). And finally methodologies in evaluation are mentioned, where "...the creative development and use of evaluation methods (e.g., experimental, analytical, observational, testing, and descriptive) and new evaluation metrics provide design science research contributions" (Hevner, March and Park 2004, p. 87).

3.1.5 Research Rigour

"Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact" (Table 3.1).

"In both design science and behavioral science research, rigor is derived from the effective use of the knowledge base, theoretical foundations and research methodologies. Success is predicated on the researcher's skilled selection of appropriate techniques to develop or construct a theory or artifact and the selection of appropriate means to justify the theory or evaluate the artifact" (Hevner, March and Park 2004, p. 88).

The question of research rigour is a trade-off between rigidity and flexibility in construction and evaluation applications. The aim is to create a design which retains the greatest amount of relevancy, where successfulness of the artefact is prioritized, and not so much to theorise about the reasons for the success (Hevner, March and Park 2004).

3.1.6 Design as a Search Process

"The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment" (3.1).

A design science research project should aspire to find some optimal design which addresses a problem in an efficient way. The design process itself is a search of this optimal design. The Generate/Test Cycle, see figure 3.2, is used as a way of describing the nature of the search process, which is defined in the above quote. Means are described as the actions and resources available to construct the solution, the laws as intractable influences from the environment, and ends as the goals and constraints. One way of implementing such a cycle is by defining subsets of means, laws and ends, initially, and by iterative progress adding value and relevance to the artefact (Hevner, March and Park 2004).

3.1.7 Communication of Research

"Design science research must be presented effectively both to technology-oriented as well as management-oriented audiences" (Table 3.1).

It is important that the communication of a design-science project consider the difference in technological competence and knowledge of the artefact and its constituents, as well as the motives for interest in the artefact. From an academic point

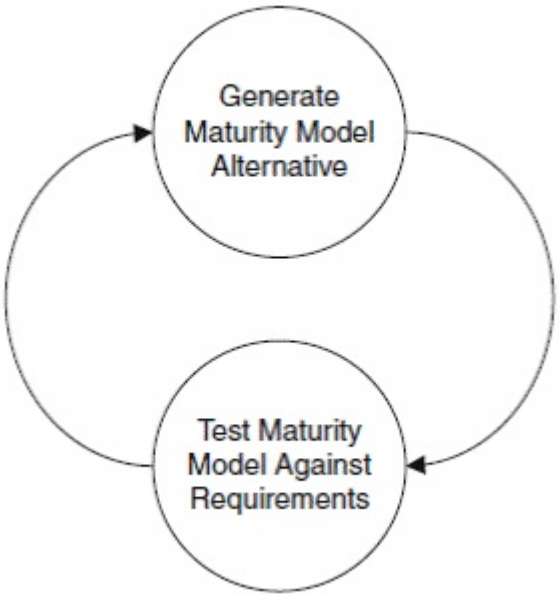


Figure 3.2: The Generate/Test Cycle (Hevner, March and Park 2004, p. 89).

of view this includes how the artefact is constructed and evaluated, abiding the principle of repeatability.

When concerning a management-oriented party the resources required to implement the artefact in a specific context should be emphasized in place of nature of the artefact itself. It is important to tailor the communication in a manner which is conscious of the recipient of the information, focusing on the aspects of the artefact which is important to them (Hevner, March and Park 2004, p. 87).

3.2 Design Science Research Implementation

In this section I will discuss how the design science research framework will be implemented as a foundation for this research. Each mentioned guideline will be discussed in relations to the development of the artefact, its evaluation, and its desired outcome as a work of research.

Guideline 1: Design as an Artefact As mentioned in section 1.3.2 there are several artefacts involved in this project. The main artefact is the facial anonymisation algorithm, while the two other artefacts are ways of implementing it for evaluation and system integration, i.e. ways of providing usability for the technology. By creating a proof of concept which demonstrates the feasibility of facial anonymisation, the main artefact will allow for research and debate, and opens for innovative applications of the technology for the community.

Guideline 2: Problem Relevance The artefact has a defined problem which it aims to solve as a research artefact for the ViSmedia research group. This is perhaps not a traditional business problem, in design science research context, but it is still a suitable problem for using a design science approach. It is assumed to be a good fit for the RRI-framework as a potentially controversial artefact in the problem domain. It also carries potential for business employment, possibly both within the domain it is to investigate, but also other domains in which it could be of use. This includes both the artefact as a system, and as a technological construct.

Guideline 3: Design Evaluation The artefact will be evaluated for utility, efficacy and quality, answering the three research questions stated in section 1.3.3.

The evaluation of designed artefacts is to utilise the methodologies which are available in the knowledge base, depending on the nature and field in which the artefact is native to. Considering the explorative nature of this project, there are few established evaluation methodologies that fully meet the criterion of rigour. However, informed arguments can be construed based on the information from the knowledge base, where "... descriptive methods of evaluation should only be used for especially innovative artifacts for which other forms of evaluation may not be feasible" (Hevner, March and Park 2004, p. 86). Additionally, as mentioned in section 3.1.5, there is a trade-off between rigidity and flexibility when designing evaluations for an artefact.

The selection of evaluation methods, and their specific criteria of success, for this project will be discussed in chapter 5 after the technical aspects of the artefacts have been presented.

Guideline 4: Research Contributions There will be several contributions as a direct result of this thesis. For one, there will be the applications created in connection to the Prosopo project. The Prosopo API will serve as a foundation for new functionality, applications and allows for further research.

Additionally, the main artefact will remain as an implemented application and as a theoretical construct. This artefact could, after further iterations of development and evaluation, progress from an investigative tool into a fully deployable business application or a theoretical foundation for other facial manipulation applications.

Guideline 5: Research Rigour This project will employ an experimental and flexible approach with regards to both development and evaluation. This is a natural consequence of the project having limited existing relatable work, a mixed background from different academic fields, and due to using technology in an untraditional way.

In terms of evaluation the project will employ a mixed method evaluation that will reduce the risk of design limitations resulting from single evaluation approaches. This is based on rigid theories from psychology and is an established approach in empirical research (Caracelli and Greene 1997). The rigidity-flexibility trade-off, which has been mentioned several times already, will be further described in relation to the evaluation design in the evaluation chapter.

Guideline 6: Design as a Search Process For this project, as it is of a somewhat exploratory nature, iterative development will be a useful approach to reaching a satisfactory design. The foundational technology must be explored and partial solutions must be implemented, and compared, while trying to determine which approaches can lead to an increasing level of success. The search process will heavily rely on expert evaluation based on visual feedback from each iteration.

Guideline 7: Communication of Research The communication of this research project will aim to take into consideration aspects which are vital to both technology-oriented and management-oriented audiences, as mentioned as part of the design science research guidelines. Dedicated chapters will deal with both technical detailing, as well as the practical possibilities of implementation. The technology will be put into a theoretical context, and the technological possibilities will be assessed in relation to other potential use cases. Furthermore, the construction, background and evaluation of the artefact will be presented in an academic manner.

3.3 Tools

In this section the tools involved in the development will be discussed. The working environment set-up will be explained, which includes the programming languages and libraries, online sources, and any application or service which have been significant in progressing development.

3.3.1 Programming languages

Programming languages are a way for programmers to write instructions which can be understood by a computer. Each language provide their own set of syntactic rules which allows for translatable code. There are different groups of languages which share an approach on how to write useful, efficient and maintainable code. Certain languages are popular within certain domains, and some are more efficient at solving a given set of problems. The development of the Prosopo applications

and the main artefact have resulted in the utilisation of three different programming languages.

Python

When deciding upon the main environment for development the choice of programming language quickly fell upon Python. Even though I lacked extensive proficiency in Python I expected that its lightweight nature would greatly increase flexibility and reduce compile time and the need for specific structure in the early phases. It also supports multiple systems and platforms and has quick access to plenty of useful libraries, one such is dlib, which is mentioned later. Most importantly, however, it is one of a few select languages that have OpenCV interfaces, the computer vision library which functions as the foundation for this entire project.

C#

It was decided to use the .NET-framework for web-development, which uses the C# programming language, as it was the only one I had experience with. As the university server runs Red Hat Enterprise Linux 7, it was required to use the very latest .NET Core version, the first version in the series that provides support for Linux-systems.

Java

When creating the front-end application to serve as a platform for demonstration and data collection, it was decided to develop a simple Android-application. I had quite a bit of experience using Java, and developing Android applications, and also had readily available relevant code from previous projects to use for quick implementation.06

3.3.2 Libraries

Programming libraries are usually a collection of related functionality built for a given programming language. These libraries can be imported into projects helping programmers find useful and verified solutions to their problems. Open source libraries are provided with full insight and allows for modification and inspiration when dealing with specific problem domains. Below are two central libraries which were used in the development of the artefact.

OpenCV

This project would not have been possible without the work put into the open source library OpenCV. The computer vision library was originally developed by Intel starting out in 1999 as a way of advancing CPU-intensive applications (Bradski and Kaehler 2008). It continued to evolve through different owners, supporters and developers, but ended up back at Intel in 2016 (D. Davis 2017), and now supports a multitude of functionality within computer vision and machine learning. OpenCV is used worldwide and has many fields of application, ranging from robotics to interactive art. It is written in C++, but provides interfaces for Java and Python, and there also exists wrappers to other languages (Bradski 2000).

dlib

Early in development the open source machine learning library dlib was used for facial landmark detection. It is written in C++, but also has support for Python, and is very well documented and provides guides for usage and building. Much like OpenCV it is a renowned tool for solving a range of complex problems, within many of the same domains (King 2009).

3.3.3 Cognitive Services

A cognitive service is not really a protected title in the academics of information systems, and is not to be confused with a medical service. A cognitive service provider, in this context, delivers computational cognitive functionality. This can be described as a computer being capable of dealing with perception and reasoning in the same way as a human would, to some extent. Within the computer vision aspect, typical services are face detection, face comparing, face searching, facial landmark detection and facial attribute analysis. Cognitive service providers may also offer functionality to analyse images (or videos) for objects which are not human faces, as well as speech, language, knowledge and search tasks.

Facial landmark detection has previously been mentioned, which is a way of estimating the location of facial features using machine learning. The concept of facial attribute analysis is fairly similar. A set of attributes are classified or given a value of probability from trained algorithms, a set which may differ from provider to provider (see example in table 4.1). Face search uses detected faces in an input image versus faces in an existing database, where another trained algorithm can return images which have a high probability of containing the same face. Face comparing, takes two faces and returns a probability value for belonging to the same person.

Input images are sent to cognitive services by posting the image as an HTTP-request. Essentially, this means sending the image on a specific format to a specific address, where the functionality linked to this address returns formatted information to the sender. This way, the sender always knows what type of information to expect in return, as long as the data sent to the service is on the correct format.

Face++

"Face++ Cognitive Services is a platform offering computer vision technologies that enable your applications to read and understand the world better. Face++ allows you to easily add leading, deep learning-based image analysis recognition technologies into your applications, with simple and powerful APIs and SDKs" (Face++ Cognitive Services 2017).

Face++ is one of two chosen cognitive services for this artefact. Face++ has existed since their first launch in 2012, and has since relaunched twice and are now in their third iteration as a cognitive service. They provide both web-based and in-application/in-device computer vision functionality, which includes face detection, face comparison, face search, facial landmark detection and facial attribute analysis (Face++ Cognitive Services 2017).

Microsoft Cognitive Services

"Detect human faces and compare similar ones, organize people into groups according to visual similarity, and identify previously tagged people in images" (Microsoft Corporation 2016).

Another provider is Microsoft Cognitive Services. Where Face++ is focused on facial computer vision, Microsoft Cognitive Services has a wider supply of functionality, divided into five categories. These categories are named vision, speech, language, knowledge, and search. Under the vision category the Microsoft's face API is found. Much like Face++ this functionality includes face detection (here within facial attribute analysis), face verification (face comparison) and face searching. They also provide face identification, which can identify faces which have been tagged by user provided data, and facial grouping which groups unidentified faces by similarity. Additionally, they provide functionality for analysing frames in videos (Microsoft Corporation 2016).

3.3.4 Online sources

As a developer you are often met with problems, of varied complexity, which can reduce the speed of progress. Chances are you are not the first one with this specific problem. Online sources, communities and tutorial sites are numerous, and can provide inexperienced and experiences developers, alike, with insight into specific problem domains. Why reinvent the wheel when you can use a well-proven and validated solution? Often the answer is, you should not reinvent the wheel. In this case, these sites are extremely helpful in finding the solution to these hiccups which surfaces along the road. Here are the ones which were essential for the development in this project.

Stack Overflow

"Stack Overflow is the largest online community for programmers to learn, share their knowledge, and advance their careers." (Stack Exchange Inc. 2017)

Stack Overflow is perhaps the programmers guide to the galaxy. Since its founding in 2008, it has become a massive community of programmers, with 40 million monthly visitors. The site is a massive Q&A (questions and answers) knowledge base of programming related questions. Programmers may ask specific questions based on specific problems, where any other programmer may opt to try and answer these questions. Other programmers can then evaluate both the question and the answers, which provide useful feedback for other programmers whom might find this of interest in the future (Stack Exchange Inc. 2017).

In the development of the artefacts, the Stack Overflow site has been invaluable to progress. It has helped solving both smaller and larger problems on a consistent basis. Particularly it has been a source of learning dealing with my initial inexperience in Python and .NET Core-development. It has also been useful in more specific tasks, e.g. as a source of finding best-practice solutions for given problems.

Learn OpenCV

Learn OpenCV is an example and tutorial based blog which provides introduction to possible use cases for the previously mentioned computer vision library OpenCV. Its creator Satya Mallick, who holds a Ph.D. in computer vision and machine learning in addition to years of experience in the field, describes it as follows: "This blog is for programmers, hackers, engineers, scientists, students and self-starters who are interested in Computer Vision and Machine Learning." The blog produces articles, or blog posts, on a regular basis covering new aspects and possibilities within OpenCV (Mallick 2017).

The blog has been a source of inspiration in implementing a lot of the computational solutions for the artefact. Mallick has covered several of the topics on facial manipulation, like face morphing and face averaging which are central to the process of facial anonymisation. The code provided in these blog posts have been used as the foundation for my own requirements. Reading the tutorial and studying the code examples has provided a very practical approach to these technologies, and a welcome complementation to the theoretical background.

PyImageSearch

Much like Learn OpenCV, PyImageSearch is a blog with the aim of teaching computer vision and machine learning by example, with a focus on image search engines. Adrian Rosebrock, the creator, describes his blog in the following manner: "This blog is dedicated to helping other programmers understand how image search engines work. While a lot of computer vision concepts are theoretical in nature, I'm a big fan of "learning by example"." Rosebrock shares a similar background as Learn OpenCV's Mallick, as they both hold a Ph.D. in computer vision and machine learning, and both are also entrepreneurs within the field (A. 2017).

PyImageSearch was the first introduction that I started using after deciding on the design of the artefact. I had previously applied some of Rosebrock's content in a different project, and it had been very helpful and provided a lot of insight into the OpenCV library. A great deal of my proficiency in OpenCV was based on of the work of the PyImageSearch blog, and I also found it to be quite useful during development. It served as a general knowledge database which worked great in tandem with the OpenCV-documentation.

3.3.5 Other

This section will include several programs and services which were utilised in some meaningful way during development.

Integrated Development Environments

The choice of which integrated development environment (IDE) to develop in is very much up to preference. However, there are certain IDEs created for specific purposes by the developers behind the system one is to develop for. One such is Android Studio by Google, a dedicated IDE for developing applications for Android-systems (Google 2017). Another is the Visual Studio series by Microsoft, which offers an environment for developing Microsoft applications like the .NET-framework (Microsoft Corporation 2017). Visual Studio also offers the possibility to develop systems for other platforms and systems. Both of these were obvious choices for

developing the web-API and the Android-application, and I was already proficient and experienced in both. Additionally, PyCharm Community version was chosen for developing the Python-scripts running computation and analysis, i.e. the main artefact, on behalf of the API. PyCharm, by JetBrains, was chosen due to positive reviews, and also due to the fact that it verifies files for PEP8 compliance, a style-guide for Python code (JetBrains 2017). This was considered beneficial due to my lack of experience with Python.

Version Control

As the Prosopo project was a collaborative effort, version control was absolutely essential. Using a Version Control system was necessary and the choice fell on Git due to previous experience. We hosted separate repositories for each module, and a repository was also used for the initial development project. To do this we used Bitbucket, a web-based hosting service for Git, free to use for academic purposes. We also used Sourcetree, which is a graphical interface for handling Git-commands.

3.4 Dataset

In order to be able to create average faces for a given input face it was necessary to have a database of pre-analysed faces. This database had to contain enough images to be able to fully represent all faces within the problem scope. Several databases were readily available spanning from early face recognition databases in the 90s, created for different purposes within face recognition (Grgic and Kresimir 2017). One such database, that stood out positively, was the **10k US Adult Faces Database**, which is described as follows:

"This database contains 10,168 natural face photographs and several measures for 2,222 of the faces, including memorability scores, computer vision and psychology attributes, and landmark point annotations. The face photographs are JPEGs with 72 pixels in resolution and 256-pixel height." (Bainbridge 2012)

The database was built in 2012 to serve as data when running several face memorability experiments. The faces were automatically downloaded from Google Image searches and observers removed images which did not meet their requirements. This excluded recognizable celebrities, low-quality images, children and occluded faces. The resulting database then represented the average US population distribution in regards to gender, age and race distribution (Bainbridge, Isola and Oliva 2013).

The database was deemed ideal as it was already quality-assured, for similar

Figure 3.3: Three faces from the 10k US Adult Faces Database (Bainbridge, Isola and Oliva 2013).

purposes to those of this project as it had already been used for facial landmark extraction and facial manipulation. I contacted Wilma Bainbridge and applied for a licence to the database with scientific intentions. The reply was positive and I was made aware of restrictions considering usage of face images in publication works, as only a specific subset of the images were available for publication. Example faces from this subset are seen in figure 3.3.

Chapter 4

Technical Presentation

This chapter aims to explain how the artefacts of this thesis are built up from a technical perspective. The main focus will lie on the anonymisation artefact, while the two auxiliary artefacts will be described in a more pragmatic manner.

4.1 Face Anonymisation

The anonymisation process consists of two sub-processes. The input image is first analysed, where all faces detected are represented as objects with a set of facial characteristics. Each face will then go through a series of manipulations based on their characteristics, leading to an output image where the all faces are anonymised. Each sub-process has its own series of steps, which will be described further. Selected code excerpts are listed in appendix C.

4.1.1 Analysis

The first step of analysis is detecting faces in the input image. Using the cognitive services mentioned in section 3.3.3, all detectable faces are identified by the sum of the left eye's left corner's landmark coordinates. This way, the ordering of faces will be the same for faces found by either cognitive service, as their combined axis location will be approximately similar for both services. Any face which is found by one service, but not the other, is ineligible for anonymisation. Each service also returns a series of facial attributes, some of which are overlapping and others unique to the respective service, as illustrated by table 4.1. The 'X' means that the cognitive service provides a value for this attribute, while blank space means that they do not. Attributes with a parenthesised keyword means that there is a difference between the providers given by this keyword. The final column shows which service was chosen for this project.

By combining the results from both cognitive services each face now has an object

representation. All attributes from age through yaw are used to describe what type of face this is, while the landmarks and face rectangle are location attributes which are used to manipulate the face in the next process. Face quality, blurriness, motion blur and Gaussian blur are all values which indicate how certain photometrical conditions have affected the analysis, and are not utilised in any way at this point.

A fairly primitive skin colour detection is then performed. Using the landmark points, a mask is created covering the parts of the face which are typically showing skin, meaning that the mouth, eyes and eyebrows are removed. The mean pixel intensity is calculated from the face underneath the mask and is represented as an RGB-value tuple. This concludes the process of creating a face object, ending the analysis process. Table 4.2 shows the facial attributes which were found for the faces in figure 3.3.

4.1.2 Manipulation

After the analysis we are left with a face object for each detected face, ready for anonymisation. Each face is anonymised separately. Prior to any manipulation each face is first cropped from the image using the detected face rectangle (which is in fact always a square). This means that all images, which are represented as a two-dimensional matrix where each cell contains a pixel/RGB-tuple, will be of equal size. An input parameter, $\boldsymbol{\alpha}$, is also provided. The alpha blending parameter could

Attribute	Face++	Microsoft	Chosen
Age	X	X	Microsoft
Gender	X	X	Microsoft
Smiling	X	X	F++
Glasses		X	Not chosen
Right eye	X		F++
Left eye	X		F++
Moustache		X	Microsoft
Beard		X	Microsoft
Sideburns		X	Microsoft
Pitch	X		F++
Roll	X	X	F++
Yaw	X	X	F++
Landmarks (count)	83	27	F++
Face rectangle	X	X	F++
Face quality	X		F++
Blurriness	X		F++
Motion blur	X		F++
Gaussian blur	X		F++

Table 4.1: Comparison between attributes from F++ and Microsoft.

Attribute	Face 1	Face 2	Face 3
Age	27.6	34.5	34.0
Gender	male	female	male
Smiling	17.47	98.13	98.55
Right eye	no_glass_open	no_glass_open	no_glass_open
Left eye	no_glass_open	no_glass_open	no_glass_open
Moustache	0.0	0.0	0.5
Beard	0.0	0.0	0.4
Sideburns	0.0	0.0	0.4
Pitch	-0.37	2.06	3.59
Roll	4.25	1.33	-0.74
Yaw	-4.32	-2.97	1.27

Table 4.2: Facial attributes from faces in figure 3.3.

be described as the degree of anonymisation, i.e. how much, or how little, of the original face is to remain in the end result.

The first step of manipulation is to create an average representation from the faces most similar to the input face, where similarity is calculated based on the facial attributes which were determined in analysis. This average face is then morphed with the original face with given the $\boldsymbol{\alpha}$ parameter. Finally, the cropped face is swapped with the morphed face and then inserted back into the original image. This entire system is illustrated in figure 4.1.

Step 1: Average Face A similarity calculation finds the five most similar faces from the database which are then used as input for the face averaging. This similarity algorithm uses a combination of exclusion and weighting of attributes to calculate a level of similarity between 0 and 1, where 0 is an identical face. This calculation is not trained or dynamic in any way, but manually weighted through trial and evaluation. See section 2.2.3 on face averaging to see how the manipulation is done computationally. The output should then be an average face with similar attributes to the input face.

Step 2: Morphed Face The next step is morphing the source face with the average face. This means that the average face itself has to be analysed, but this time we are only interested in the landmark coordinates. The morphing stage is in a sense the true anonymisation stage, as it is here the retained percentage of the original face and the average face is established. The theory behind morphing is described in section 2.2.4. The output is a face with a mixture of $\boldsymbol{\alpha}$ percent average face, and $\mathbf{1} - \boldsymbol{\alpha}$ percent original face.

Step 3: Swapped Face Finally we need to replace the original face with the morphed face. Similarly to the average face, we need the landmarks of the morphed face to be able to do any sorts of manipulation. The morphed face is simply swapped, see section 2.2.5, into the cropped input face and then placed back into the original image matrix from where was initially cropped. The output is now the original image where all detected faces have been anonymised to a degree of $\boldsymbol{\alpha}$ percent.

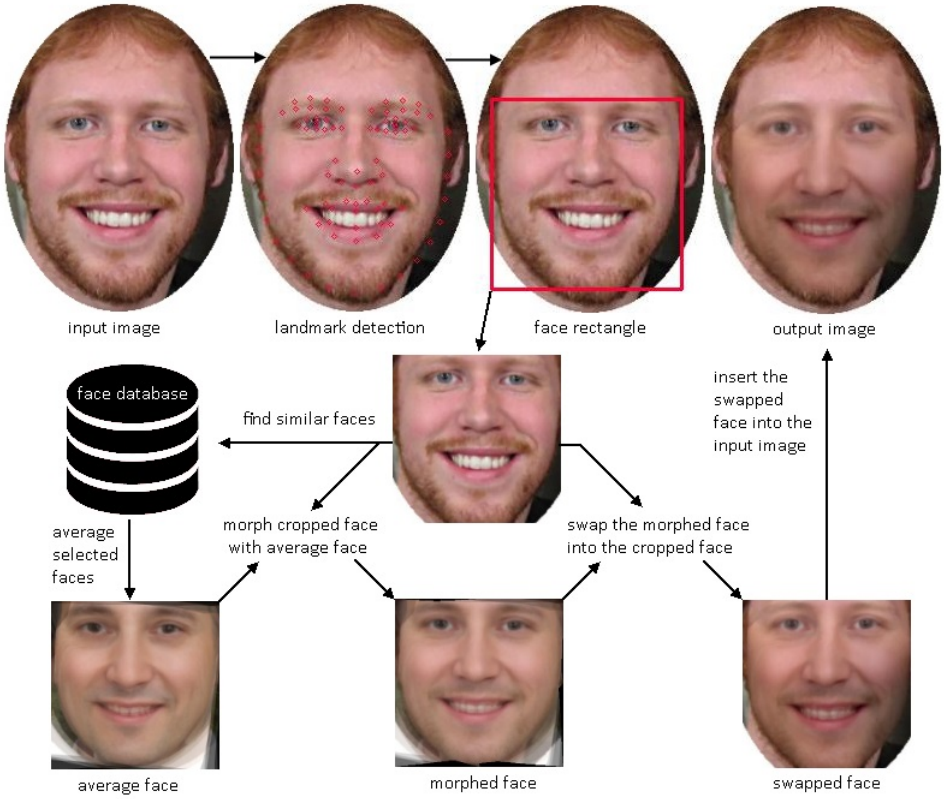


Figure 4.1: An illustration of how the anonymisation system works.

Figure 4.2: A comparison of prototype outputs. From the left: input image, featural anonymisation, holistic anonymisation.

4.2 Prototypes

As a result of the search process, three separate prototypes of the anonymisation algorithm materialised. During the initial development, a basic application was created to evaluate and quickly visualise the output of both analysis and manipulation. After exploring the possibilities and experimenting with different techniques and solutions, two separate prototypes were designed. Based on the findings on human perception, it is possible that simply manipulating facial features would not yield a sufficient anonymisation result. If face perception is a holistic process, then attempting a holistic approach to anonymisation would provide an interesting comparison. Additionally, it is worth reiterating the criteria of anonymisation as mentioned in 1.1.2.

- De-identification of the face
- Preserving face realism and image quality

The first prototype relied on feature based manipulation and aimed to have a minimal impact on the overall image quality. The second prototype had an holistic approach which focused on a larger scale of manipulation, with correspondingly higher impact on the original image quality. The differences are illustrated in figure 4.2. The prototypes will be discussed in greater detail in the following sections.

4.2.1 Development Prototype

The first development prototype was essential as a visualisation tool for providing output. Its main function was to produce evaluable data of the sub-processes and steps mentioned in section 4.1. The prototype contained tests for each of the manipulation steps, allowing for evaluation, debugging and experimentation with minimum effort. It also allowed for visualisation of additional information, such as Delaunay Triangulation and facial landmark coordinates.

Figure 4.3: An early version of the development prototype.

The application was implemented as a simple graphical user interface, created with a python-library called TkInter (shown in figure 4.3). By providing an input image, each step could be modified according to desired parameters. Whenever a new problem would arise, this application could easily be modified to provide useful debugging options and visualise the given problem accordingly. The prototype also allowed for demonstration of the artefact state to the members of the Prosopo group.

4.2.2 First Prototype: Featural Anonymisation

The featural anonymisation prototype was the first prototype which was created after reaching an acceptable level of completeness. It is designed to provide anonymisation of the area within the facial landmarks, which contains the facial features, hence the name.

Concept The concept for the first prototype is to limit the manipulated area to within the face rectangle which contains the landmark points of the facial features. These are the jawline, eyebrows, eyes, nose and mouth. Since it is possible to segment these features as a convex hull, it is possible to allow for keeping the remainder of the face rectangle fairly untouched. See figure 4.4 for example outputs.

Strengths The perceived strengths of this prototype is its ability to reduce manipulation of the input image to a very small area. This increases the likelihood of preserving face realism and also keeping the original image as unaffected as possible.

Weaknesses It is fair to suspect that the output might not reach a sufficient de-identification level. When ignoring large areas of the face (e.g. hair, facial hair, forehead and neck), it is not unreasonable to assume that there could be a lot of visual data which can provide foundation for identification. Considering that there is good evidence for face recognition as a holistic process, as mentioned in section 2.1.1, this will be particularly interesting.

Figure 4.4: Example anonymisations from the featural anonymisation prototype. Can you see who they are? See appendix D for their identities.

4.2.3 Second Prototype: Holistic Anonymisation

Though the development of the holistic anonymisation was conceptualised early on, the actual prototype is based on correcting the issues of its predecessor. It aims to increase the scope of manipulation, in accordance with the theory of holistic facial perception.

Concept The second prototype will not limit its manipulation to the original face rectangle. It will instead expand this rectangle allowing it to contain the entire face. This way, there will be fewer unmanipulated facial areas. As such, this can be called a holistic approach, accounting for the possible identifying information which can be located in all parts of the face. The prototype also crops the head, limiting the effect of contextual data from non-facial areas, such as background and clothing (see figure 4.2). See figure 4.5 for example outputs.

Strengths The strength of the holistic approach is the limited scope, and the increase of manipulated areas. According to human face perception, the face could be recognisable even when only parts of a face is visible (Matlin 2014). In this case, it should provide a higher likelihood of success in de-identification.

Weaknesses There is no technology implemented which accurately allows for detection of the entire head, allowing for segmenting the head from the background. This means that the rectangle will include a series of boundary points, points which are not facial landmarks, possibly outside the actual head. When doing manipulation with these boundary points, areas surrounding the face will also be affected. This will often result in obvious signs of manipulation, and will, as such, increase the risk of negative impact on face realism and image quality.

Figure 4.5: Example anonymisations from the holistic anonymisation prototype. Can you see who they are? See appendix D for their identities.

4.3 Implementation

The implementations of the main artefact comes in two separate auxiliary artefacts. These are the Prosopo API and the Prosopo Android which are mentioned in section 1.3.2. Where the Prosopo is an artefact for accessing the main artefact, Prosopo Android is a method for demonstrating or visualising it.

4.3.1 Prosopo API

The Prosopo API is a web-application which is responsible for all facial analysis and manipulation for the Prosopo projects. The anonymisation artefact is the first implementation using this API. The web application uses the MVC-pattern, which is a very popular architecture for web applications (Leff and Rayfield 2001). This means that is uses a model-view-controller architecture for its user interface. Each component is responsible for its own independent part of the system. The model contains the program logic, and the representation and storage of data, of the system, e.g. the representation of a face, or the rules for facial manipulation. The view is what visualises output data to the user of the system, e.g. the anonymised output image sent to Prosopo Android (section 4.3.2). Finally, the controller operates as the connection point from user input and the other components, for instance by sending the input image to the anonymisation process, or by retrieving some stored data for a view on a user's request. The API is also RESTful, which means that it uses Representational State Transfer. This is a popular architectural style of communicating to and from a web-service. This means that it should be easy for new Prosopo developers to communicate with the Prosopo API. The creator of REST, Roy Thomas Fielding, summarises REST as follows:

"This chapter introduced the Representational State Transfer (REST) architectural style for distributed hypermedia systems. REST provides a set of architectural constraints that, when applied as a whole, emphasizes scalability of component interactions, generality of interfaces, independent deployment of components, and intermediary components to reduce interaction latency, enforce security, and encapsulate legacy systems. I described the software engineering principles guiding REST and the interaction constraints chosen to retain those principles, while contrasting them to the constraints of other architectural styles" (Fielding 2000).

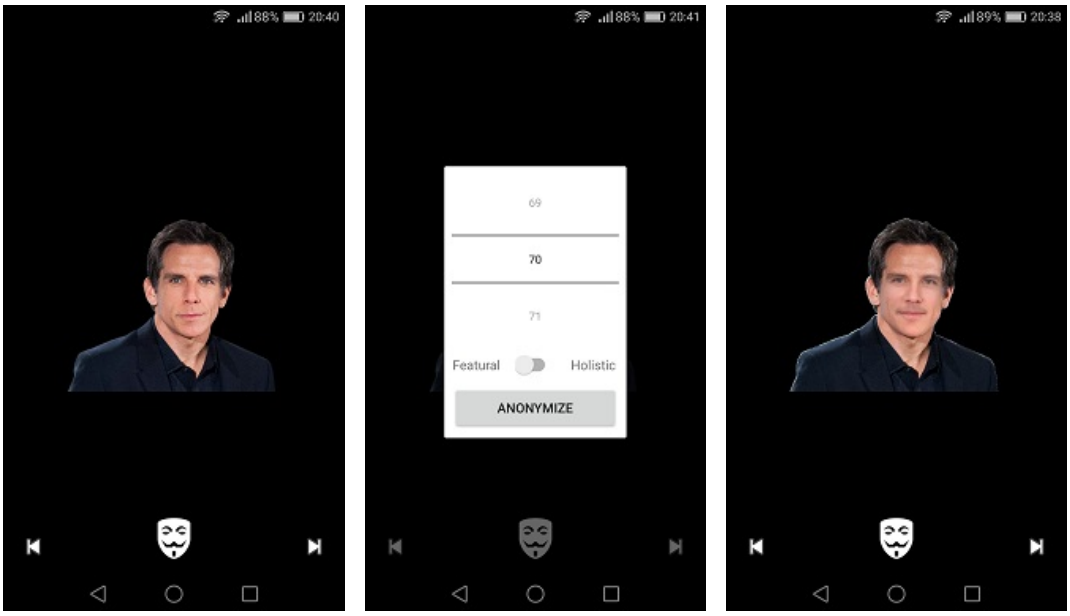
All facial analysis and manipulation is done in a virtual Python environment, and not by the API itself. The controller creates processes which executes the relevant python scripts located within the model. Both the model and the Python environment has connection to a MySQL-database which stores information about faces,

images and face sets. The actual images are stored on the server which hosts the web-application, and are represented as own objects with metadata and the path where it is stored. Face sets contains references to certain faces, used for instance if a developer want to create a very specific average face using only faces which can fulfil one or more given requirements.

4.3.2 Prosopo Android

The Prosopo Android application is, as mentioned, created for demonstrating the anonymisation. It is in fact the first application that interacts with the Prosopo service. It works by allowing its users to either snap photos of people in the wild or using a web-search for face images on the internet. After producing or selecting an input image (figure 4.6a), a degree and method of anonymisation is chosen (figure 4.6b), which results in the corresponding output image (figure 4.6c). The application also allows for navigation between the input and output images (the arrows in the bottom left and right corners).

The application works for all Android devices using Android 4.2, also known as Jelly Bean (API level 17), and up. It obviously requires internet connection, as none of the facial manipulation is done locally. This makes Prosopo Android very lightweight. It is also structured with ease of further development in mind, using standard Android coding conventions and with extensive code documentation.



(a) Input image (b) Degree and method (c) Output image

Figure 4.6: Captures from the Android demonstration application.

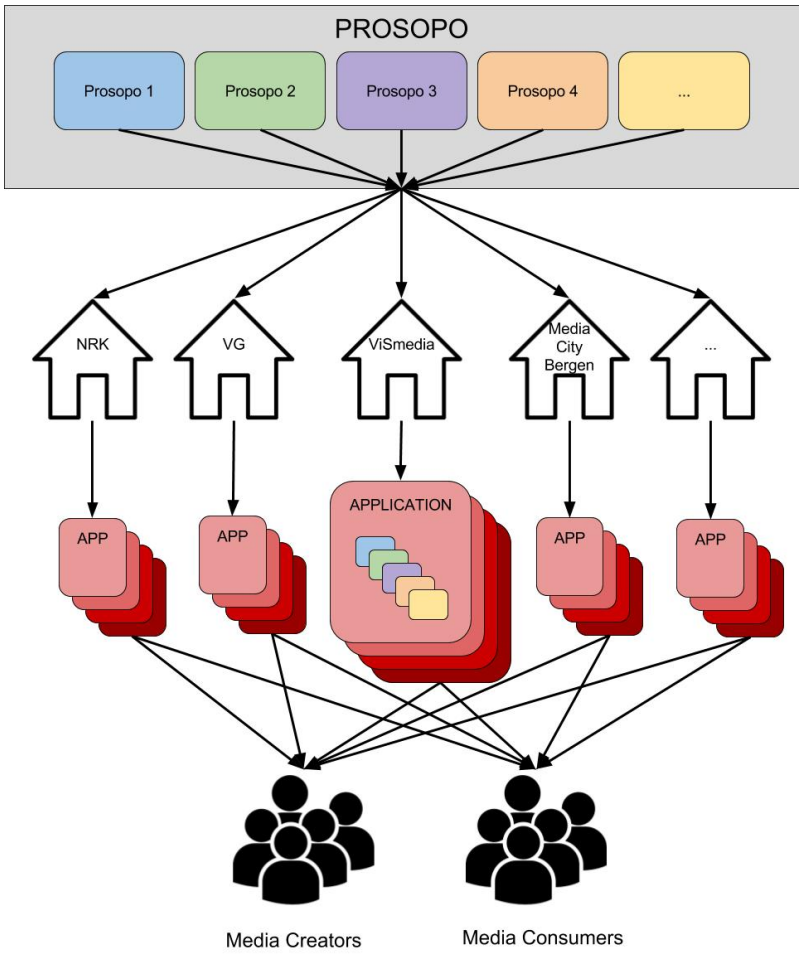


Figure 4.7: An example of the Prosopo architecture. For clarity, the only current adopter is the ViSmedia research group.

4.4 Prosopo Architecture

The Prosopo architecture, which is illustrated in figure 4.7, contains four layers. The top layer contains the different functionality of Prosopo (the Prosopo API), provided as specific services which can be accessed through API-requests. The facial anonymisation artefact is an example of this. The next layer contains the adopters of Prosopo, the businesses and organisations which have access to the upper level, such as the ViSmedia research group. The application level is where the adopters can create applications which can utilise the Prosopo functionality for their own purposes. The Prosopo Android demonstration application is one such application, provided by ViSmedia for research purposes. The bottom layer contains users of the applications, both in terms of content producers and consumers. The evaluators of this thesis could be described as an example of consumers from this group.

Chapter 5

Evaluation

As a design science research artefact, the research questions will be answered by evaluating usage in the problem domain. This section will reiterate these research questions, and then look at how to answer these questions optimally. The selection of informants and design evaluation will be discussed, separated by each iteration of the search process.

5.1 Research Questions

Q1: Is it possible to create an artefact capable of automatically anonymising faces using an average face morphing approach?
Q2: Can the artefact be used to investigate the views on normative practices in manipulation and usage of face photographies in the news media?
Q3: Are there other possible use cases for the artefact or its technology?

5.2 Evaluation Overview

According to Hevner, March and Park (2004) the utility, quality, and efficacy of an artefact should be rigorously demonstrated using established evaluation methods available in the relevant knowledge base.

There was two stages of development, where each stage differed greatly in evaluation. The initial explorative development phase was evaluated in a more flexible and diligent manner. As a researcher and developer, I used my own feedback, and that of my work group, as a way of quickly getting to a first prototype that was purposeful to evaluate in a more controlled setting. This type of evaluation was based on the theory of face perception, where development choices reflect the perceived challenges of how humans recognise faces.

After having reached this stage, the next goal was to enlist informants which could qualify as members of the problem domain. Through demonstration and use of the artefact, these informants were then asked to evaluate both the success of anonymisation (Q1), and asked to use their background to reflect upon the technology according to Q2 and Q3. In this way it would be possible to demonstrate the mentioned qualities, and accumulate the data which was required to answer the research questions of the thesis.

5.3 Development Prototype

The evaluation of the development prototype was on-going from the initiation of the project. Where the initial work was based on the theory of face perception, the succeeding development was based on the continuous findings from the development prototype.

This prototype, though not completely adhering to the rigorousness of design science research evaluation, could be considered as a series of search iterations in regards to guideline 6 (see section 3.1.6). This is where the rigour-flexibility trade-off was most in favour of a flexible approach, by delaying the use of rigid evaluation methods until development could produce a "complete enough" prototype. By this, it is meant that the prototype was to reach a state where it was possible to extract valuable data from an evaluation.

5.4 First Prototype: Featural Anonymisation

The following is a description of the evaluation of the initial artefact prototype. This section will go through the selection and recruitment of informants, execution and its structure.

Recruiting Informants

A total of eight students from the Bachelor Programme in New Media from the University of Bergen was to attend pairwise in four separate evaluations. The students were informed that they were being approached due to their background and knowledge of the problem domain. They were informed that it was expected that they were able to reflect and discuss the technology based on this. A brief introduction to the technology was also provided as part of the recruitment effort. In order to increase the likelihood of achieving the desired number of informants, an evaluation remuneration of 500 NOK was offered. This remuneration was provided by the Prosopo work group.

Execution

The evaluation was to be carried out in Lauritz Meltzer's house at the University of Bergen, in suitable rooms providing privacy and a stable internet connection. The total evaluation was predicted to last about an hour and a half. There was two phones and one laptop available to the informants in order to perform the required tasks.

Evaluation Structure

Effectively, the evaluation was separated into three parts. This included an introduction to the technology, an anonymisation task with a corresponding questionnaire in the shape of a Google Form, and an interview for discussion and reflection on the technology. The parts were carried out in that respective order and were time-boxed to ensure a purposeful progression velocity.

5.4.1 Demonstration

Each evaluation consisted of a briefing, followed by demonstration of the technology using the Prosopo Android artefact. This would allow the users to use and experience the technology without any interference or direct influence. The participants was also informed of some of the limitations and requirements of the technology. This introductory part should take no longer than 10-15 minutes.

5.4.2 Anonymisation Tasks

Next followed a session of anonymisation tasks. In turn, each participant was to anonymise a face, which was likely to be familiar to both participants, and presented to the other. By choosing familiar faces, the anonymisation algorithm would be put against the worst case scenario for de-identification. This was interesting in regards to human face perception, which claims that holistic processing is more involved with familiar faces.

The participants were provided a random order of predefined values for the degree of anonymisation, which were set at 50, 60, 70 and 80 percent, helping to create a larger variance for analysis. Based on the output, the other participant was to answer a series of questions related to the performance of the anonymisation, i.e. evaluating the success of the two criteria of successful anonymisation (see 1.1.2). This was repeated four times for each informant, for a total of eight answers per evaluation. This was to take approximately 30 minutes.

In order to be able to answer Q1 it was essential to have human feedback on the output of the anonymisation algorithm. By pairing up the participants, it was

possible to allow for independent use of the technology. The only limitation given to the informants were that they should evaluate the quality of the manipulation prior to presenting the output. If the output was deemed of low quality (e.g. the face was heavily distorted), which was subject to the individual's judgment, then it should be discarded and a new face selected. This was due to the artefact being a proof of concept and it was more valuable to produce data when the algorithm worked as intended. In other words, it was purposeful to limit the likelihood of having the informants evaluate cases where the algorithm failed during one of its steps.

The form was divided into three parts. First it was determined if the anonymised face was identifiable and what degree of anonymisation was used. Based on whether or not the person was recognised, a second part aimed to have the informant evaluate the quality of manipulation. The form consisted mainly of short questions aimed to provide short qualitative observations. This included what made them recognise the person, whether or not the face looked natural, and if the image looked odd in any way. If the person was not recognised, the evaluating informant was asked if they knew who the person was after being informed of the identity. Finally, the output image was compared to the original and the informant was asked to compare the similarity of the faces, in order to put anonymisation into a face perception context.

The primary outcome of the anonymisation task was to generate data which could describe the current status of anonymisation performance. This means data which could help answer Q1, and demonstrate the efficacy and quality of the artefact. A secondary outcome was also to serve as a foundation for the reflective interview in the final evaluation stage.

5.4.3 Interviews

Finally, a joint semi-structured interview was carried out where the informants were asked a series of questions revolving around both the performance of the technology and its potential as an artefact in the problem domain, as well as other possible use cases. Additionally, the informants were asked to reflect around the artefact, attempting to challenge their views of the press photography and its usage in different news media. This was a way of evaluating the artefact's investigative potential, or RRI-potential, as a possibly controversial artefact in the problem domain. Using a joint interview, though maybe unusual in most experiments, are considered to be suited for this purpose. As the informants would be cooperating in the preceding parts of the evaluation, the continuation of this format would allow for a cooperative reflection space. This format also allowed for collecting observational data concerning the intercommunication of the informants throughout the evaluation. The interviews were to take about 45 minutes.

The interview guide contained three topics of discussion. The first part was aimed at discussing the immediate reactions and thoughts of the informants, after having gone through the first two parts of the evaluation. The informants were asked to describe the process, hopefully putting them in a mindset for reflection. They were then asked to vocally evaluate the technology in its current state, positives and negatives, allowing for suggestions for improvement or feedback in any form. As part of this, it was natural to ask questions related to specific cases which they had experienced during the demonstration and the anonymisation tasks. This was valuable for answering Q1.

After having concluded the first part, the informants were asked to envision a state of the artefact which performs close to perfect on the two anonymisation criteria, i.e. considering the ultimate solution for the problem that the proof of concept demonstrates. This would be helpful in order to perhaps push the informants to visualise more extreme or provoking thoughts in relations to the artefact.

The second part opened with a general question on use cases for the technology, followed by a discussion on problems and challenges that could arise in the given scenarios. Based on the responses and cases provided by the informants, there was eventually put specific focus on potential use areas within both social and news media. This was the main way of answering Q3.

The final part aimed specifically to investigate the value and position of the photography, and in particular with a perspective from photojournalism. An interesting question was to compare or contrast the anonymisation technology to traditional face de-identification methods (see figure 1.2). Furthermore, by contrasting this technology to the each individual's own perception of the photograph and perhaps points from the NPPA code of ethics, it could be possible to attempt to understand how the innovation of such an artefact could be met. This was relevant as an RRI-artefact, as discussed in section 1.2.1, and was how answering Q2 was possible.

5.5 Second Prototype: Holistic Anonymisation

The following is a description of the evaluation of the second artefact prototype. There are considerable similarities between the two evaluations, and only the differences will be highlighted in this section. The main point of the second evaluation was to investigate if there were any differences in anonymisation performance between the prototypes. The evaluation was thereby limited to a performance analysis, which limited the post-interview to how the artefact performed in terms of anonymisation. This greatly reduced the threshold for recruitment and complexities of execution, while still providing important data to the anonymisation performance.

Recruiting Informants

As mentioned, there was only a limited interview as part of the second evaluation. This removed the requirement of having to select and recruit informants tied to the original problem domain which was investigated in the first evaluation. It was decided to recruit a total of eight informants using my personal network, increasing the likelihood of sufficient participation. The chosen participants was not to be explicitly familiar with the artefact, excluding those who had previously been involved in demonstration and discussion.

Execution

The evaluation was carried out over two sessions with four participants each. The evaluations were to take about an hour, and held in a location which could provide privacy and a stable internet connection. One phone and one laptop was provided to each pair, which would allow the informants to carry out the required tasks.

Evaluation Structure

The evaluation of the second prototype shared the first two parts in terms of structure with the first prototype, with a demonstration and a series of anonymisation tasks. Besides having two pairs instead of one, the structure was otherwise similar to that of the first prototype. Having two pairs at a single evaluation was due to practical reasons, making execution easier. The benefits of a small evaluation groups were also somewhat diminished with the reduction of the post-task interview.

5.5.1 Demonstration

The demonstration followed the same structure as in the first evaluation, with a quick application demonstration, a brief explanation of the limitations and finally a set time of exploration and usage of the artefact.

5.5.2 Anonymisation Tasks

The anonymisation tasks used the same Google Form and the same structure as the first evaluation, where the participants worked in two pairs following the same progression. The pairs were held from communicating with each other in order to limit influences between them.

5.5.3 Interviews

The interviews only used the first part of the interview guide, which was an open discussion involving the thoughts and questions the informants had after using the artefact. Since the evaluation was dedicated to anonymisation performance, and the informants were not part of the problem domain (the journalism and media domain), the interviews were concluded after the first part.

Chapter 6

Results

This chapter will present the research results of this project, looking primarily at the data collected from the evaluation of the prototypes. The first section will discuss the results that emerged as part of the preliminary development process, which resulted into the two prototypes mentioned in section 4.2.

6.1 Development Prototype

It was through evaluation of the continuously evolving development prototype that the two anonymisation prototypes materialised. It is based both on visual feedback of output images and on the theories of human face perception that these two prototypes emerged as viable solutions. The visual feedback showed that it was difficult to create a prototype that performed on a high level on both anonymisation metrics.

- De-identification of the face
- Preserving face realism and image quality

Initially, the featural anonymisation prototype was developed. It performed very well on the second criterion. As mentioned in section 4.2.2 the strengths of this approach is primarily success in the second criterion. The second prototype was created in order to increase the performance regarding the first criterion. It was quickly evident that these criteria were opposing each other. In order to retain realism and reduce overall image quality loss, it was necessary to limit manipulation to the facial landmark coordinates. However, in order to anonymise the entire head region, it was not possible to rely solely on the segmentation available from facial landmark detection. In this regard, it felt natural to create two prototypes demonstrating the inevitable trade-off between the two criteria.

Comparing both a featural and holistic approaches in anonymisation allows for demonstrating the current possibilities and challenges of the artefact. It will also

be possible to evaluate which prototype is the best fit for the problem in its current state, and possibly determine which prototype should be developed further.

6.2 First Prototype: Featural Anonymisation

6.2.1 Execution

The evaluation of the first prototype was executed over two days, with two evaluation sessions each day. The eight desired informants were successfully recruited from the Bachelor Programme in New Media. However, out of the eight recruited informants only seven showed up at the agreed date and time. This lead to a single improvised evaluation where the role of the partner had to be simulated, and the interview was carried out in a one-to-one format.

6.2.2 Anonymisation Tasks

Each evaluation produced eight form answers on anonymised faces, except for the improvised one which only resulted in four. In total this gives 28 unique form answers. The immediate finding is fairly obvious. The first prototype performs very poorly on the de-identification criterion, where 86 percent of the cases lead to successful identification of the anonymised face (see figure 6.1). Additionally, 25 percent of those who failed to identify the face did not know the person after being informed.

Figure 6.2 shows at which degree of anonymisation the artefact managed to successfully de-identify the input face. Successful de-identification is defined as cases where the informants could not recognise the person's identity, while also stating that the person was known to them after being informed. As such, any anonymised person which are not known to the informant, is not considered as a successful de-identification.

Identification Sources

After having successfully identified an anonymised face, the informant is asked to describe what allowed for identification of the face. Based on these answers, several features and influences emerged as more common factors of identification. By analysing and categorising the answers, nine identifying sources were established. The categories are generalised to their theme, where the stated source could be a variant of this theme. For instance, a general statement like 'hair' and a more specific statement like 'hairdo' are both placed in the hair category.

- Hair
- Eyes
- Head shape
- Non-facial
- Mouth
- Overall similarity
- Nose
- Expression
- Skin

The most commonly stated source of identification was related to the hair of the individual, followed by the eyes and the head shape. The second most frequent source is the non-facial source category, which includes contextual information, clothing, background and other image qualities. The mouth region is next, alongside a category which is called overall similarity. The latter category consists of cases where the face is described as very similar to that of the original person who was attempted anonymised, without stating specific reasoning for the experienced similarity. At the bottom, the nose, facial expressions and skin rounds off the identification source frequency listings. The distribution is shown in figure 6.3.

Face Realism and Image Quality

In order to evaluate the experience of face realism and image quality, two questions were asked. The first one asked for a description of the face, and whether or not the image looked natural. The second asked for the informants to describe if there was anything wrong with the image. The first question yielded answers which typically entered into one of three categories of feedback.

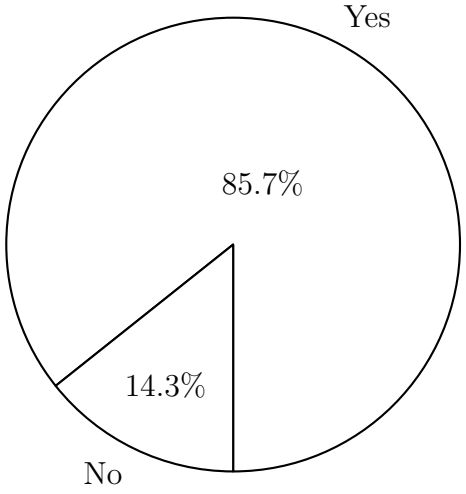


Figure 6.1: Face identification distribution for the Featural Anonymisation Prototype. Did you manage to identify the person in the image?

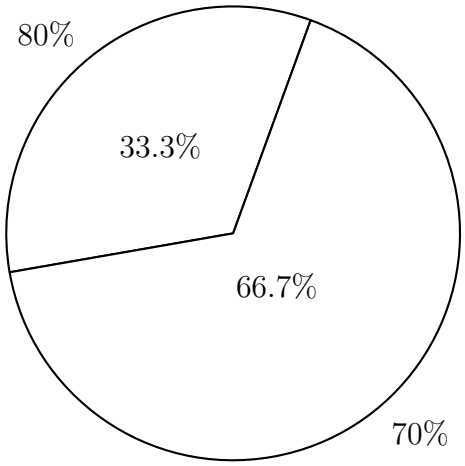


Figure 6.2: Anonymisation degree distribution for successful de-identification of the Featural Anonymisation Prototype.

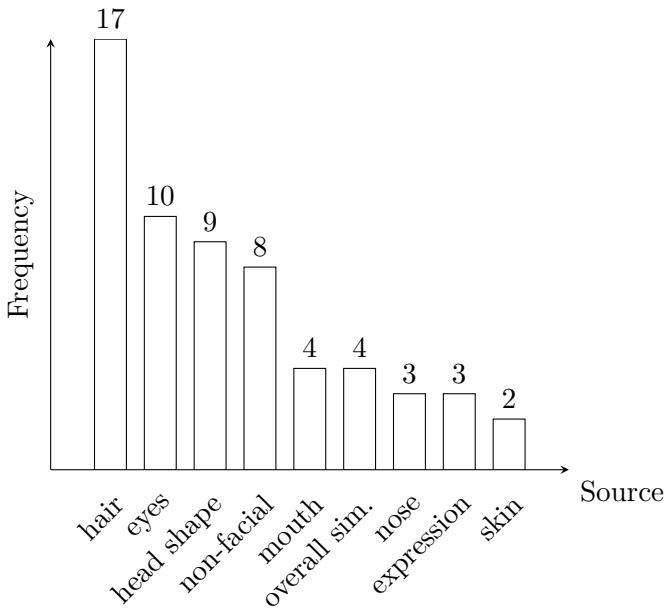


Figure 6.3: Identifying sources for the Featural Anonymisation Prototype. The number represents how many times this feature was given as a source of identification.

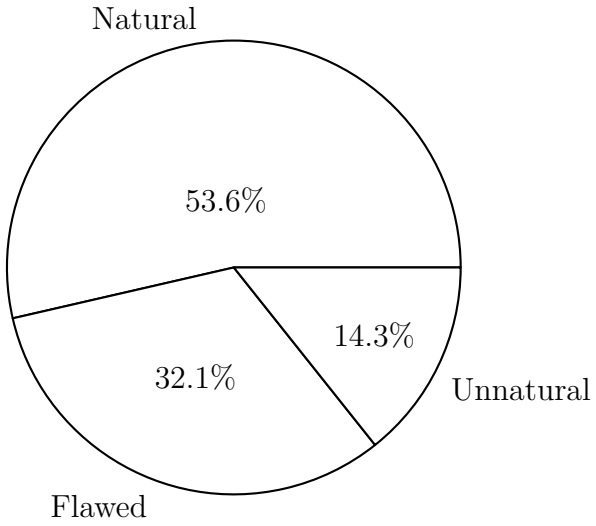


Figure 6.4: Face realism experience distribution for the Featural Anonymisation Prototype. What do you think of the face in the image? Does it look natural?

- Natural
- Natural, but flawed
- Unnatural

54 percent consider the output face to look natural and human, as shown in figure 6.4. A further 32 percent agree that the face looks natural, but also state that there are some considerable flaws and signs of manipulation or inconsistencies. The remaining 14 percent perceive the face to look unnatural. These flaws and inconsistencies are provided as an answer to the second question. What is wrong with the image? After analysing and categorising the data, seven main sources were found which contributed to reduced image quality.

- Blurring/focus
- Skin/colour
- Eyes
- Head shape
- Mouth
- Feature positioning
- Manipulation lines

The two biggest influences on image quality are blurring issues of the manipulated area, alongside visual inconsistencies regarding the skin of the person. These two categories are very much related, as they tell of differences between the area which have been manipulated opposing areas which has not gone through any form of alteration. Less frequent are accounts of facial features. This includes eyes and mouth looking unnatural, unfocused and distorted, while the head shape sometimes mixes with the background. Strange facial feature positioning is mentioned, which is predominately related to the previously mentioned category of unnatural faces (section 6.2.2). Furthermore, occasional seams can be located between manipulated and unmanipulated parts of the face. This is particularly evident when facial hair continues beyond the jawline, or when hair extends into the manipulated areas. See figure 6.5 for the frequency of reported sources.

Facial Similarity

In order to contextualise de-identification issues informants were asked to evaluate the similarity between the input image and the anonymised output image. Figure 6.6 shows the distribution of how the informants evaluated similarity of the face images, from very unsimilar (1) to very similar (5).

Considering 86 percent were able to identify the anonymised person, it is worth noting that 40 percent considered the faces to be more unsimilar than similar. Comparatively, 30 percent thought that the faces were more similar than unsimilar. Ad-

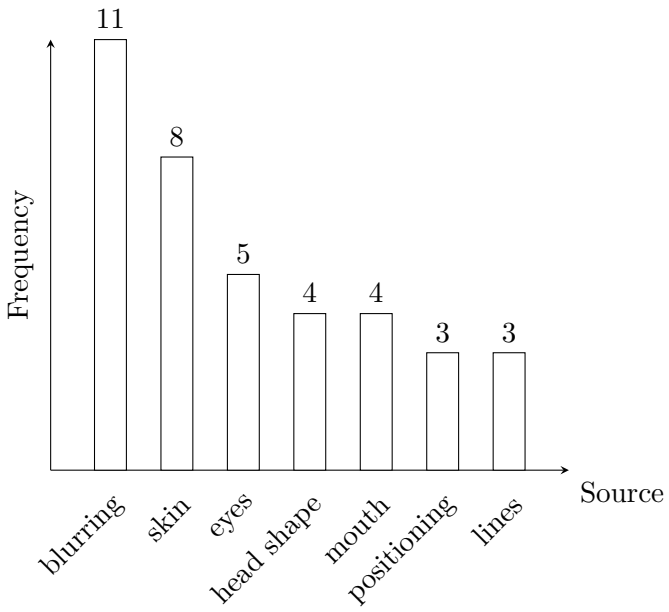


Figure 6.5: Image quality reducing sources for the Featural Anonymisation Prototype. The number represents how many times this feature was given as a source of quality reduction.

ditionally, only 14 percent considered the faces to be unnatural, and there is little correlation between the evaluation of a face as unnatural and as unsimilar.

6.2.3 Interviews

The interviews, though free flowing and flexible in progression, attempted to gather information about four main themes. First, the immediate thoughts and reactions after using the technology. Next was a discussion around possible use cases and the potential challenges related to implementation of artefacts using the technology. The final topic was a more in-depth discussion regarding potential for usage in news media, including a specific discussion around the value and position of the press photography with the artefact as the catalyst.

Reactions to Technology

The immediate reactions are generally similar across the different evaluations. The technology is considered interesting, exciting and entertaining despite only being a limited proof of concept, where the varying success of both manipulation and the low success rate of de-identification are mentioned. A specific issue which is given much interest is how the faces are identifiable despite not looking all that similar. Some of the identified sources are reiterated as probable reasons, where hair and contextual information are highlighted. Some of the informants believe that the technology would work better for unknown faces, specifying that they believe that

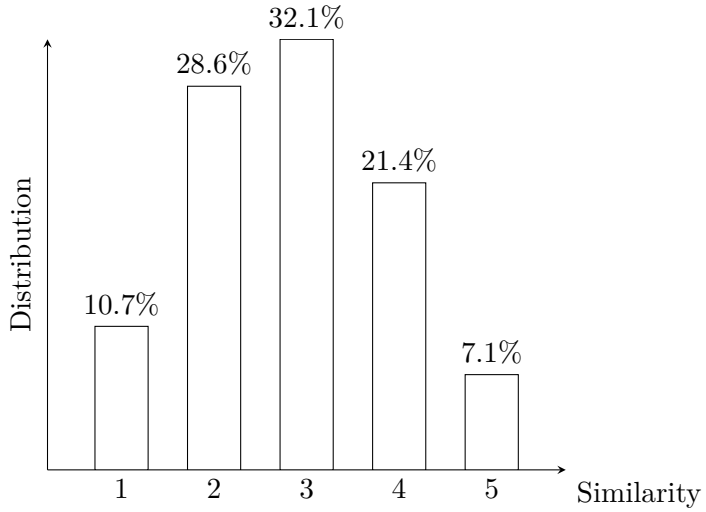


Figure 6.6: Face similarity distribution for the Featural Anonymisation Prototype, where 5 is very similar and 1 is very unsimilar.

the setting allowed for connotative reasoning and as such may have influenced the results. Additionally, some of the image quality reducing factors are mentioned. However, not all informants believe this to be negative, claiming that the ability to recognise manipulation is perhaps a good thing.

Use Cases

Quite a few cases were proposed by the informants regarding usage, both of problem domains and specific implementations. A particular domain, entertainment, was prevalent throughout all of the evaluations. Several other commercial ideas were also proposed, using the technology to solve problems not directly related to anonymisation. Additionally, several de-identification scenarios were proposed, including different usages in both news and social media, as well as in documentaries. Some informants even discussed possible undesirable usages, perhaps interesting to criminals and people who would have interest in fake identities.

Entertainment/Commercial The informants were very creative and experimental in thought when considering general use cases. A particular implementation was the game show "Guess Who?", which had multiple conceptions during the interviews. Other suggestions included realistic generation of profile faces for computer games, a proposal for a semi-blind dating application, and an application for previewing the result of cosmetic surgery. Several informants also thought of use cases for the technology in video and live camera capture.

De-identification Considering the technology as an anonymisation technique, the informants managed to identify both aspects and implementations valuable for

1) Can you see any use cases for this technology in the news media?
2) How would you compare this technology to traditional de-identification techniques (blurring, pixelation etc.)?
3) Can you see challenges or problems in using this technology in the news media?
4) How do you compare this anonymisation to other non-computational image manipulation in journalism?
5) Would you agree with a statement claiming that the press photography is a realistic representation of the world?

Figure 6.7: The investigatory part of the interview guide.

demonstrating the utility of the artefact. One of the most interesting being de-identification in documentaries, or other forms of expressive outlets, for touchy or personal stories in which the storyteller would prefer anonymity. This could help convince people to come forward to challenge or highlight injustice, corruption and exploitation in domains where it would otherwise not be safe. The case of Edward Snowden, who leaked confidential papers documenting the surveillance of the American people by the National Security Agency, was used as an example of this. Furthermore, groups like Anonymous and WikiLeaks were highlighted as types of organisations which could personalise their agenda while still maintaining anonymity and protection for members and sources.

Also, the technology was mentioned as a possible solution to allow for some privacy in social media, where sharing of pictures, often of private nature, is a prominent activity. If anonymisation could be performed without reducing the quality of the image, it could be possible to share photos without provoking concern of people who might not want it out for public viewing.

News Media and the Press Photography

The most discussion and controversy was found when talking about usage in the news media. Two main discussions were prominent after having defined certain use cases and implications for the news media, such as anonymisation of people in need of identity protection and increase in engagable story illustration. By comparing traditional anonymisation, or the lack of using illustration at all, a trade-off between value and risk was identified. As a follow-up to this, the usage of press photographies were debated and put into context by the image manipulation of the artefact. The part of the interview guide which was used to facilitate these investigations is shown in figure 6.7.

Comparing Anonymisation Techniques Concerning anonymisation, or de-identification, of individuals in news reports, the traditional methods has often been usage of blurring or pixelation techniques (as illustrated in figure 1.2), or simply using neutral avatars only pertaining to gender and age. When comparing the traditional anonymisation approach to the one of this artefact, the informants were somewhat divided in their response. The informants agreed that the idea of using illustrations in a story could increase its value. At the same time, using photographies in sensitive cases also includes risk. As such, anonymisation is a trade-off between the value it offers to a story, and the risk of breaching the ethical and legal implications of identification. Comparing the traditional methods of anonymisation to the artefact, the risks connected to the artefact are considered to be greater. At the same time, the value of the illustration is considered, by the majority of the informants, as much higher. Some of the informants would even go so far as to claim that the story is told more honest and unbiased by using actual footage/depiction (though anonymised) of a scene or person. Others would question if the value is all that important when existing techniques can do the job in a satisfactory way. This is particularly the case in regards to a specific challenge, which is the first one to be discussed in section 6.2.3.

The Position of the Press Photography In a way to spark debate regarding the press photography, a discussion on manipulation and editing, such as points 5 and 6 from the NPPA Code of Ethics (see figure 1.1), was initiated. The discussion started off by creating an example of a pre-capture situation specifically orchestrated to give specific meaning to a photography and comparing it to the post-capture processing of the anonymisation artefact. This way, the informants started discussing the differences between pre-capture and post-capture ”manipulation”. Though the informants were somewhat agreeing to the fact that pre-capture orchestration could be called manipulation, there was still essential differences to post-capture manipulation or editing. It was claimed that even though it might be unethical to perform such orchestration, it is a type of manipulation it is possible to deal with or be made aware of.

In order to give some more context to the discussion, post-processing techniques like colour correction and light intensity adjustments were also compared to the image manipulation of the artefact. This is also post-capture editing which most of the informants found to be acceptable, at least in most cases. What should be considered unethical manipulation could vary from case to case, according to some. For instance, a scenario put forward by an informant says that there should be a different threshold for ethical editing if a photograph depicts a war situation as opposed to a politician in parliament.

When asked if they perceived the photography as a realistic representation of the world, there were some differences in the response. While some agreed that a photograph depicts the world at a given location and time, and given the background and conditions behind its capture, others questioned whether or not a photograph could truly represent the world.

One of the main messages, was that the image is not supposed to be altered or manipulated in a way that it changes its original meaning. It became apparent that the informants were often discussing their opinions, and not necessarily stating them as their own believes. It seemed to be that they were starting to reflect on the matters that where asked. There was even a couple of quotes stating that they were not sure if they would have the same opinion if they were to be asked the same questions in a couple of days, and as such supports the potential of the RRI-perspective of the artefact.

Challenges

The most apparent and immediate concern of the informants involved is the possible scenario where the algorithm would create faces which could resemble unrelated individuals. Creating suspicions, and possibly implicating these people in cases of undesirable nature, would naturally not be a wanted outcome. Through this challenge, the informants highlighted the need for an identifier for facial anonymisation, i.e. the anonymised result needs to be easily recognisable in the output image.

Another worry was that the generated faces would serve as a source of generalisation and generation of stereotypes. Using an anonymised face as illustration in a story about a paedophile or rapist, this could possibly create a persona of such a person. It would be undesirable if certain ethnic groups, particularly distinctive groups of the population, or any other visually identifiable stereotype, were to be negatively generalised due to utilisation of the artefact.

Exploitation, or misuse, of the technology was also discussed. Allowing criminals, or people with unscrupulous intentions, to essentially create fake personalities could help increase trust to the potential victims.

A more unique and intricate problem was presented by a single informant. This individual questioned whether or not it was positive to have created faces as a normal part of every day image consumption. If the consumers of, say an advert, are not consciously aware of the artificiality involved, could this affect how we relate to stories? For instance, by using actors who are supposed to represent a product or organisation, the advertisers are not really advertising their true nature. The same could be said about McDonald's who are known for retouching their burger illustrations, as mentioned by another informant. All informants stated the absolute importance of having some evident way of recognising an anonymised face.

6.3 Second Prototype: Holistic Anonymisation

6.3.1 Execution

Out of the eight desired informants, it was only possible to recruit four. As such, the two planned evaluations were limited to a single evaluation. In order to compensate for the lack of participants, it was agreed upon to perform two rounds of the anonymisation tasks. This way, the two evaluations would have an approximate amount of answers. This was not an ideal solution, but it was deemed acceptable considering the practical difficulties of a second recruitment effort and the following evaluation.

6.3.2 Anonymisation Tasks

The Google forms used to document the anonymisation performance are the exact same as those in the featural prototype evaluation. As each pair performed two rounds of the anonymisation tasks, a total of thirty-two form responses were produced. Interestingly, 53 percent of the answers resulted in a failure to identify the person which had been anonymised (see figure 6.8), which is a considerable higher number than that of the first prototype. Out of these cases, 20 percent did not know the person after being informed of his/her identity. Still, 47 percent of the faces were identified, which means that the de-identification performance is still fairly unreliable.

Figure 6.9 shows the degree of anonymisation at which the artefact managed to successfully de-identify the source face. Successful de-identification is defined in the same manner as described in section 6.2.2.

Identification Sources

Similarly to the first evaluation, the answers were analysed and categorised into related groups which can help understand the sources of identification. An interesting aspect of the second prototype evaluation is the difference in identifying sources. A total of six categories were found and are listed below.

- Overall similarity
- Eyes
- Mouth
- Hair
- Head shape
- Nose

The most frequent source of identification is overall similarity. Two facial features also proved to be quite frequent, which were the eyes and the mouth. Hair is found next, followed by the head shape and nose. See figure 6.10 for the frequency distribution.

Face Realism and Image Quality

The same categories for face realism evaluation were found in the second evaluation, reiterated below.

- Natural
- Natural, but flawed
- Unnatural

50 percent considered the produced faces to be natural and humanlike, whereas 31 percent pointed out that the the faces had some problematic inconsistencies. 19 percent of the answers claimed that the face was unnatural. This is shown in figure 6.11. Out of all the answers, six categories emerged as problematic in terms of the image quality.

- Blurring/focus
- Skin/colour
- Manipulation lines
- Eyes
- Mouth
- Ears

The three largest groups were about equal in size. These were blurring issues, skin and image colour inconsistencies, and manipulation lines which highlighted the seams of the areas which had been altered. The three other categories were also fairly similar in frequency which were facial features, namely the eyes, mouth and ears. The frequency is depicted in figure 6.12.

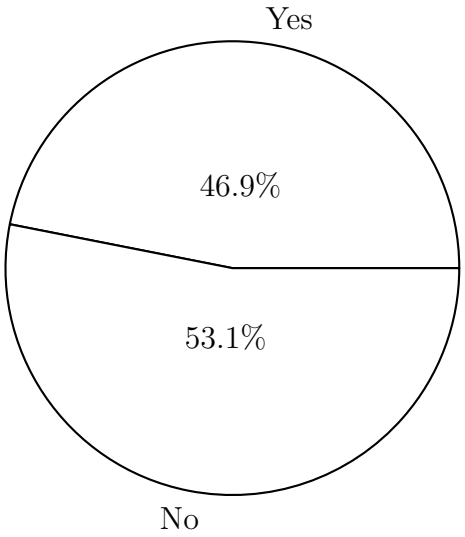


Figure 6.8: Face identification distribution for the Holistic Anonymisation Prototype. Did you manage to identify the person in the image?

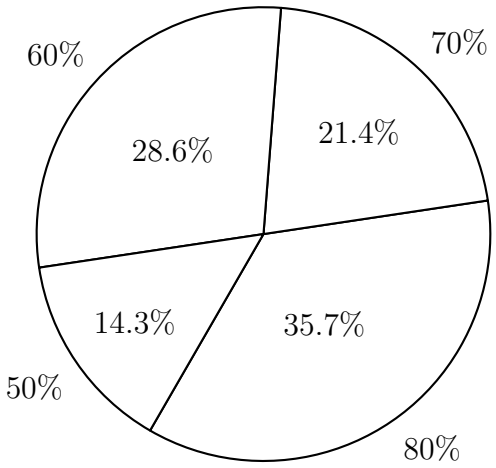


Figure 6.9: Anonymisation degree distribution for successful de-identification of the Holistic Anonymisation Prototype.

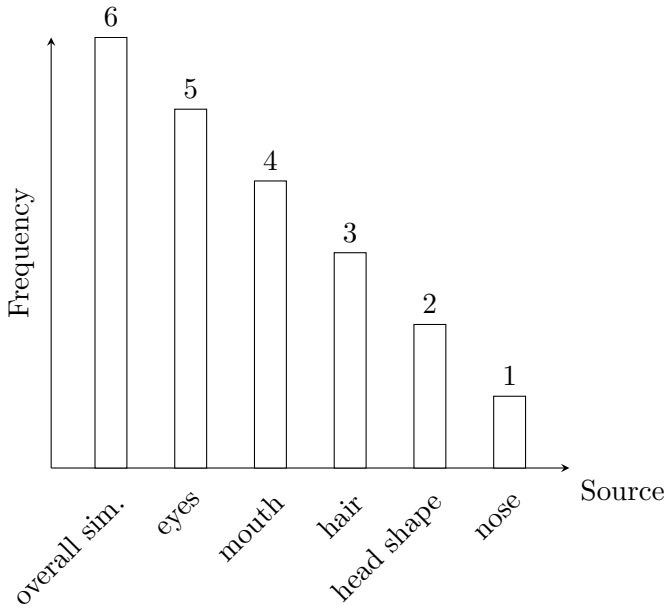


Figure 6.10: Identifying sources for the Holistic Anonymisation Prototype. The number represents how many times this feature was given as a source of identification.

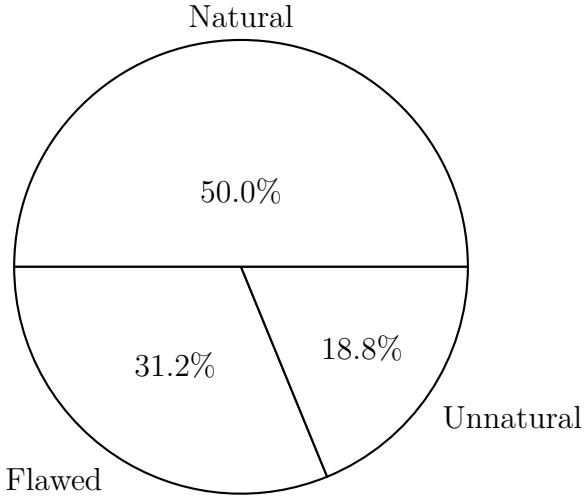


Figure 6.11: Face realism experience distribution for the Featural Anonymisation Prototype. What do you think of the face in the image? Does it look natural?

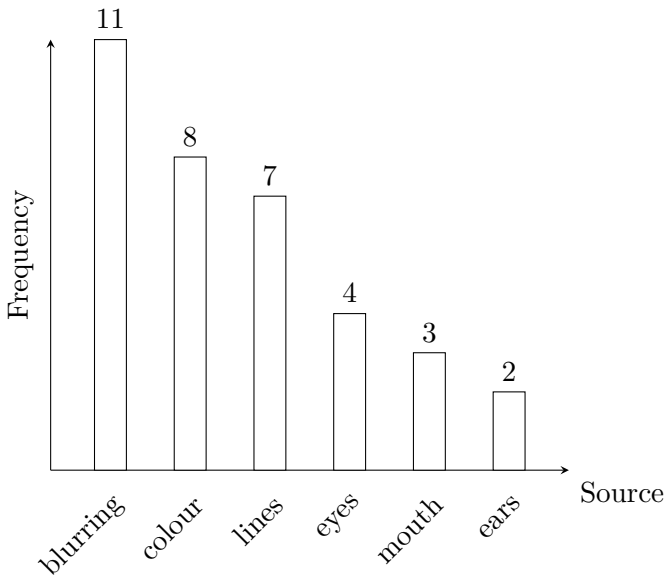


Figure 6.12: Image quality reducing sources for the Holistic Anonymisation Prototype. The number represents how many times this feature was given as a source of quality reduction.

Facial Similarity

Figure 6.13 shows the distribution of experienced facial similarity of the holistic prototype, from very unsimilar (1) to very similar (5).

It is interesting to note that a total of 34 percent experienced that the facial similarity was below similar, while only 28 percent considered the faces to be similar. Only a total of 3 percent considered them to be very similar. This could possibly mean that the holistic prototype is better than the featural prototype at creating a markedly different face.

6.3.3 Interviews

The interviews were carried out just like in the first evaluation, with a fluid development allowing the informants to talk about what they found to be most interesting. The main focus was naturally on the performance of the artefact, both regarding de-identification and aspects of face realism and image quality.

Reactions to Technology

Much like in the first evaluation, the immediate reactions were directed at the amusement and entertainment value which it provided during the anonymisation tasks. The informants were quick to discuss the perceived identification sources which they had found while evaluating the faces. They were also very interested in how fairly different faces could still look like the same person. When concerned with the image

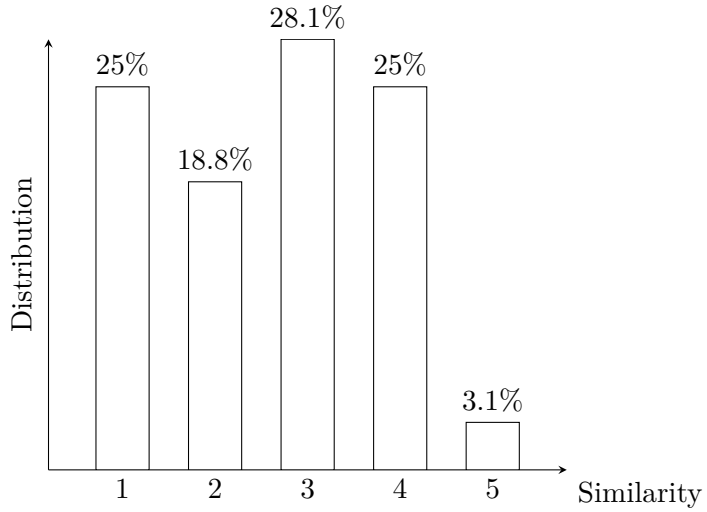


Figure 6.13: Face similarity distribution for the Holistic Anonymisation Prototype, where 5 is very similar and 1 is very unsimilar.

quality, certain areas were highlighted. This included blurring and manipulation lines in particular. The holistic prototype created some cases where the face itself looked very natural, but traces of manipulation were quite evident.

There was also an agreement that the faces were either quite similar from input to output, or there was some minor detail or feature which led to identification. In one case, where Albert Einstein was anonymised (figure 6.14), parts of his very characteristic hair was visible post manipulation. This informant was convinced that without this detail it would not have been possible to identify him. Several similar cases were mentioned. There were also cases where the informants were genuinely impressed by the result, where both de-identification, and the face realism and image quality criteria, had been achieved well.

Figure 6.14: Albert Einstein anonymised holistically.

Chapter 7

Discussion

This chapter is dedicated to discussing the results, methodology, technology and research theory involved in the project. By reflecting upon evaluation results the research questions will be answered, challenges and problems identified, and future work and development proposed.

7.1 Research Questions

These are the research questions which are to be answered. This will be done by looking at the theoretical background, while also analysing the results of the evaluations.

Q1: Is it possible to create an artefact capable of automatically anonymising faces using an average face morphing approach?
Q2: Can the artefact be used to investigate the views on normative practices in manipulation and usage of face photographies in the news media?
Q3: Are there other possible use cases for the artefact or its technology?

7.1.1 Answering Q1: Anonymisation Technology

In order to discuss the performance of the artefact as an anonymisation technology, it is necessary to look back at the definition that was presented as part of the introduction of this thesis. The success of anonymisation was defined as follows.

- De-identification of the face
- Preserving face realism and image quality

An additional requirement was that the algorithm should be fully automatic from input to output, where the user only provides an input image and the degree of anonymisation.

It should be emphasised that the findings in the following sections are not generalisable and they are not sufficient to provide a definite answer to the question at hand. However, they do provide a foundational insight into the current state of the artefact and serves as a foundation for discussing aspects which could be valuable for understanding how to better achieve a satisfactory level of anonymisation. This means identifying the obvious problem areas and challenges which are possible to find based on the design of the evaluation.

Automation

The featural prototype uses established landmark detection techniques, which allows for manipulation on sophisticated premises. This is the foundation for automatic manipulation. The holistic prototype uses the same landmark detection, but also requires less sophisticated methods for creating landmarks for detecting the entire head. This could be done quite successfully by removing the requirement of automation, i.e. by manually plotting coordinates for facial features which are not automatically detected by landmark detection algorithms, such as the ears and hairline.

As of now, both the prototypes have managed to reach the level of automation as initially described. However, there are cases which will lead to failure in producing an output image. This challenges the concept of full automation, meaning it will not always produce an output for a given face image. Still, the artefact should be considered to have achieved a satisfactory level of automation as a proof of concept. The limiting factors for completeness, and possible ways of solving them, will be discussed in section 7.5.

De-identification

Looking at the evaluation of the two prototypes it is apparent that the success of de-identification is far from reliable. The first prototype performed very poorly, where the informants successfully identified 86 percent of the faces. The second prototype performed considerably better, where only 47 percent were able to identify the person in question. The performance difference between featural and holistic anonymisation was also highlighted by user interviews. The featural prototype was generally labelled as unsuccessful, while the holistic one was considered incomplete.

It is worth nothing that both prototypes were put to the ultimate de-identification task, where widely familiar faces were anonymised. According to human face perception studies, as mentioned in section 2.1.1, there is evidence supporting that holistic processing is more dominant in recognition of the familiar. This might be why the holistic method performs better, but could also be due to the informants

choosing different images and that some informants are better at recognising faces than others.

It might still be fair to assume that there is potential located at looking into a more refined version of holistic anonymisation. This means handling all sources of identification, not limiting the anonymisation to a mere facial feature restructuring. As such, this approach is proposed to be the best solution for achieving a reliable anonymisation algorithm.

Degree of Anonymisation It is difficult to find evidence that supports that the degree of anonymisation has affected identification distribution, and it would be presumptuous to attempt to attribute causality to this, rather than co-variation of the limited sample size and the relatively uncontrolled evaluation setting. However, it would still be interesting to look into this in a dedicated study with a more complete and consistent prototype, i.e. where the threshold for when anonymisation can be achieved is located, if one even exists. Based on informant feedback certain faces require a larger anonymisation degree than others, which indicate that there is no general percentage threshold for anonymisation.

Identification Sources When comparing the two prototypes and the identification sources which are reported in their respective evaluations, several differences can be found, as demonstrated by table 7.1. The most dominant identification factor from the featural prototype is almost eliminated in the holistic anonymisation. The prevalence of hair and non-facial sources are drastically reduced from 71 to 20 and 33 to 0, respectively. This is likely in large part due to the head cropping of the holistic prototype, limiting the majority of contextual information, while hair is visibly reduced to smaller parts surrounding the head (as shown in figure 4.2). A similar result can be seen for the head shape, which is also considerably lower at 13 in the holistic prototype opposed to the 38 of featural anonymisation. This could be attributed to the holistic prototype's approach of inserting a head, instead of a face, into the original image.

An interesting source is the overall similarity category. It is possible that the holistic approach makes it harder to point out specific sources when trying to determine the reason for identification, but this could also be due to having different groups involved in evaluating the two prototypes. There is also no guarantee that the proposed identification sources are real, meaning that even though the informants perceive a certain feature to have identifying qualities, there might be more complexity involved without the informant being able to express it in writing or verbally. In practice, this means that even when all problems stated by the informants have been dealt with, there are no guarantees that de-identification will be

Source	Featural (%)	Holistic (%)
Hair	71	20
Eyes	42	33
Head shape	38	13
Non-facial	33	0
Mouth	17	27
Overall similarity	17	40
Nose	13	7
Expression	13	0
Skin	8	0

Table 7.1: Comparison of the percentile distribution of identification source prevalence based on all successful identifications from both prototypes.

successful.

It is apparent from the interviews from the holistic evaluation that certain faces have very distinctive features which sometimes are visible through anonymisation. At the same time, some faces are just very similar as a whole after manipulation. For instance, George Bush and Will Smith are recognised immediately, Albert Einstein is only recognised due to his characteristic hair, and other faces are very hard to identify. However, again, it is hard to claim that these cases have a clear identification cause, but could just as likely be related to the personal evaluative capability of the informant involved.

Face Realism and Image Quality

The findings on face realism and image quality shows more promise compared to de-identification. In both prototypes the experiences of face realism are fairly high, where 80 percent of the faces are characterised as natural and humanlike.

Natural Anonymisation In both prototypes approximately 50 percent of the faces were described as convincing face representations with limited or no obvious inconsistencies, uncanniness or signs of manipulation. These faces usually performed well on seamless cloning, which is when the anonymised face is placed back in the original image. This technique is crucial in colour correction and adjusting differences in illumination and texture, allowing for the new face to smoothly blend into the destination face. It is assumed that the best faces, in terms of realism and image quality, are those where each step of anonymisation have had generally favourable outcome (see figure 4.1 for these steps).

Flawed Anonymisation Of the remaining 50 percent, approximately two thirds claimed that despite evidential signs of manipulation, the anonymised faces still

looked natural and human to a certain degree. The different flaws include manipulation lines and significant differences in skin colour and texture, which are most commonly stated, in addition to smudged eyes, inconsistencies regarding the mouth and teeth, and head shape. When looking at the anonymisation system (figure 4.1), it is likely that for these flawed outputs, one or more of the anonymisation steps have encountered a challenge which it has failed to handle optimally. More specific analysis of each anonymisation, and its steps, would be required to determine the complexities of these challenges. However, many of these challenges have been identified as a result of the search process of the development prototype, and are presented in section 7.2.

Unnatural Anonymisation The rest of the cases described their relevant faces as strange, malformed, unnatural and inconsistent. This was often the case when facial features were out of proportion or position, due to poor matching between the input face and the faces of the database. This implies that the weighting of attributes does not always produce the best set of similar faces for each input. The only mentions of uncanniness are documented in these cases, where the faces are deemed to be strange, unnatural, and similar descriptions which relate to the uncanny. Many of these faces originate from cases where the anonymisation algorithm has failed substantially on one or more of its step. For instance, the average face and the cropped face might be very different, meaning facial similarity calculation has performed poorly. This could lead to strange positioning and proportionality of facial features in their morphed outcome.

In an attempt to reduce the amount of these types of faces, as it is most beneficial to look at the cases where a proof of concept system works reasonably well, informants were discouraged to use images which resulted in a ”poor” outcome. However, this was subject to the judgment of the informants, who might have had different ideas of how to define a ”poor” outcome. Still, these faces help put the performance of the other anonymisations into perspective.

Image Quality The most consistently reported issue is blurring or lack of focus in the manipulated areas, which is mentioned as the most disruptive effect on image quality. It is also mentioned as a factor in reducing face realism, particularly in cases where the blurring leads to a very distinct separation of the manipulated area and the original image.

For both prototypes manipulation lines are mentioned as image quality reducing, and particularly in the holistic prototype these lines are apparent. For the latter, this is a result of landmark misalignment, which is unavoidable with a primitive method for creating the additional landmarks around the head. For the former,

Source	Featural (%)	Holistic (%)
Blurring	39	34
Colour/skin	29	25
Eyes	18	13
Head shape	14	0
Mouth	14	9
Positioning	11	0
Manipulation lines	11	22
Ears	0	6

Table 7.2: Comparison of the percentile distribution of image reduction source prevalence based on all answers from both prototypes.

this is often the case when landmarks separate facial hair into manipulated and unmanipulated sections. See table 7.2 for a comparison of image quality reducing sources between the two prototypes.

Conclusion

It is difficult to conclude anything of significance from the findings of this thesis. However, there are some indications that a holistic approach, using more sophisticated technologies and techniques, could achieve success according to the definition of facial anonymisation. This is based on its increase in success regarding de-identification, while still providing a fairly high experience of facial realism and image quality. It is not unreasonable that parts of this improvement can be attributed to the cropping of the head from the background, as in the holistic prototype. Arguably, this could be described as circumventing the problems, rather than solving them. However, as a proof of concept, this is considered a way of maintaining a reasonable level of feasibility, while still investigating the subject.

Another thing which has to be pointed out is that the design of the anonymisation tasks used in the evaluation will not lead to a complete understanding of all possible influences and problems related to the anonymisation criteria. It is not unreasonable to suspect that tattoos, piercings, birthmarks and other distinguishing features could lead to challenges, despite the informants not mentioning them in a considerable regard. More of these issues could become apparent after more usage and evaluation.

However, in section 7.5 I will argue that there are many ways to develop a more sophisticated holistic anonymisation algorithm, one that might reach a level of sufficient performance for solving the anonymisation problem, or at a level that could at least serve as a catalyst for reflection or further/new development, either as an extension or in a tangential manner.

7.1.2 Answering Q2: Investigative Tool

This research question is answered by looking at the interviews from the featural anonymisation prototype. The question asks if the artefact has any potential to identify and challenge the views on usage and manipulation of facial photographies in the news media. This means encouraging the informants to reflect upon the existing standards, as well as their own beliefs and convictions, regarding the ethical and practical sides to the utilisation of press photographies.

The first thing to highlight is the selection of informants. The participants of the evaluation are not journalists, but instead belong to a study which looks at new ways of communicating media content, as students of the Bachelor Programme in New Media. However, their background is found to be of a good match in order to test out the investigative qualities of the artefact within the problem domain, as it is very much relatable to that of a journalist.

The section of the interview which lead the informants into the domain of the press photography is divided into several questions and scenarios. Figure 7.1 shows the pre-defined questions which served as a foundation for this investigation. As a semi-structured interview, these questions were often related to statements and scenarios which had surfaced previously, for instance from intercommunication during the anonymisation tasks.

The most interesting findings from the interview is the conception of anonymisation as a trade-off between risk and value. There is considerable risk involved in using a face-retaining anonymisation technology, mostly due to the fact that individuals with no connection to the case could be unintentionally implicated. There is also a value perceived, where the majority claims that a genuine face does offer something to a story in terms of creating stronger emotional engagement, and that it providing more information to the consumer.

The informants discussed the aforementioned trade-off amongst themselves, dis-

1) Can you see any use cases for this technology in the news media?
2) How would you compare this technology to traditional de-identification techniques (blurring, pixelation etc.)?
3) Can you see challenges or problems in using this technology in the news media?
4) How do you compare this anonymisation to other non-computational image manipulations in journalism?
5) Would you agree with a statement claiming that the press photography is a realistic representation of the world?

Figure 7.1: The investigatory part of the interview guide.

cussing the pros and cons and the challenges that surfaced. This was also the case when they discussed whether or not the photograph is a representation of the world, and where the borders are drawn regarding manipulation and photometrical correction when post-processing these photographies. Where some informants agreed that the border is fairly easily drawn in accordance with the traditional colour correction and light adjustment statement, some argued that the border could be more fluid depending on the nature of the story.

It was interesting to see how the stories evolved and transformed across a relatively short timespan. Though there were some very foundational values at base, such as an absolute requirement of honesty in presentation, and protection of the identification of individuals, it was still evident that some informants experimented with the idea of such a technology in its full potential. This argues that the artefact possess the ability to provoke controversial thought.

Conclusion

Despite that the informants involved in the evaluation could be considered as only a stepping stone into the intended problem domain, it is still fair to claim that the artefact manages to incite reflection when put into context of a hypothetical implementation in the problem domain. As such, it is argued that the utility of the artefact is demonstrated according to Q2. It would, however, be interesting to see how journalists and the editorial environment would respond in terms of the implicated ethical issues, and if they share the same views on the mentioned risk-value trade-off.

7.1.3 Answering Q3: Use Cases

Based on the feedback from the informants, it was quickly evident that they believed the artefact to have potential in other domains. In particular, several entertainment concepts were pursued, particularly the face guessing game which was dubbed ”Guess Who?”. This underpins the general reaction which was immediately mentioned by the informants, that the artefact is fun to engage with.

Some informants found specific commercial ideas like a beautifier application, or the cosmetic surgery preview application mentioned in section 6.2.3. One informant stated that they would have used such an application, even if just out of curiosity.

These artefacts were also mentioned in a social media context, providing an alternative for bloggers or content creators from nations or social groups which could be put in danger, or unfortunate situations, based on the content they produce. By anonymising their faces, they would be able to communicate their content without putting them in risk of unfavourable consequence, while still having the possibility

to create a personal relation with their intended audience.

Related to the above case, and comparable to the social media de-identification algorithms mentioned in section 1.1.1, is the potential to combat privacy issues in social media image sharing. Such a service could be created where users of social media applications could anonymise faces of individuals which may not be interested in having their faces broadcast without their knowledge or consent.

These scenarios themselves do not prove that the artefact has utility within any of the mentioned domains, but could serve as examples of where the technology involved may find a purpose. By providing these possible use cases, the artefact increases its relevancy, even though the relevance is not directly linked to the specific problem domain. Particularly interesting, are the use cases which are connected to social media content production and sharing, which is a genuine problem in terms of de-identification, as discussed in section 1.1.1.

Conclusion

Several scenarios have been theorised where the artefact could be implemented to solve a problem. The use cases are diverse, but specifically highlighted is the potential within entertainment, commercial facial manipulation services, social media de-identification, and anonymisation functionality for non-traditional media content creators. This argues that the utility of the artefact is demonstrated, even though the potential might lie outside the initial problem domain. Further and more distinctive work must be put into this effort in order to validate these scenarios.

It has to be specified though, that the demonstration of utility regarding Q2 and Q3 are not to be compared. Where Q2 demonstrates the utility of the original problem domain, Q3 is an exploration into other possible relevant domains.

7.2 Problems and Challenges

This section address the technical problems and challenges regarding performance of the facial anonymisation algorithm, meaning obstacles and incompleteness in relationship to the definition of facial anonymisation from section 1.1.2. Several problems have been identified from user feedback, and some have been identified through the creation and evaluation of the development prototype.

7.2.1 Blurring

Blurring was mentioned in both evaluations, and was a problem which was also observed in the development prototype. There are two obvious reasons behind this problem. Firstly, face averaging creates a smoother face where distinctive facial

textures are watered down. For each face used as input for the averaged face, this effect increases. This will naturally lead to less distinctive faces, something which is also considered beneficial, as this helps the de-identification process. However, it is also a problem that the faces in the 10K US Adult Faces Database are very small, being only 72 pixels in resolution and having a height of only 256 pixels. This means that when providing input faces which are of a much larger size, there will be resizing of the faces in the database used for face averaging. When this averaged face is then placed in the much larger face, there will be fewer pixels in the manipulated area then in those surrounding it, resulting in a blurred/unfocused face. A good example of this can be seen in figure 7.3b.

7.2.2 Colour/skin

Another frequent problem which was evident based on informant feedback was inconsistencies in skin colour and texture. For the holistic prototype this was particularly evident, due to the seamless cloning not being able to easily identify that the new face should ”belong” in the destination face, but rather as an object to place into the background. This is due to having an entire head, which includes parts of the background, hair and areas of mismatched facial features (particularly neck and ears), which does not necessarily fit into the original image.

For the featural prototype the problem is mainly due to the lack of holistic manipulation, meaning that when swapping in a much more smooth morphed face (where face means all features within the convex landmark hull, as described in

Figure 7.2: An example which illustrates problems regarding skin colour. Notice how the skin differs greatly in both colour, illumination and texture from the forehead and down. This example is from the featural anonymisation prototype.

(a) The beard is separated into manipulated and unmanipulated parts, and a seam in the skin is visible on the left side of the face. This example is from the featural anonymisation prototype.

(b) Seams can be seen around the head, particularly on the neck and forehead. This example is from the holistic anonymisation prototype.

Figure 7.3: Examples which illustrate manipulation lines, which contribute to evident signs of manipulation. Both faces are also overly blurred, which amplifies this effect.

section 2.2.2), it will result in an unnatural change in skin texture. This can be seen in figure 7.2. Even though seamless cloning does a good job at correcting colour differences there will often be certain details that visually separates the original and manipulated areas, such as considerable differences in texture and illumination.

7.2.3 Manipulation Lines

Manipulation lines are defined as seams and borders which are very disruptive to image quality and highlights inconsistency between manipulated and unmanipulated areas. These are very visible in the holistic prototype due to using a large rectangular area around the head for use in manipulation. In the featural prototype these lines are most visible when objects within the convex landmark hull extend beyond this area, where examples are glasses, facial hair, and hair which covers parts of the hull. These lines are hard to guard against as long as it is not possible to coordinate the entire face into landmarks which can be used to map all facial features across different faces. Examples can be seen in figure 7.3.

Figure 7.4: The disproportionate positioning of the facial features lead to a strange-looking face. This example is from the featural anonymisation prototype.

7.2.4 Facial Features

Some informants mention that there are problems with certain facial features. The eyes are sometimes described as bland, or inanimate, with almost no distinctive eye colour. This mostly due to the mixture of many different eyes, resulting in what could be described as impersonal eyes lacking the uniqueness that the eye colour provides to a face. The mouth and teeth are occasionally overlapping, and for open mouths the teeth are sometimes seen as a continuous row, rather than a row of individual teeth. The latter case is simply a result of not being able to map each tooth from all the faces involved in the anonymisation process, while the former is a result of one or more of the selected faces used in the face averaging having an open mouth when the mouth of the input image is closed. Facial features are also occasionally out of position or proportion, as illustrated in figure 7.4. This highlights that the facial similarity calculation is flawed and does not work optimally for all cases.

7.2.5 Scope

There are several cases, meaning some faces in potential input images, where the current solutions are not able to produce any output, or where the output is of such a poor quality that the image is ruined. These cases should be handled in order

to fulfil the concept of a fully automated anonymisation algorithm. Many of these cases were intentionally circumvented as part of this project, due to this being a proof of concept, where the scope had to be limited in order to maintain a feasible complexity level for this project.

Another concern regarding the thesis scope is that the evaluations are very limited, including a small sample size and an uncontrolled explorative research design. Even though this was intentional, in order to focus on the natural variety of feedback when used in an uncontrolled setting, this means that there may be additional undiscovered problems. Uncovering these problems might increase the complexity in developing a robust solution in the future. Additionally, this implies that the findings are not generalisable and limits the conclusive power they provide.

7.3 Discussion on Methodology

This section will look at the research methodology involved in the project. This includes a look at the results of using the design science research framework, alongside a discussion of the artefact evaluations.

7.3.1 Design Science Research

The design science research framework was a natural choice of methodology for developing an innovative artefact, like this thesis aimed to do. However, due to the explorative nature of the artefact, a more flexible approach was chosen for the design science research implementation. In section 3.2 the guidelines of Hevner, March and Park (2004) were discussed in relations to the development of the artefact, its evaluation, and its desired outcome as a work of research. In this section, the results of using these guidelines will be assessed.

Guideline 1: Design as an Artefact All the artefacts mentioned in section 1.3.2 have been developed to the standards which were decided upon. These are:

- Anonymisation artefact - proof of concept
- Prosopo API - complete implementation
- Prosopo Android - complete implementation

As such the requirements of this guideline is considered to be successfully met, having produced the proof of concept artefact which was the initial goal of the project, alongside the auxiliary artefacts required to demonstrate the state of the main artefact.

Guideline 2: Problem Relevance The main relevance of the artefact is as an RRI-artefact which aims to investigate the reactions and controversies involved in its usage in the problem domain. An attempt has also been made to identify more areas of relevance, in which a more complete artefact could be of interest. This includes the artefact as an anonymisation system which is explored in Q1, as well other use cases, such as those proposed as an answer for Q3.

Guideline 3: Design Evaluation A flexible and explorative approach was taken in regards to the evaluation design. The artefact was first evaluated during its infancy using an assess-refine search process, attempting to understand how the knowledge from the relevant research fields could best lead to an effective solution. Following, an evaluation was designed to allow for feedback from members of the problem domain. Finally, another evaluation was executed in order to look at a second iteration of the artefact. This is described in chapter 5.

Guideline 4: Research Contributions The mentioned artefacts are the main contributions of this project. The main artefact will remain as an a proof of concept service and as a theoretical construct. The Prosopo API will serve as a foundation for new functionality and applications, and as a facilitator for further research within the Prosopo group, while the Android application can be used for further evaluation and demonstration of the main artefact.

Guideline 5: Research Rigour Due to the experimental and flexible approach regarding development and evaluation, the research rigour of this project is somewhat problematic considering the point of relying on well established methods. This has been a challenge, as there are few academic works to compare the artefact and its technology against. However, it has been a focus to use well established theory and techniques from psychology, computer vision and machine learning. A mixed method approach in evaluation has also been utilised, based on rigid theory used in empirical research evaluation. A trade-off between rigour and flexibility has been necessary to explore the problem, which resulted in the development and evaluations which are mentioned in chapters 4 and 5.

Guideline 6: Design as a Search Process The main artefact of this project has continuously undergone evaluation as part of a search process, and is still located somewhere in between a concept and a complete and effective solution to the anonymisation problem. The artefact requires additional iterations of development and evaluation, as the search process needs to continue in order to reach this final solution.

Guideline 7: Communication of Research As part of a master's degree, the project will be presented in the form of a master thesis. This thesis has been written in order to take into consideration both a technological audience, as well as a business audience. As such, the communication of this research should be fathomable for those with a relation to the technology involved, or those affiliated with the problem domain.

7.3.2 Choice of Evaluation

This section will look at the evaluations which were designed to understand the state of the artefact, and for producing data to answer the research questions of the thesis.

The initial part of evaluation, being the development prototype, has been discussed previously. It stands out as the initial search process, combining the technological possibilities with the knowledge of the problem, in order to create a prototype which was at a level where it could be used to create valuable data. There was not much of an evaluation design involved, but rather continuous cycles of experimentation and exploration of the possibilities followed by analysis of the visual outcome it produced.

The later prototype evaluations were designed in a way which focused on uncontrolled in uninfluenced usage. The data collection relied heavily on qualitative feedback based on this usage of the artefact. The decision to use a partner-based evaluation was two-part. Firstly, it was a good way of allowing uninfluenced use, as the informants created the anonymisation cases for each other. Secondly, it was considered a beneficial way of sparking discussion and debate on the matters of the different interview themes, where the informants could come up with statements to be answered or challenged by their partner. The impromptu intercommunication between informants of such a design was also considered to be beneficial.

By combining two different types of evaluation, being the questionnaire form for the anonymisation task and the follow-up interview, this fulfilled the goal of having a multi-method approach. The benefits of mixed-method approaches are described as "... increased validity, more comprehensiveness of findings, more insightful understandings and greater value consciousness and diversity" and "... mixed-method approaches can enhance validity and offer greater certainty when results from two approaches reveal similar conclusions" (C., Benjamin and Goodyear 2001). Ideally, the mixing of methods should perhaps have resulted in two separate evaluations. However, this being a master thesis, where informants are hard to recruit, particularly when requiring a specific set of informants, a pragmatic view was taken. Additionally, the development prototype could also be considered as part of a mixed-

method approach, where the feedback from users, i.e. the informants, was compared to the discoveries of the initial development phase. Ultimately, this helped to better understand what the most apparent challenges were to consider for the continued development of the artefact.

The research questions were used as a basis, when considering the design of the anonymisation tasks and the interview. The definitions of success which had been decided upon, were used as foundation for creating formulations which were in accordance to the relevant success metrics. A pilot study was even employed to make sure that the questions were formulated in a way which was expedient to the intended interpretations. The design was also, in part, created in cooperation with the Prosopo group, which resulted in a more reflected product. Many different designs could have been employed, and further evaluations should be conducted, with the potential for both general and explorative designs, as well as more controlled experiments on anonymisation degree or specific identification sources. However, it should be stated that the evaluations of these first prototypes have been helpful in understanding the artefact in their respective states.

7.4 Discussion on Theory

To put the thesis into a theoretical perspective it is interesting to look back at the research literature review. Three academic backgrounds were highlighted as imperative to understand and solve the facial anonymisation problem. Human face perception was introduced to understand how humans process faces, and how this could be utilised to help maintaining face realism while providing sufficient de-identification. Computer vision and machine learning was introduced as the most promising areas of providing practical means of facilitating for the manipulation, which is done by using face morphing concepts.

7.4.1 Face Perception

Several cognitive phenomenon were highlighted as relevant for facial anonymisation. The most central in this topic is the theory of holistic facial perception, illustrated in the face inversion effect shown in figure 2.1. This thesis further supports this theory when comparing the differences in de-identification performance between the two prototypes. All prototypes, including the development prototype, have been consistent with this theory, which also initially lead to the decision to create the more experimental holistic prototype.

Another phenomenon which was talked about was the cross-race effect. Though this thesis has not addressed or catered for this effect in its research design, the

artefact could definitely investigate this phenomenon further in dedicated evaluation. This implicates that there could be potential for the artefact as a way of studying human face perception phenomena.

The uncanny valley hypothesis was also a potential challenge which was considered during the development. It was a genuine worry that faces would end up looking strange and inconsistent, and thus illicit a negative experience in the artefact's users. Though there were some reports that could easily be characterised as within this eerie and unnatural threshold, the majority found the faces to be natural and even sometimes attractive. The uncanny valley has been a problem for many artificial faces in the entertainment business, and as such will be important to address if the technology is to be used in this domain.

7.4.2 Computer Vision and Machine Learning

Computer vision and machine learning are perhaps some of the most promising academic and business domains in computer technology, with approaches like deep learning implemented in image recognition systems as the pinnacle of the development (He et al. 2016). Cognitive services like Microsoft Cognitive Services and Google Cloud Vision API openly provide ways of utilising computer vision and machine learning functionality to solve both recognition and classification problems which can be implemented in innovative new applications.

These problems have for a long time been considered as very difficult for computers to handle with precision and efficiency, while being trivial for humans to solve. The advances in these fields are very interesting for being able to solve these problems for computers, and this artefact is definitely a part of this movement, which means that the theory and technology involved could be useful for solving similar problems. With the continued advances of these fields, new, more efficient, and more complete solutions will surface. It is not unlikely that the facial anonymisation artefact in the future might take advantage of these advances, allowing for a more complete and reliable solution to be developed.

Many of the current problems and challenges which have been discovered while developing the anonymisation artefact, might very well be solved by using good machine learning algorithms. In the same way machine learning is used to find the landmark coordinates in a face, machine learning could be implemented to train the facial similarity algorithm to better determine what "similarity" really is in this context. These are interesting endeavours which could help create a very solid facial anonymisation artefact given sufficient time and resources for such a task.

7.4.3 Face Morphing

The decision to use face morphing as the manipulation technique was based on the works by Karungaru, Fukumi and Akamatsu (2006), Korshunov and Ebrahimi (2013) and Zanella and Fuentes (2004), and for being a good fit considering the possible challenges connected to holistic face recognition. The largest initial fear was that anonymised faces would be subject to the uncanny valley problem, something which thus far has not shown any prominence based on informant feedback.

There are few documented cases where the face morphing technologies are responsible for poor performance and reliability. The problems have usually been found when looking at the resources and techniques used to facilitate the manipulations. A somewhat problematic database of faces was used, and there is only a limited amount of reliable landmark points which are automatically obtainable by landmark detection algorithms. Face morphing might still be a viable method for facial manipulation in a facial anonymisation context. However, issues may yet arise, which could be difficult for such an approach to handle.

7.5 Future Work

This section will suggest future work which could be of use to solve current limitations and challenges, as well as looking into how the technology could be used in other approaches.

7.5.1 Face Database

In order to deal with the excessive blurring, it is necessary to revamp the current database of the faces used in averaging. Either a new database would have to be constructed using better suited face photos of higher pixel density and size, or another existing database would have to be found to replace the current one. When looking for suitable databases during the development it was found that most face databases were used for facial recognition problems, which do not share the same requirements as a facial anonymisation dataset would require.

7.5.2 Refining the Holistic Approach

To improve the holistic prototype, which seems to be the only real solution to a successful facial anonymisation system, it is required to develop a solution which can locate additional landmarks surrounding the head. This includes hair, hair line, facial hair, ears and the neck of the face. These landmarks will combat both problems of feature identification and image quality reduction sources due to mismatched

facial landmarks. An example of hair detection is a that of Yacoob and L. Davis from 2006 which aimed at detecting and analysing the hair of 524 subjects. Looking into similar works in order to improve the primitive holistic prototype is crucial for a complete and precise solution to the facial anonymisation problem.

7.5.3 Expanding the Scope

In order to reach a complete solution, the current scope has to be expanded. One of the current limitations of the artefact is the inability to detect all faces in an input image. The artefact will not reach a complete state until a fully automatic, undiscriminating detection algorithm is in place. For a facial anonymisation algorithm to be considered dependable it needs to be consistent and complete, and as such it is necessary to deal with currently unhandled situations like side-view faces. A fairly simple example of this is demonstrated by Santemiz, Spreeuwers and Veldhuis (2013) using a three point detection algorithm for side-view face landmark detection intended for use in house safety applications. How to perform manipulation on such faces must also be addressed.

7.5.4 Improving Naive Solutions

A possible explanation to the occasional failure in retaining face realism, is due to the primitive facial similarity evaluation system. A suitable solution to this limitation is to utilise machine learning. A possible implementation could be a supervised artificial neural network, where human face perception input could teach the system to more accurately group similar faces, given their respective facial attributes. A similar approach could be used for creating a more accurate skin colour classification algorithm, as well as other problems which are hard to explicitly describe for a computer.

7.5.5 3D Model Conversion

A somewhat different approach would be to convert the project to a 3D model concept. By converting 2D faces into 3D models, usage could be extended to videos, live capture, computer game modelling, animation and similar use cases where a three-dimensional approach would be required. An example is provided by Jeni, Cohn and Kanade (2015) in which a single 2D image of a person's face is registered as a dense 3D shape in real time across the frames of a 2D video. This could be a very fruitful approach considering some of the ideas for new use cases for the technology, such as those which were provided as answers to Q3.

Chapter 8

Conclusion

The conclusion will attempt to put the thesis into a logical ending, formulating compelling answers to the research questions.

Q1: Is it possible to create an artefact capable of automatically anonymising faces using an average face morphing approach?
Q2: Can the artefact be used to investigate the views on normative practices in manipulation and usage of face photographies in the news media?
Q3: Are there other possible use cases for the artefact or its technology?

Q1: Anonymisation Technology

The artefact in its current state does not comply to the definition of a fully automatic facial anonymisation system. However, there are no findings, at this stage, which implies that a considerable increase in performance is infeasible given further development. At the same time, it is also hard to conclude that it will be possible to reach this stage with a face morphing approach. Further work and additional evaluations will allow for more insight into the problem space and the possibilities for solving the challenges which have been identified.

As such, it is plausible that this artefact could reach a state where the answer to this question would be positive.

Q2: Investigative Tool

Despite that the informants are only a small and peripheral representation of the problem domain, it is claimed that the artefact has a potential to serve as a catalyst for reflection on the implications of its hypothetical implementation in this domain. It would be very interesting to see if this potential is realised with members of the journalistic field.

Based on this, it is claimed that the artefact has the possibility to be relevant in an RRI-perspective, and as such, it is argued that it fulfils the purpose of an investigative tool in its dedicated problem domain.

Q3: Use Cases

Since the artefact originates from a steadily advancing technology field, it was also investigated if there are other use cases which could be interesting in regards of artefact utility. Several scenarios, and possible problem areas, have been presented. There have not been any attempts as to validate the potential in any of these scenarios or problem areas, and further investigation is required to conclude if there really is potential utility for the artefact outside the initial problem space.

Epilogue

The common dominator for the future of the artefact is that it relies on considerable amounts of additional work. This includes the developing of solutions to identified problems and challenges, new explorative evaluations which can further the current knowledge base, and also future investigations into new problem areas to improve the overall utility of the artefact. The road is fairly open in terms of this continuation where several proposed works have been presented.

Appendix A

Evaluation Form

The following pages contain the anonymisation task form. It was produced using Google Forms and exported as a PDF. It is provided in Norwegian, as it was presented to the informants in both evaluations.

Test av Prosopo - Anonymisering av ansikter

Her fyller du ut svarene dine for hver runde av forsøket. Svararket er helt anonymt.

Fyll ut et nytt svarark for hvert bilde.

*Må fylles ut

1. **Hvem tror du personen på bildet er?** *

2. **Hvem er personen på bildet?** *
   Spør partneren din.

3. **Hvilken anonymiseringsgrad ble brukt på bildet?** *
   Spør partneren din.
   Markér bare én oval.
   - ◯ 50%
   - ◯ 60%
   - ◯ 70%
   - ◯ 80%

4. **Klarte du å kjenne igjen personen på bildet?** *
   Markér bare én oval.
   - ◯ Ja *Hopp til spørsmål 5.*
   - ◯ Nei *Hopp til spørsmål 8.*

Svar på disse spørsmålene dersom du kjente igjen personen som ble anonymisert.

5. **Hva gjorde at du kjente igjen personen?** *

6. **Hva synes du om ansiktet på bildet?** *
   Bruk gjerne stikkord. Ser ansiktet naturlig/menneskelig ut?

7. **Hva er eventuelt "feil" med bildet?**

Hopp til spørsmål 11.

Svar på disse spørsmålene dersom du ikke kjente igjen personen som ble anonymisert.

8. **Visste du hvem personen var etter at det ble fortalt?** *
   Markér bare én oval.
   - Ja
   - Nei

9. **Hva synes du om ansiktet på bildet?** *
   Bruk gjerne stikkord. Ser ansiktet naturlig/mennesklig ut?

10. **Hva er eventuelt "feil" med bildet?**

Hopp til spørsmål 11.

Sammenlign bildene før og etter anonymisering.

11. **I hvor stor grad ligner fjesene på hverandre?** *
    Markér bare én oval.

	1	2	3	4	5	
Veldig ulike						Veldig like

Appendix B

Evaluation Interview Guide

This following page contains the interview guide for the first evaluation. The first topic was also used in the second evaluation. It is provided in Norwegian.

Prosopo 1 - Spørsmål til intervju

Tema: Reaksjon

Hva er din umiddelbare reaksjon etter å ha brukt applikasjonen/teknologien?

- Hva er det som skjer?

Hvordan vurderer du denne teknologien som en ansiktsanonymiseringsteknologi?

- Forbedringer?
- Hva fungerer?

Tema: Sosiale medier

Ser du noen bruksområder for denne teknologien i sosiale medier?

Kan teknologien sammenlignes med lignende tjenester?

- Snapchat?

Ser du noen problemer med bruk av denne teknologien i sosiale medier?

Tema: Nyhetsmedia

Ser du noen bruksområder for denne teknologien i nyhetsmedia?

Hvordan vil du sammenligne denne teknologien med andre, vanlige anonymiseringsteknikker (som sladding, blurring, pikselering ol.)?

- https://www.facepixelizer.com/

Ser du noen problemer med bruk av denne teknologien i nyhetsmedia?

Hvordan ser du på denne anonymiseringen i forhold til andre former for ikke-komputasjonelle bildemanipuleringer i journalistikk?

- Etisk vurdering
- Følelsen av hva det er som er virkelig?

Er du enig i utsagnet at et pressefotografi er en realistisk framstilling av verden?

Tema: Andre bruksområder

Ser du andre bruksområder for teknologien?

- Video/dokumentar/fiksjonsfilm?
- Underholdning/kommersialisering?

Ser du andre problemer med teknologien?

Appendix C

Code Excerpts

The following code listings are excerpts from the facial anonymisation artefact. These selections have been chosen to demonstrate some of the solutions which were employed during development.

```
def face_detect_fpp(image_path):
  """ Returns all faces detected by F++ as a dictionary. """

  des_attributes = "&return_landmark=1&return_attributes=age,gender,
    smiling,headpose,facequality,blur,eyestatus"

  r = requests.post(fpp_detect_url.format(k=fpp_api_key, s=
    fpp_api_secret) + des_attributes, files={"image_file": open(
    image_path, "rb")})

  # Load the result into a dictionary
  faces = simplejson.loads(r.text)

  return faces
```

Listing C.1: F++ face detection request

```
  # Corners of the eyes
  eye_corner_src = [faces[i].get_left_eye_corner(), faces[i].
    get_right_eye_corner()]

  # Compute similarity transform
  transform = similarity_transform(eye_corner_src, eye_corner_dst)

  # Apply similarity transformation
  img = cv2.warpAffine(images[i], transform, (width, height))

```

```
# Apply similarity transform on points
points2 = np.reshape(np.array(points1), (n, 1, 2))

points = cv2.transform(points2, transform)

points = np.float32(np.reshape(points, (n, 2)))
```

Listing C.2: Similarity transform

```
def get_delaunay_triangles(rect, points):
  """ Returns the delaunay triangles of a face given its landmarks.
  """
  # Create sub-div
  sub_div = cv2.Subdiv2D(rect)

  # Insert points into sub-div
  for p in points:
  sub_div.insert((p[0], p[1]))

  # List of triangles where each triangle is a list of 3 points
  triangle_list = sub_div.getTriangleList()

  # Find the indices of triangles in the points array
  delaunay_tri = []

    for t in triangle_list:
    pt = [(t[0], t[1]), (t[2], t[3]), (t[4], t[5])]

    pt1 = (t[0], t[1])
    pt2 = (t[2], t[3])
    pt3 = (t[4], t[5])

    if rect_contains(rect, pt1) and rect_contains(rect, pt2) and
  rect_contains(rect, pt3):
      ind = []
      for j in xrange(0, 3):
        for k in xrange(0, len(points)):
          if abs(pt[j][0] - points[k][0]) < 1.0 and abs(pt[j][1] -
  points[k][1]) < 1.0:
            ind.append(k)
      if len(ind) is 3:
        delaunay_tri.append((ind[0], ind[1], ind[2]))

  return delaunay_tri
```

Listing C.3: Delaunay triangulation

```
def morph_triangle(image1, image2, image_out, t1, t2, t, alpha):
  """ Warps and alpha blends the found delaunay triangles from image 1
    and image 2 to image. """

  # Find bounding rectangle for each triangle
  r1 = cv2.boundingRect(np.float32([t1]))
  r2 = cv2.boundingRect(np.float32([t2]))
  r = cv2.boundingRect(np.float32([t]))

  # Offset points by left top corner of the respective rectangles
  t1_rect = []
  t2_rect = []
  t_rect = []

  for i in xrange(0, 3):
    t_rect.append(((t[i][0] - r[0]), (t[i][1] - r[1])))
    t1_rect.append(((t1[i][0] - r1[0]), (t1[i][1] - r1[1])))
    t2_rect.append(((t2[i][0] - r2[0]), (t2[i][1] - r2[1])))

  # Get mask by filling triangle
  mask = np.zeros((r[3], r[2], 3), dtype=np.float32)
  cv2.fillConvexPoly(mask, np.int32(t_rect), (1.0, 1.0, 1.0), 16, 0)

  # Apply warpImage to small rectangular patches
  img1_rect = image1[r1[1]:r1[1] + r1[3], r1[0]:r1[0] + r1[2]]
  img2_rect = image2[r2[1]:r2[1] + r2[3], r2[0]:r2[0] + r2[2]]

  size = (r[2], r[3])
  warp_image1 = math_util.apply_affine_transform(img1_rect, t1_rect,
    t_rect, size)
  warp_image2 = math_util.apply_affine_transform(img2_rect, t2_rect,
    t_rect, size)

  # Alpha blend rectangular patches
  img_rect = (1.0 - alpha) * warp_image1 + alpha * warp_image2

  # Copy triangular region of the rectangular patch to the output image
  image_out[r[1]:r[1] + r[3], r[0]:r[0] + r[2]] = (image_out[r[1]:r[1]
    + r[3], r[0]:r[0] + r[2]] * (1 - mask) + img_rect * mask)
```

Listing C.4: Triangle morphing

```
  # Clone seamlessly
  output = cv2.seamlessClone(np.uint8(input_warped), image2, mask,
    center, cv2.NORMAL_CLONE)
```

Listing C.5: Seamless cloning

```
def __compare_face_similarity(face1, face2):
  """ Uses a predetermined set of weights to calculate the level of
    similarity between two faces given their attributes. The lower the
    value, the more similar, where 0 is identical. """

  skin_colour_w = 0.05
  age_w = 0.5
  smiling_w = 0.06
  moustache_w = 0.5
  beard_w = 0.5
  sideburns_w = 0.5
  pitch_w = 0.35
  roll_w = 0.9
  yaw_w = 0.9

  skin_colour = (__compare_skin_colour_similarity(face1.skin_colour,
    face2.skin_colour, False)) * skin_colour_w
  age = (face1.age - face2.age) * age_w
  smiling = (face1.smiling - face2.smiling) * smiling_w
  moustache = (face1.moustache - face2.moustache) * moustache_w
  beard = (face1.beard - face2.beard) * beard_w
  sideburns = (face1.sideburns - face2.sideburns) * sideburns_w
  pitch = (face1.pitch - face2.pitch) * pitch_w
  roll = (face1.roll - face2.roll) * roll_w
  yaw = (face1.yaw - face2.yaw) * yaw_w

  similarity = (abs(skin_colour) + abs(age) + abs(smiling) + abs(
    moustache) +
  abs(beard) + abs(sideburns) + abs(pitch) + abs(roll) + abs(yaw))

  return similarity
```

Listing C.6: Facial similarity calculation

```
def calculate_skin_mean(face):
  """ Finds the mean skin RGB-value of a face given its landmarks. """

  image = face.image

  # Create the mask
  mask = np.zeros(image.shape[:2], np.uint8)

  # Find the distinct landmarks that make up the relevant facial
    features
  border_landmarks = face.read_landmark_coordinates()
  left_eyebrow_landmarks = face.get_left_eyebrow()
  left_eye_landmarks = face.get_left_eye()
```

```
13  right_eyebrow_landmarks = face.get_right_eyebrow()
14  right_eye_landmarks = face.get_right_eye()
15  mouth_landmarks = face.get_mouth()
16
17  # Find the convex hull indices for each feature
18  border_hull_indices = cv2.convexHull(np.array(border_landmarks),
      returnPoints=False)
19  ...
20
21  # Create lists for convex landmarks for every feature
22  border_poly = []
23
24  for hull_index in border_hull_indices:
25    index = hull_index[0]
26    border_poly.append(border_landmarks[index])
27  ...
28
29  # Fill the mask and remove disturbing features
30  cv2.fillConvexPoly(mask, np.int32(border_poly), (255, 255, 255))
31  cv2.fillConvexPoly(mask, np.int32(left_eyebrow_poly), (0, 0, 0))
32  cv2.fillConvexPoly(mask, np.int32(left_eye_poly), (0, 0, 0))
33  cv2.fillConvexPoly(mask, np.int32(right_eyebrow_poly), (0, 0, 0))
34  cv2.fillConvexPoly(mask, np.int32(right_eye_poly), (0, 0, 0))
35  cv2.fillConvexPoly(mask, np.int32(mouth_poly), (0, 0, 0))
36
37  # Calculate the mean from the masked image
38  return cv2.mean(image, mask)
```

Listing C.7: Skin colour calculation

Appendix D

Anonymisation Identities

The identities in figure 4.4 are from left to right: Marilyn Monroe, Ashton Kutcher and Angelina Jolie.

The identities in figure 4.5 are from left to right: Tom Cruise, Emma Watson and Usain Bolt.

Were you able to recognise any of them?

References

A., Rosebrock (2017). *PyImageSearch About*. URL: http://www.pyimagesearch.com/about/ (visited on 31/03/2017).

Bainbridge, W. A. (2012). *10k US Adult Faces Database*. URL: http://wilmabainbridge.com/facememorability2.html (visited on 22/03/2017).

Bainbridge, W. A., P. Isola and A. Oliva (2013). 'The intrinsic memorability of face images'. In: *Journal of Experimental Psychology: General. Journal of Experimental Psychology: General* 142.4, pp. 1323–1334.

Benson, P. and D. Perrett (1991). 'Computer averaging and manipulations of faces'. In: *Photovideo: Photography in the age of the computer*, pp. 32–38.

Bradski, G. (2000). 'The OpenCV Library'. In: *Dr. Dobb's Journal of Software Tools*.

Bradski, G. and A. Kaehler (2008). *Learning OpenCV*.

C., Greene J., L. Benjamin and L. Goodyear (2001). 'The Merits of Mixing Methods in Evaluation'. In: *Evaluation* 7.1, pp. 25–44.

Caracelli, V. J. and J. C. Greene (1997). 'Crafting Mixed-Method Evaluation Designs'. In: *New Directions for Evaluation* 1997.74, pp. 19–32. (Visited on 01/05/2017).

Davis, D. (2017). *Intel Acquires Computer Vision for IOT, Automotive*. URL: https://newsroom.intel.com/editorials/intel-acquires-computer-vision-for-iot-automotive/ (visited on 22/03/2017).

Driessen, B. and M. Dürmuth (2013). 'Achieving Anonymity against Major Face Recognition Algorithms'. In: *Communications and Multimedia Security Lecture Notes in Computer Science*, pp. 18–33.

Face++ Cognitive Services (2017). *Everything You Need on Facial Recognition*. URL: https://www.faceplusplus.com/ (visited on 29/03/2017).

Face Recognition with OpenCV (2016). URL: http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html (visited on 08/03/2016).

Feusner, J. D. et al. (2007). 'Visual Information Processing of Faces in Body Dysmorphic Disorder'. In: *Archives of General Psychiatry* 64.12, pp. 1417–1426.

Fielding, R. T. (2000). 'Architectural Styles and the Design of Network-based Software Architectures'. Ph.D. University of California, Irvine. Chap. 5.

Frischholz, R. (2017). *The Face Recognition Homepage Databases*. URL: `https://facedetection.com/` (visited on 22/03/2017).

Galton, F. (1878). 'Composite portraits'. In: *Journal of the Anthropological Institute of Great Britain and Ireland* 8, pp. 132–142.

Goffaux, V. and B. Rossion (2006). 'Faces are "spatial"–holistic face perception is supported by low spatial frequencies'. In: *Journal of Experimental Psychology: Human Perception and Performance* 32.4, pp. 1023–1039.

Google (2017). *Android Studio*. URL: `https://developer.android.com/studio/features.html` (visited on 22/03/2017).

Grgic, M. and D. Kresimir (2017). *The Face Detection Homepage*. URL: `http://www.face-rec.org/databases/` (visited on 22/03/2017).

Halberstadt, J. and G. Rhodes (2000). 'The Attractiveness of Nonface Averages: Implications for an Evolutionary Explanation of the Attractiveness of Average Faces'. In: *Psychological Science* 11.4, pp. 285–289.

— (2003). 'It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive'. In: *Psychonomic Bulletin & Review* 10.1, pp. 149–156.

He, K. et al. (2016). 'Deep Residual Learning for Image Recognition'. In: *The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

Hevner, A. R., S. T. March and J. Park (2004). 'Design Science in Information Systems Research'. In: *MIS Quarterly* 28.1, pp. 75–105.

Jeni, L. A., J. F. Cohn and T. Kanade (2015). 'Dense 3D Face Alignment from 2D Videos in Real-Time'. In: *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*.

JetBrains (2017). *PyCharm*. URL: `https://www.jetbrains.com/pycharm/` (visited on 22/03/2017).

Karungaru, S., M. Fukumi and N. Akamatsu (2006). 'Automatic Human Faces Morphing Using Genetic Algorithms Based Control Points Selection'. In: *International Journal of Innovative Computing, Information and Control* 3.2, pp. 247–256.

King, Davis E. (2009). 'Dlib-ml: A Machine Learning Toolkit'. In: *Journal of Machine Learning Research* 10, pp. 1755–1758.

Koeslag, J. H. (1990). 'Koinophilia groups sexual creatures into species, promotes stasis, and stabilizes social behaviour'. In: *Journal of Theoretical Biology* 144.1, pp. 15–35.

Korshunov, P. and T. Ebrahimi (2013). 'Using face morphing to protect privacy'. In: *10th IEEE International Conference on Advanced Video and Signal Based Surveillance*.

Leff, A. and J. T. Rayfield (2001). 'Web-Application Development Using the Model/View/Controller Design Pattern.' In: *IEEE Enterprise Distributed Object Computing Conference*, pp. 118–127.

Li, S. Z. and A. K. Jain (2011). *Handbook of Face Recognition*. London: Springer London.

Little, A. C., B. C. Jones and L. M. DeBruine (2011). 'The many faces of research on face perception'. In: *Philosophical Transactions of the Royal Society B: Biological Sciences*, pp. 1634–1637.

Mallick, S. (2016). *Average Face: OpenCV (C++ / Python) Tutorial*. URL: `http://www.learnopencv.com/average-face-opencv-c-python-tutorial/` (visited on 23/03/2017).

— (2017). *Learn OpenCV About*. URL: `https://www.learnopencv.com/about/` (visited on 31/03/2017).

Matlin, M. W. (2014). *Cognitive Psychology*.

Microsoft Corporation (2016). *Face API*. URL: `https://www.microsoft.com/cognitive-services/en-us/face-api` (visited on 29/03/2017).

— (2017). *Visual Studio*. URL: `https://www.visualstudio.com/` (visited on 22/03/2017).

Newton, E., L. Sweeney and B. Malin (2005). 'Preserving privacy by deidentifying face images'. In: *IEEE Trans. Knowl. Data Eng. IEEE Transactions on Knowledge and Data Engineering* 17.2, pp. 232–243.

NPPA (2017). *Code of Ethics*. URL: `https://nppa.org/code_of_ethics` (visited on 16/04/2017).

O'Toole, A. J. (2011). 'Face Recognition by Humans and Machines'. In: *Handbook of Face Recognition*, pp. 597–614.

Pérez, P., M. Gangnet and A. Blake (2003). 'Poisson Image Editing'. In: *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2003* 22.3, pp. 313–318.

Peterson, S. (2017). *Computing Constrained Delaunay Triangulations*. URL: `http://www.geom.uiuc.edu/~samuelp/del_project.html` (visited on 26/03/2017).

Russell, S. and P. Norvig (2014). *Artificial Intelligence: A Modern Approach*. 3rd ed.

Santemiz, P., L.J. Spreeuwers and R.N.J. Veldhuis (2013). 'Automatic landmark detection and face recognition for side-view face images'. In: *2013 International Conference of the BIOSIG Special Interest Group*.

Stack Exchange Inc. (2017). *Stack Overflow*. URL: `http://stackoverflow.com/company/about` (visited on 31/03/2017).

Surynek, P. and I. Lukšová (2011). 'Automated Classification of Bitmap Images using Decision Trees'. In: *Polibits* 44, pp. 11–18.

Takahashi, Y. et al. (2006). 'Feature Point Extraction in Face Image by Neural Network'. In: *2006 SICE-ICASE International Joint Conference*.

Tanaka, J. W., M. Kiefer and C. M. Bukach (2004). 'A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study.' In: *Cognition* 93.1.

The EU Framework Programme for Research and Innovation (2017a). *Responsible research and innovation*. URL: `https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation` (visited on 01/05/2017).

— (2017b). *What is Horizon 2020?* URL: `https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020` (visited on 01/05/2015).

Tinwell, A. (2014). *The Uncanny Valley in Games and Animation*.

ViSmedia (2017). *About | ViSmedia*. URL: `http://vismedia.org/about/` (visited on 21/03/2017).

Weinstein, E. W. (2017). *Affine Transform*. URL: `http://mathworld.wolfram.com/AffineTransformation.html` (visited on 26/03/2017).

Yacoob, Y. and L. Davis (2006). 'Detection and analysis og hair'. In: *EEE Transactions on Pattern Analysis and Machine Intelligence* 28.7, pp. 1164–1169.

Zanella, V. and O. Fuentes (2004). 'An Approach to Automatic Morphing of Face Images in Frontal View'. In: *MICAI 2004: Advances in Artificial Intelligence Lecture Notes in Computer Science*, pp. 679–687.

Zhang, Z. et al. (2014). 'Facial Landmark Detection by Deep Multi-task Learning'. In: *Computer Vision – ECCV 2014 Lecture Notes in Computer Science*, pp. 94–108.