Realistic face manipulation by morphing with average faces

Joar Midtun, Bjørnar Tessem, Simen Karlsen, Lars Nyre
Department of Information Science and Media Studies, University of Bergen, Norway

Abstract

Face manipulation has become a standard feature of many social media services. Most of these applications use the feature for entertainment purposes. However, such manipulation techniques could also have potential in a journalistic setting. For instance, one could create realistic, anonymized faces, as an aesthetic alternative to the coarse techniques of blurring or pixelation normally used today. In this paper, we describe how we can use algorithms for face manipulation from computer vision to anonymize faces in journalism. The technique described uses morphing with average faces from a selection of faces that is similar to the original face, and alters the faces in the original pictures into realistic-looking face manipulations. However, it struggles with sufficient anonymization due to identifiable non-facial features of persons in an image.

Keywords: computer vision, face manipulation, face morphing, average faces, anonymization

Introduction

With the advent of social media services like Snapchat, face manipulation has become a readily available tool for any smartphone user. This feature is mostly used for entertainment; to make funny faces, although it has also been used to anonymize people in interview situations on sensitive topics [1]. This news story shows that face manipulation in images may have more serious purposes, for instance anonymization of persons covered in crime journalism. We ask if it is possible to make anonymization tools more optimized for this task than those provided by for example Snapchat. Is it possible to create or alter faces that are realistic, but still unrecognizable for the viewer?

In this paper, we present an approach to anonymizing faces in images using known algorithms for face manipulation, like face detection, face landmark detection, face averaging, face morphing and face swapping. The goal of the approach presented here is to be able to create an artificial face similar to a true face represented in a photo. The new image maintains its life-like, realistic look, but has been subtly altered and is sufficiently different for the true person not to be recognizable.

We start by presenting the techniques that are used in face recognition and face manipulation. We then describe the approach chosen, and present some examples which display how our technique performs. We present the technological tools used, before we present some user responses to the technology. In the discussion we focus on technological weaknesses and potential, before we conclude.

Face manipulation techniques in images

Face recognition

According to Dr. Robert Frischholz' face detection and recognition information website [2] the main problem of face recognition is in fact to detect if there is a face in the image. As soon as the faces are located, the features we need to identify people are readily available. Li and Jain further define face recognition as a pattern recognition

This paper was presented at the NIK 2017 conference. For more information see http://www.nik.no/

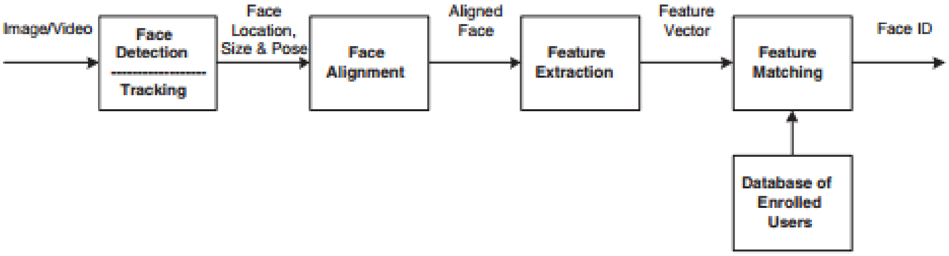


Figure 1: Face recognition process flow. From [3]

problem [3]. A face can be represented as a three-dimensional object characterized by varying attributes, e.g. illumination, pose, expression, and others. The task of face recognition therefore involves uncovering these attributes, and matching them with those of previously known faces.

A face recognition system thus uses a four-step approach to a recognition process, as can be seen in Figure 1. These steps, or modules, are: detection, alignment, feature extraction and matching. Face detection and alignment can be considered as pre-processing requirements before recognition can take place, where recognition consists of feature extraction and matching.

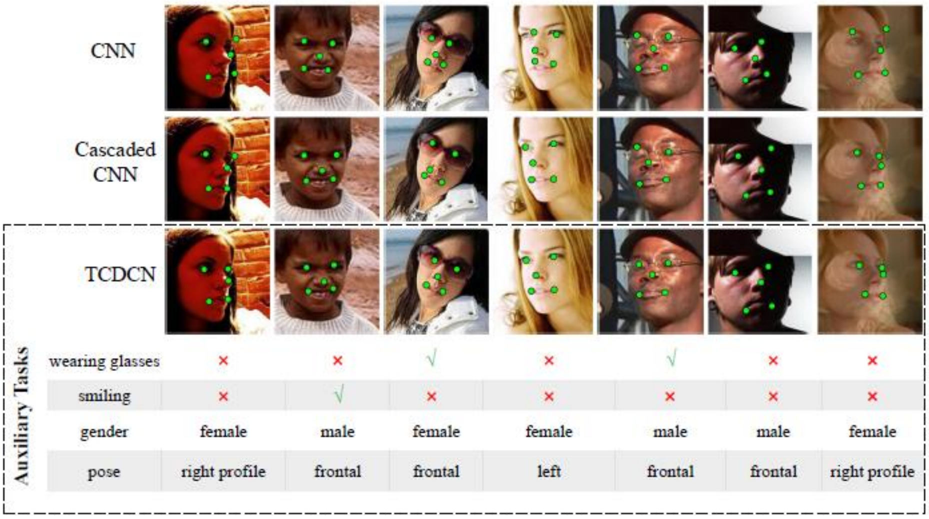
Face detection is responsible for segmenting the face, or more specifically a face area, from the background. Face alignment is performed in order to more precisely pinpoint face location, and also to normalise the faces as data for the next stages. This is done by performing morphing or geometrical transforms on the different features of the face, as well as normalising with regards to photometrical properties, such as illumination and grey scale. As such, detection and alignment work in tandem to provide estimates of the location and scale of faces detected in the input data.

After having been geometrically and photometrically normalised, the face object is ready to undergo feature extraction. In the case of face recognition, the interesting features to extract are those that are useful and consistent for differentiating between faces, in regards to the geometrical and photometrical variation. The final module of matching involves comparing the extracted feature vector against some applicable database of similarly processed faces. This will finally 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.

Facial landmark detection

Facial landmark detection has already been mentioned as an integral part of the process in recognising faces. For a computer, the facial landmark points make up the identifying points of a face, allowing for the use of faces as data. Zhang et al. describe the importance of, and challenges involved in, facial landmark detection [4]. Even though considerable amounts of work has been performed 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.



wearing glasses	×	×	√	×	√	×	×
smiling	×	√	×	×	×	×	×
gender	female	male	female	female	male	male	female
pose	right profile	frontal	frontal	left	frontal	frontal	right profile

Figure 2. Landmark detection by three different algorithms; traditional CNN, Cascaded CNN, and TCDCN. From [4].

Historically, there has been two main categories of detection methods: Regression-based and template fitting. 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, more recent approach is to use cascaded Convolutional Neural Networks (CNNs) [5]. 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.

Traditionally, landmark detection has been treated as an isolated and independent problem, something Zhang et al. claim to be a shortcoming [4]. 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."* Zhang et al. name their approach a Tasks-Constrained Deep Convolutional Network (TCDCN). Figure 2 show how various algorithms compare according to Zhang et al. [4].

Average faces

The average face is a concept that has been of interest in several disciplines. It has been subject to much debate in psychology, where several studies have shown that computationally averaged faces are generally regarded as more aesthetically pleasing [6]. This phenomenon is often credited to the fact that through averaging, individual imperfections and asymmetry are watered down. *Koinophilia* is an evolutionary hypothesis, postulating that an average looking specimen is more often preferred as a mate as it is less likely to have undesirable mutations [7]. 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 [8, 9]. The composite technique had a resurgence in the 1990s when computers could take over these operations, and it is now often referred to as face averaging [10].

The concept of computationally averaging a set of faces is fairly similar to that 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). 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, or some other points of reference, are defined, the original image is warped, and the landmarks are transformed using a similarity transform [10]. 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, and aligning the rest of the facial features is also required. This is done by calculating the mean landmark coordinates of each reference frame and then calculating a Delaunay Triangulation [11]. 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 3. Given this triangulation it is possible to warp the face triangles to match the mean average face landmark points using an affine transform [12]. Given a source plane, and a destination plane, this transformation preserves collinearity, which means that all points lying on a single line in the source plane still lies on a single 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 [12]. 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 averaged and added onto an output image [10].

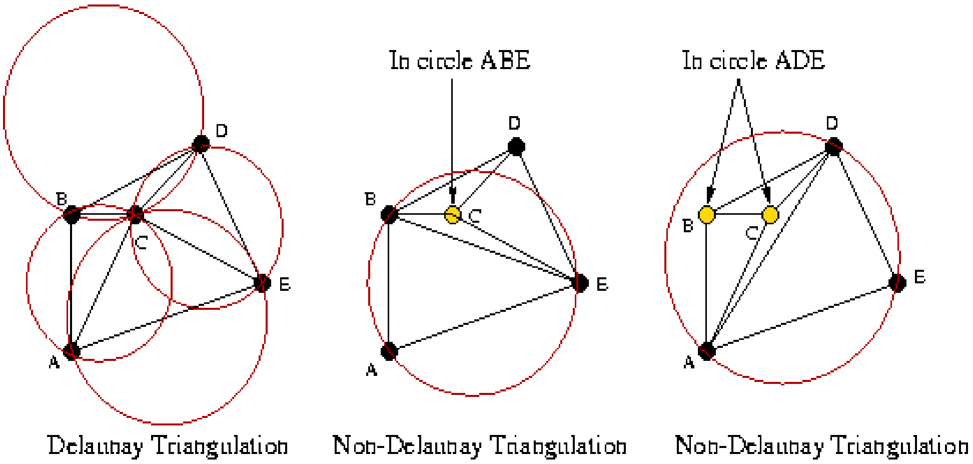


Figure 3. Triangulations for A through E. From [13]

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 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 [10].

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 ‘Poisson Image Editing’ idea of Pérez et al. [14]. 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 [10].

Anonymization with the use of average faces

The anonymization process we have implemented 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 all the faces are anonymised. Each sub-process has its own series of steps, which will be described further.

Face analysis

The first step of analysis is detecting faces in the input image. For this we have used free ‘cognitive’ services from commercial providers Microsoft Cognitive Services [15] and Face++ [16]. From these services we find all detectable faces. For matching, we order them by the sum of the landmark coordinates of the leftmost corner of the left eye. This way, the ordering of faces will be the same for faces found by either 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 anonymization. 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 1. The ’X’ means that the cognitive service provides a value for this attribute, while blank space means that they do not. The final column shows which service was chosen for this project.

Table 1. Cognitive services attributes provided and chosen provider for our application

Attribute	Face++	Microsoft	Chosen
Age	X	X	Microsoft
Gender	X	X	Microsoft

Smiling	X	X	Face++
Glasses		X	Not used
Right eye	X		Face++
Left eye	X		Face++
Moustache		X	Microsoft
Beard		X	Microsoft
Sideburns		X	Microsoft
Pitch	X		Face++
Roll	X	X	Face++
Yaw	X	X	Face++
Landmarks count	83	27	Face++
Face rectangle	X	X	Face++
Face quality	X		Not used
Blurriness	X		Not used
Motion blur	X		Not used
Gaussian blur	X		Not used

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 it 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 currently not utilised in any way.

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. To exemplify, Figure 4 shows three example faces, and Table 2 shows the facial attributes discovered for these faces.

Table 2. Cognitive services attributes detected for the faces in Figure 4.

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
Sidburns	0.0	0.0	0.4
Pitch	-0.37	2.06	3.59
Roll	4.25	1.33	-0.74
Yaw	-4.32	-2.79	1.27

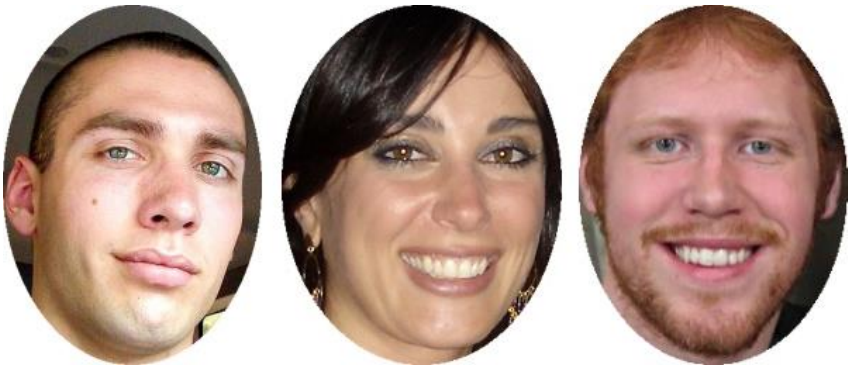


Figure 4. Faces analyzed for face attributes in Table 2

Manipulation

After the analysis, we are left with a face object for each detected face, ready for anonymization. Each face is anonymised separately. Prior to any manipulation, each face is first cropped from the original image using the detected face rectangle (which is in fact always a square). This means that all images, which are represented as two-dimensional matrices where each cell contains a pixel/RGB-tuple, will be of equal size. An input parameter, α, is also provided. This α blending parameter could be described as the degree of anonymization, 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 α. 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 5. A description of the important steps follow:

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 difference between 0 and 1, where 0 is a face with identical attributes. This calculation is not trained or dynamic in any way, but manually weighted through trial and evaluation (see section on face averaging to see how the manipulation is done computationally). The output then, is 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 anonymization stage, as it is here the retained percentage of the original face and the average face is established. The output is a face with a mixture of α percent average face, and $100 - \alpha$ 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 manipulation. The morphed face is simply swapped 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 α percent.

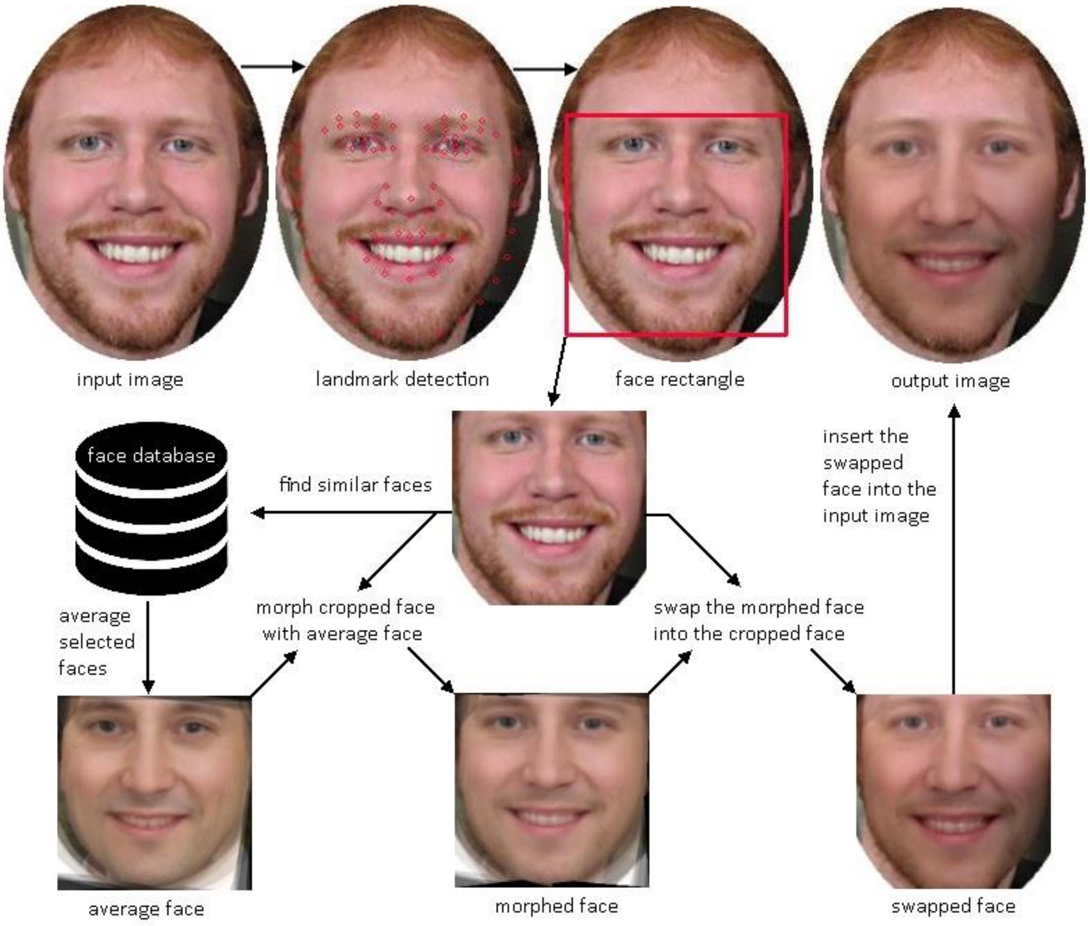


Figure 5. The anonymization process

Feature anonymization

The feature anonymization prototype was the first prototype which was created. It was designed to provide anonymization of the area *within* the facial landmarks, which contains the identifying facial features, hence the name. The prototype was compared to a holistic anonymization prototype described below. Both prototypes are presented according to variation in the concepts, and their strengths and weaknesses.

Concept

The concept for the first prototype is to limit the manipulated area to within the 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 6 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 keeping the original image as unaffected as possible.

Weaknesses

It is fair to suspect that the output might not reach a sufficient level of de-identification to suffice as an anonymization technique. When ignoring large areas of the face (e.g. hair, facial hair, forehead and neck), it is not unreasonable to assume that there is a lot of unaltered visual data which can provide foundation for identification. Considering that there is good evidence for face recognition being a holistic process [17], this is in all likelihood a significant shortcoming.

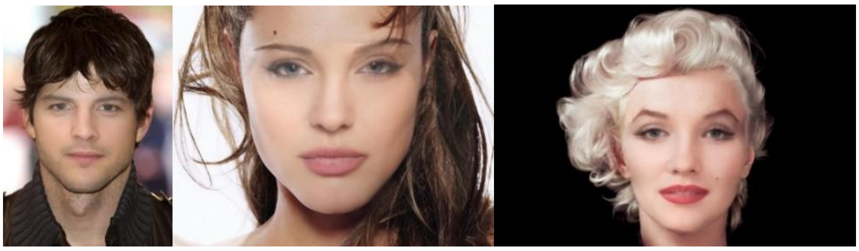


Figure 6. Example anonymizations from the feature anonymization prototype

Holistic anonymization

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

Concept

This second holistic 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 7 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 [18]. In this case, it should provide a higher likelihood of success in de-identification/anonymization.

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.

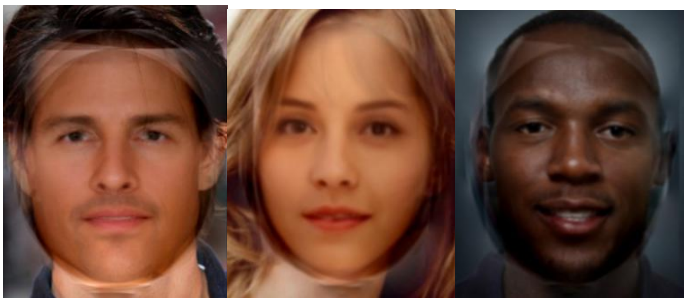


Figure 7. Three faces anonymized with holistic anonymization

Implementation

To test the approach we implemented the facial anonymization in an Android mobile application (named Prosopo, meaning ‘face’ in Greek). The app makes it possible to take a picture of a person or search the web for pictures of persons, and then run the anonymization process on these pictures through a web service.

The service is implemented as a RESTful service and is coded in Python. It receives the picture and runs the anonymization process by using several free services, as well as algorithms implemented in OpenCV, an open source collection of software for computer vision [19]. The first two steps of the algorithm, landmark detection and extraction of the face rectangle is handled by the use of OpenCV. As mentioned, Microsoft Cognitive Services [15] and Face++ [16] services are utilized to identify facial features of the input face.

The 5 faces used for averaging is selected from a freely available database of face images. The database is the 10K US Adult Faces Database [20] where more than 10.000 faces have been collected for use in psychology, cognitive science and computer vision. The selection of similar faces from the database is based on features obtained from the Face++ and Microsoft services, and then linearly weighted according to manually chosen parameters.

The further sub-processes of averaging of the 5 database faces, morphing with the original face and face swapping is handled by the OpenCV software.

User responses

The focus of this article has been on the solution for the anonymization process, and not on the more usefulness issues. For instance, we still don't know whether it would be ethically sensible for a news medium to implement Prosopo in their crime journalism. Here we summarize the findings from evaluation of the tool, in two separate groups consisting of new media students at the bachelor level and three seasoned photographers.

Concerning the quality of anonymization, the new images can be assessed according to two dimensions: face realism/image quality, and ability to anonymize the face. Our respondents indicate that the anonymized faces are natural, perhaps with some reduced image quality. Examples of quality issues were blurring, skin colour, and visible manipulation boundaries. Ability to anonymize gave more unclear results. The respondents seem to be able in most cases to recognize the anonymized persons when they are well-known actors or politicians. This may be considered less encouraging as for the usefulness of the technology in journalism. On the other hand, the use of celebrities may be questionable as for measuring ability to recognize a person from a manipulated picture of the person.

The journalist respondents were mostly critical of using this technology in journalism. The journalistic ideal is that a photography should present an unedited version of reality, and if the photo has indeed been manipulated, this should be labelled and tagged in the photo itself. The professionals were determined in their views about this requirement, and would prefer the use of existing techniques for anonymization. The media students, however, were more inclined to imagine documentary or otherwise realistic genres where such an approach to anonymization could be useful.

The respondents were also asked to reflect on use of the technology outside of journalism, and they had ideas about entertainment apps like “Guess who”-quizzes or realistic avatar faces for computer games. Additionally, applications in semi-blind dating applications, previewing of cosmetic surgery, and live camera captures, indicate that there are possibilities for this technology beyond professional journalism, as well as the social media niche it occupies today.

Conclusion

In this article, we have described an approach to manipulation of faces with the aim of anonymization of persons involved in crime journalism stories. The technique itself is promising and results in faces that have a large degree of realism, while at the same time becoming more or less unrecognizable. The average face computation and consecutive morphing itself creates realistic faces, and could most likely be improved with larger face databases. We acknowledge that in some cases the person might not be sufficiently anonymized. There is still a problem about hair, clothes and other non-facial attributes that are often enough for readers to identify a person in an image. This poses challenges not yet solved in computer vision and image manipulation research.

References

1. Omar,Y., (2016). Using Snapchat to give a voice to sexual abuse survivors. *The Guardian*. https://www.theguardian.com/media-network/2016/aug/24/snapchat-give-voice-sexual-abuse-survivors
2. Frischholz, R., (2017), The Face Recognition Homepage Databases. https://facedetection.com/
3. Li, S.Z. and A.K. Jain (2011). Handbook of Face Recognition. Springer, London.
4. Zhang Z., Luo P., Loy C.C., Tang X. (2014) Facial Landmark Detection by Deep Multi-task Learning. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham.
5. Y. Sun, X. Wang and X. Tang (2013), "Deep Convolutional Network Cascade for Facial Point Detection," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 3476-3483.
6. Halberstadt, J. and G Rhodes (2000). The Attractiveness of Nonface Averages: Implications for an Evolutionary Explanation of the Attractiveness of Average Faces. Psychological Science Vol 11, Issue 4, pp. 285 – 289
7. Koeslag, J.H. (1990), Koinophilia groups sexual creatures into species, promotes stasis, and stabilizes social behaviour, Journal of Theoretical Biology, Volume 144, Issue 1, 1990, Pages 15-35.
8. Benson P. J. and D.I. Perrett D I, (1991) Computer averaging and manipulation of faces” in Wombell P. (ed) Photovideo. Photography in the Age of the Computer (London: Rivers Oram Press), pp. 32-38
9. Galton, F. (1878). Composite portraits. Journal of the Anthorpological Institute og Great Britain and Ireland, vol.8, pp. 132-144.

10. Mallick, S (2016) Average Face: OpenCV (C++/Python) Tutorial. http://www.learnopencv.com/average-face-opencv-c-python-tutorial/
11. Delaunay, B (1934). *"Sur la sphère vide"*. Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles. **6**: 793–800.
12. Weinstein, E.W. (2017). Affine transform. http://mathworld.wolfram.com/AffineTransformation.html
13. Peterson, S. (2017) Computing Constrained Delaunay Triangulations. http//www.geom.uiuc.edu/~samuelp/del_project.html
14. Pérez P., M. Gangnet., and A. Blake. (2003). Poisson image editing. *ACM Trans. Graph.* 22, 3 (July 2003), 313-318.
15. Microsoft Cognitive Services Face API. https://azure.microsoft.com/en-us/services/cognitive-services/face/
16. Face++. https://www.faceplusplus.com/
17. Goffaux, V. and B. Rossion (2006). Faces are “spatial”- holistic face perception is supported by low spatial frequencies. Journal of Experimental Psychology: Human Perception and Performance, 32(4), pp. 1023-1039
18. Matlin, M.V. (2013) Cognitive Psychology. Wiley.
19. OpenCv (2017). http://opencv.org/
20. Bainbridge, W.A. 10k US Adult Faces Database. http://wilmabainbridge.com/facememorability2.html