Morphing robust face recognition

KRISZTIÁN SZABÓ, University of Twente, The Netherlands

The increased accessibility of face morphing software poses a threat to face recognition systems due to their vulnerability to face morphing attacks. An attacker can merge a picture of themselves with another similar but distinct-looking person, and current face recognition systems would consider the picture to match both persons. After creating an ID with such picture, the ID would no longer be considered unique. This research aims to combat such attacks by creating a shape-free representation for face recognition and measuring its accuracy against other open-source face recognition software. Using the PUT face database, 50 peoples faces have been morphed and warped to generate 4 datasets, to find out if warping could help with improving facial recognition systems by making them robust against morphing attacks.

Additional Key Words and Phrases: Biometrics, Face Morphing, Face Morphing Attack, Morphing attack detection.

1 INTRODUCTION

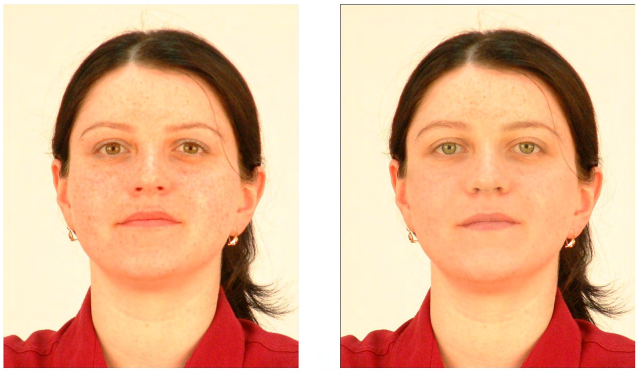


Fig. 1. Two faces which look similar

Are the two pictures in Fig. 1 the same person? However subtle, taking a closer look would show that the iris colors differ and the skin tone has become more washed on the second picture. In Fig. 2 it can be seen that the second picture is a Morph, created from two different identities. In an experimental evaluation [3], it is shown that different types of face manipulation, i.e. retouching, face morphing, and swapping, can significantly affect the biometric performance of face recognition systems and hence impair their security. Face morphing is a technique to blend facial images of two or more subjects such that the result resembles both subjects. Face morphing attacks pose a serious risk for any face recognition system. Without automated morphing detection, state-of-the-art face recognition systems are extremely vulnerable to morphing attacks.

The most common approach to face morphing is by first detecting landmarks in both contributing faces, then define triangles in the images, determining an average geometry, and mapping the averages of the textures in the triangles from the contributing faces to the averaged geometry. An example of a face morphing attack could be thought of as follows: taking two pictures of different people's faces (closely resembled in age, gender, and/or ethnicity) for example in Fig. 2 *Person 1* and *Person 2*, and creating a morphed image M using them. Using a current face recognition software one may find that the M image is able to dupe the facial recognition into thinking that the pairs (*Person 1*, M) and (*Person 2*, M) are matching faces. If the quality of the morph M is high enough it is able to create problems with identification in visa applications and at border control [1]. To combat this, this paper investigates if warping (Appendix B) the faces helps to combat this shortcoming. From the PUT [4] 50 persons faces have been taken to create a baseline performance then warping the faces and comparing the performance at various threshold via a simple facial recognition system.

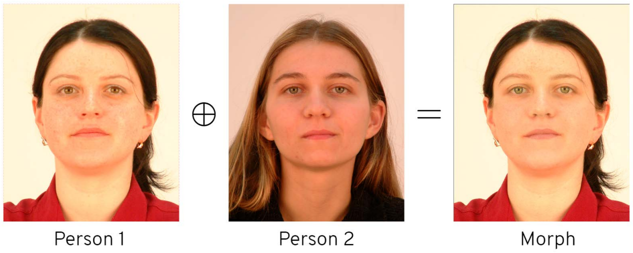


Fig. 2. Morph composition

To construct a convincing morphed face that appears to be genuine, the contributing faces from *Person 1* and *Person 2* have to be cropped to somewhere below the lip and above the chin, should exclude the ears and no hair, then the morphing can be applied, creating a mask M_{mask}. Afterwards, either one of the contributing faces could be merged with the mask, and could need to be post-processed by hand, such a post-process could be blending the edges of the skin color on the M_{mask} with the skin color of the contributing face of *Person 1*.

2 RESEARCH QUESTION

- **Main question** To what level will the proposed system increase the robustness against morphing attacks of Face Recognition Systems?
- **RQ1** How well can the resulting face recognition system perform in the task of face recognition?

3 RELATED WORK

3.1 Facial recognition

Facial recognition is a computer vision task that maps an individual's facial features mathematically and stores the data as a faceprint.

TScIT 37, July 8, 2022, Enschede, The Netherlands

It is done via extracting facial features and comparing them to each other. In this research a simple facial recognition system is used in the following manner: faces are encoded (some info about face encodings in Appednix A) are used to get a faceprint (a 128 dimensional feature vector) using $face\_distance()$ from the python library *face_recognition*. Using the encodings, euclidean distances can be calculated between two face encodings, where lower values mean the faces are more similar. A cutoff point (threshold) can be chosen to distinguish identities, in general *0.6* is suggested. This means that if the distance between two faces would be *0.4*, at *threshold = 0.6* the facial recognition system would consider the two pictures to show the same person (same identity).

3.2 Performance metrics

To be able to evaluate of biometric system accuracy many genuine and impostor attempts are made with the system and all similarity scores are saved. By applying a varying score threshold to the similarity scores, pairs of FRR and FAR (or FNMR and FMR) can be calculated. Results are presented either as such pairs, i.e. FRR at a certain level of FAR, or in plots (see below). Rates can be expressed in many ways, e.g. in percent (1%), as fractions (1/100), in decimal format (0,01) or by using powers of ten (10 -2). When comparing two systems, the more accurate one would show lower FRR at the same level of FAR. Some systems don't report a similarity score, only the match/non-match decision. In that case it is only possible to gain a single FRR/FAR pair (and not a continuous series) as result of a performance evaluation. If the mode of operation (the security level) is adjustable (i.e. we have a means of controlling the internally used score threshold), the performance evaluation can be run again and again in different modes to obtain further FRR/FAR pairs. There are two common ways of plotting performance evaluation results:

- DET graph (Detection Error Trade-off) plots FRR (Y -axis) vs. FAR (X -axis), i.e. false negative vs. false positive rate, often using logarithmic scale (at least for the FAR axis). As the Y -axis shows the number of match errors, the curve that is closest to the bottom of the plot corresponds to the best biometric performance.
- ROC graph (Receiver Operating Characteristic) plots true positive (1 - FRR) vs. false positive rate (FAR). Best biometric performance near the top of the plot.

To measure the facial recognition systems performance against morphing, in this research three metrics were taken into account:

- TMR = # TM / # genuine comparisons
  - True Match (TM) = 2 original images with the same identity have a score < threshold
- FMR = # FM / # impostor comparisons
  - False Match (FM) = 2 original images with different ids have a score < threshold
- MAR = # MA / # mated morph comparison
  - Morph Accept (MA) = comparison of morph of images with id1 and id2 with another image with id1 or id2 resulting in a score < threshold

In order to understand the performance metrics used, face comparison types are included as reference in 3.

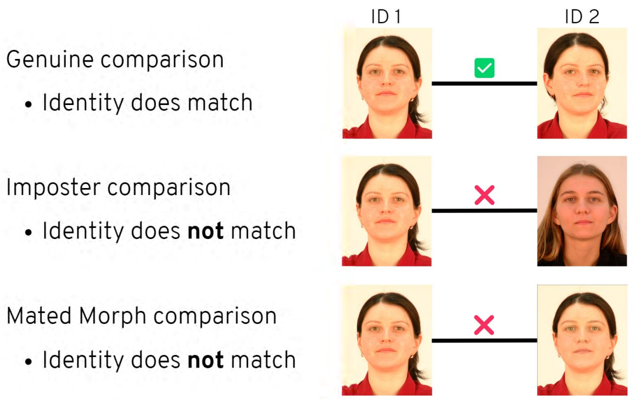


Fig. 3. Types of comparison

3.3 Previous works

In recent years, various papers have proposed the detection of face morphing attacks with different detection approaches. Andrey et al. has made an overview of recent advances in assessing and mitigating the face morphing attack [5]. Tables with matching performance, measured in morph acceptance rate (MAR) and false rejection rate (FRR) can be viewed of recently released related works. In their 2019 paper, Scherhag et al. [8] used photo response non-uniformity (PRNU) analysis which comes from the slight variations among individual pixels during the photoelectric conversion in digital image sensors, achieving in their best configuration a D-EER of 11.2%. In their 2020 paper, Scherhag et al. [10] utilized the information of the embeddings (feature vectors) of the ArcFace algorithm Deep Face Representation (DFR) achieving a D-EER of 3%. In the above papers, D-EER (Detection Equal Error Rate) is used as the accuracy of the detection algorithms, which is at the decision threshold where the proportion of attack presentation incorrectly classified as bona fide presentations (APCER) is as high as the proportion of bona fide presentations incorrectly classified as presentation attack (BPCER). Single-image Morphing Attack Detection (S-MAD) can be grouped to the following three classes: texture descriptors, e.g. in [6, 9, 10], forensic image analysis, e.g. in [8, 11], and methods based on deep neural networks, e.g. in [2, 7]. These differ in the artefacts they can potentially detect.

4 PROBLEM STATEMENT

The aim of this assignment is to investigate robustness of morphing face recognition. For any facial recognition system, getting the threshold right is challenging enough, and introducing the possibility of morphs further complicates the situation. To showcase this problem, in Fig. 4 the faces inside the red area are considered to be close enough to be the same person, showing a face distance threshold of less than 0.5. However on the edge of the red area, a red arrow points to a face which has a distance of 0.671 while it is visible to us that it is the same person. If the threshold would be above that number, the facial recognition system would consider it the same identity.

In Fig. 5 the threshold has been increased to 0.7, and now the previously outlier picture is included in the same identity. However a morph can be seen, with a face distance score of 0.437 which is inside the red area, and is considered by the facial recognition system to have the same identity. This is false however, since the eyes are different color, and the skin has become more white-washed.

The proposed approach is to create warps of the faces. First, a reference geometry is defined e.g. based on landmarks of "the average face". Next, the texture of all faces is mapped on this reference geometry, resulting in a set of faces with the same geometry, but different textures.

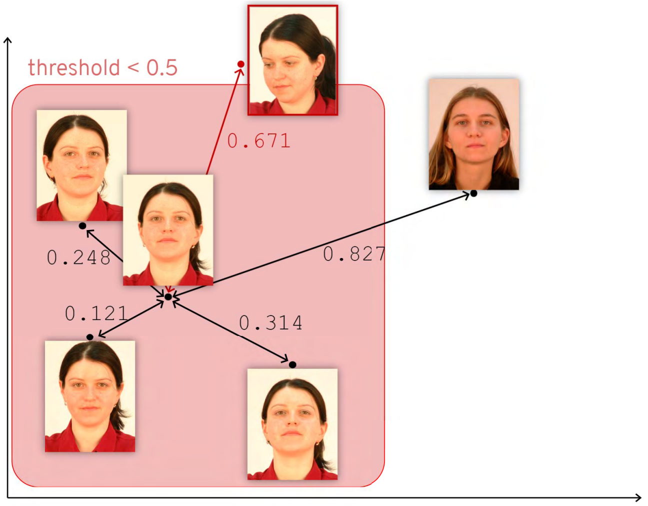


Fig. 4. Faces in red area are below threshold

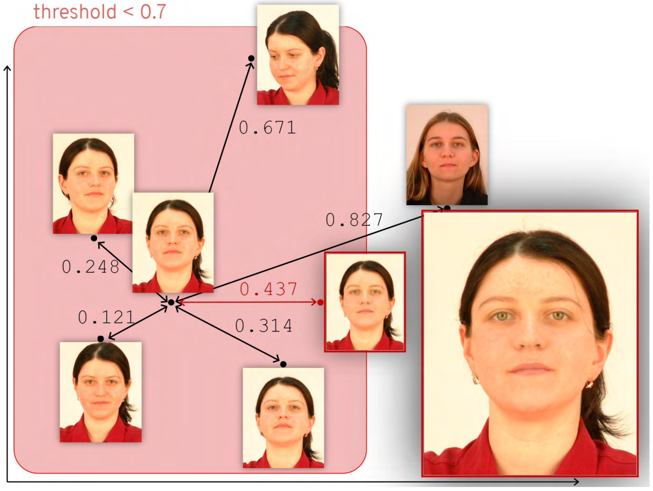


Fig. 5. A morphed face is introduced

5 PROPOSED METHODOLOGY

This section will detail the steps taken to answer the proposed research questions. Due to limited time, this research takes one picture as a reference from the PUT database, and uses that to warp the other images.

5.1 Process

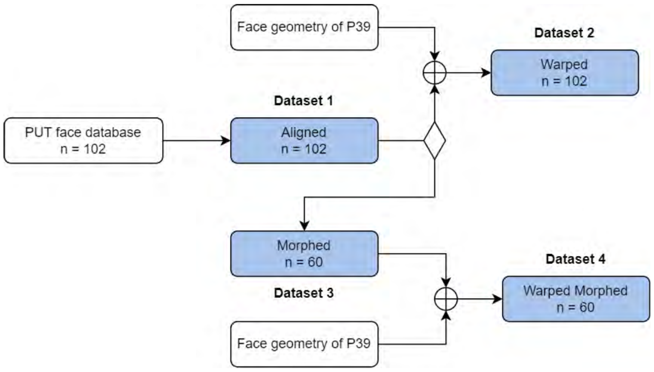


Fig. 6. Creation of datasets

Fig. 7. Person 1 from PUT database

For this research, the PUT database [4] was used, which contains 10000 images of 100 people in different face positions, and the images are of size 2048 × 1536. From it, the first 50 people's frontal picture was chosen by hand, such that the face and the eyes were both facing as forward as possible, and for each person at least 2 picture was chosen, giving a total of 102 pictures, example picture shown in Fig. 7. An overview for the creation of datasets can be seen in 6

The research was conducted on a virtualized Ubuntu 20.04 LTS, programmed in python, with the libraries that are included in appendix A. First, the images were aligned and resized to the same dimensions of 800 × 650 pixels (from now on Aligned images), as Fig.

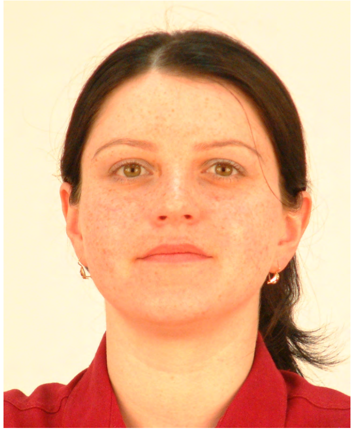
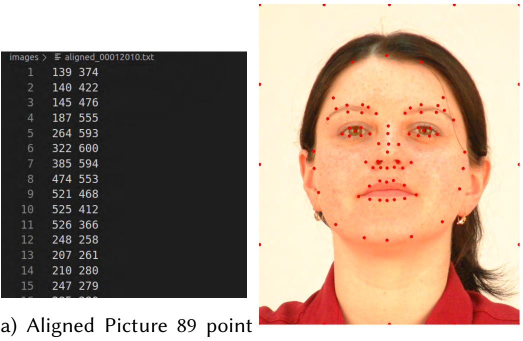


Fig. 8. Aligned Picture of Person 1



```
139 374
140 422
145 476
187 555
264 593
322 600
385 594
474 553
521 468
525 412
526 366
248 258
207 261
210 280
247 279
```

(a) Aligned Picture 89 point landmarks

(b) Landmarks on Person 1

Fig. 9. Landmarks

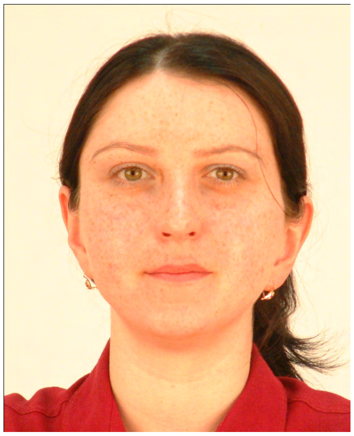
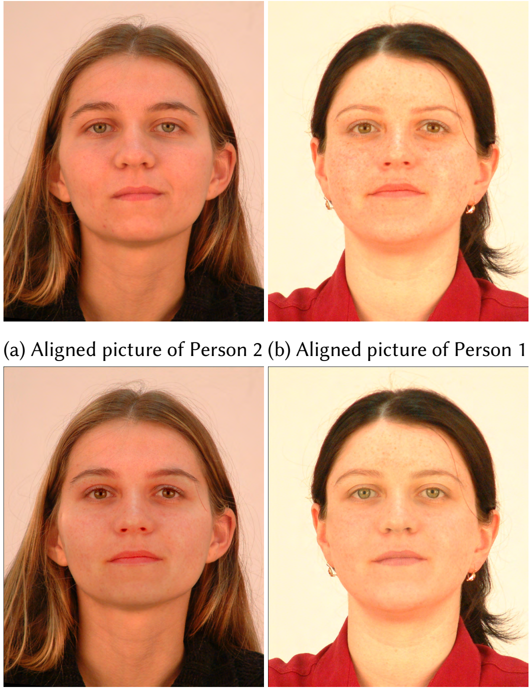


Fig. 10. Warped picture to chosen face



(a) Aligned picture of Person 2 (b) Aligned picture of Person 1

(c) Morph of Person 1 and Person 2 (d) Morph of Person 2 and Person 1

Fig. 11. Morphs

8 illustrates, then landmarks were extracted with Dlib (68 points) and STASM (76 points) and combined into 89 landmark points saved in txt files for each face (Fig. 9a). At this point, the first dataset has been created, consisting of aligned faces (from now on **Dataset 1**). Then an average was taken of all of the 89 landmarks for the 102 faces, and saved in a txt file.

At this point, due to time constraints, the most average-looking face (according to personal opinion) was chosen from the 50 people, and the average landmark points were exchanged for the landmark points of the chosen person. Then, taking the **Dataset 1** faces, each face has been warped to the landmark points of the chosen person, as Fig. 10 illustrates, thus the second dataset has been created (from now on **Dataset 2**). For more information about warping

Next, from **Dataset 1** faces, the faces have been paired such that the most similar faces become paired, and since each person has at least 2 pictures, the pairings that constitute the same person have been disregarded. Now each pairing is morphed into each other, creating 2 faces, in the following manner: Person 1's face Fig. 11b is morphed onto Person 2's face such that Person 1's face texture is more prominent, illustrated by Fig. 11c (take note on eye and mouth color change), all while retaining the original geometric shape, then the same thing happens vice versa Fig. 11d. Now we have our third dataset (from now on **Dataset 3**), containing images such as Fig. 11c and 11d. Finally, **Dataset 3** is also warped to the chosen face's geometry, creating our fourth dataset, as illustrated by Fig. 12 (from now on **Dataset 4**).

In the following, the findings will be discussed.

6 FINDINGS

The face recognition system as previously described, uses face distances to measure euclidean distance via *face_encoding()*, giving a

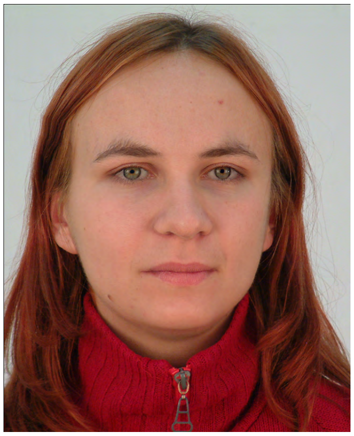
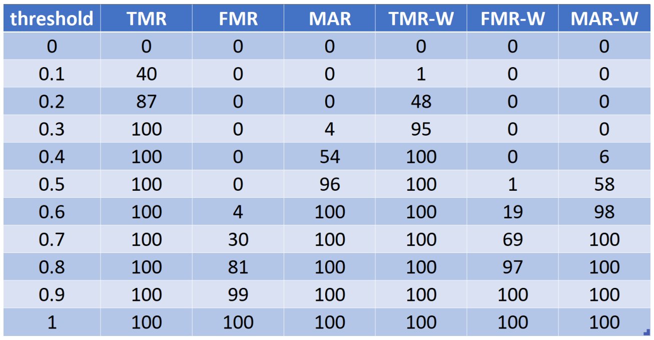


Fig. 12. Warped morphed picture of Person 1

measurement from *0.0* meaning exactly same picture, to an undefined upper limit. To answer the research question **Dataset 1** with **Dataset 3** has been compared against **Dataset 2** with **Dataset 4**. **Dataset 1** with **Dataset 3** shows the baseline performance, **Dataset 2** with **Dataset 4** shows what performance would one get if the faces were warped (signed by ending the performance metric column name with "-W"). Performance metrics for measuring effectiveness of the facial recognition system is done via the metrics TMR, FMR, while MAR measures the robustness against morphing attacks.
TMR of 0 means no correct matches have been made, higher is better.
FMR of 0 means no false matches have been made, lower is better.
MAR of 0 means no morphs have been accepted, lower is better.

Thresholds were defined from 0.0 to 1.0, and the measurements were compared.

(1) Dataset 1 with Dataset 3 on Fig. 14
(2) Dataset 2 with Dataset 4 on Fig. 15



threshold	TMR	FMR	MAR	TMR-W	FMR-W	MAR-W
0	0	0	0	0	0	0
0.1	40	0	0	1	0	0
0.2	87	0	0	48	0	0
0.3	100	0	4	95	0	0
0.4	100	0	54	100	0	6
0.5	100	0	96	100	1	58
0.6	100	4	100	100	19	98
0.7	100	30	100	100	69	100
0.8	100	81	100	100	97	100
0.9	100	99	100	100	100	100
1	100	100	100	100	100	100

Fig. 13. Performance metrics over thresholds 0.0 - 1.0 where warped results are marked with "-W"

In Fig. 13 the following can be noticed:

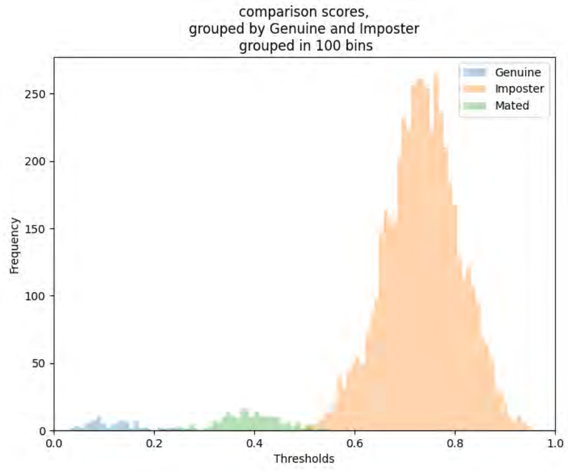


Fig. 14. Histogram of Aligned (D1) with Morhped (D3)

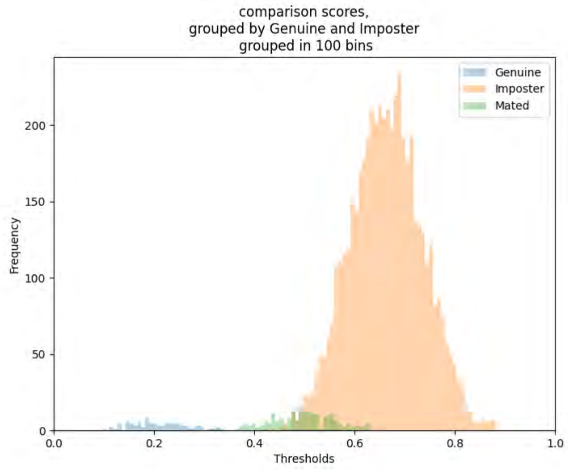


Fig. 15. Histogram of Warped Aligned (D2) with Warped Morhped (D4)

(1) comparing TMR with TMR-W the performance of facial recognition has changed, a threshold of 0.3 no longer accepts every identity that should be accepted
(2) comparing FMR with FMR-W the performance of facial recognition has changed, a threshold of 0.5 would allow for one face to be falsely matched with a different identity
(3) comparing MAR with MAR-W the detection of morphs has become more accessible, at threshold of 0.3 no morphs are accepted while faces were warped, and at 0.4 only 6 were accepted as opposed to MAR's 54, and at 0.5 only 58 instead of 96, showing a clear trend of correctly classifying while warped.
(4) based on the above, a threshold of 3.5 would be ideal.

Comparing Fig 14 and 15:

- in the former a large spread of Genuine can be noticed, and Mated morphs are more mixed

- in the later line between Mated morph and Genuine seems to be more clearly separable

7 DISCUSSION

7.1 Conclusion

The lower morph accept rate (MAR) at the same threshold when the picture is warped indicate that warping pictures does create more robust face recognition system. It is noteable that at the same threshold, the facial recognition system performs worse if the pictures are warped.

7.2 Future work

Due to limited time, some decisions had to be made that leave some of the research space unexplored.

- After creating **Dataset 1**, the faces were warped to a preexisting face shape, here instead could be chosen the average face of the sample or a different face geometry.
- The sample size was 102 in the research which could be either increased or decreased.

REFERENCES

[1] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. 2014. The magic passport. *IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics*. https://doi.org/10.1109/BTAS.2014.6996240

[2] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. 2021. Face morphing detection in the presence of printing/scanning and heterogeneous image sources. *IET Biometrics* 10 (5 2021), 290–303. Issue 3. https://doi.org/10.1049/BME2.12021

[3] Mathias Ibsen, Christian Rathgeb, Daniel Fischer, Pawel Drozdowski, and Christoph Busch. 2022. Digital Face Manipulation in Biometric Systems. *Advances in Computer Vision and Pattern Recognition* (2022), 27–43. https://doi.org/10.1007/978-3-030-87664-7_2/TABLES/2

[4] Andrzej Kasiński, A Florek, and Adam Schmidt. 2008. The PUT face database. *Image Processing and Communications* 13 (01 2008), 59–64.

[5] Andrey M and Andreas W. 2018. An Overview of Recent Advances in Assessing and Mitigating the Face Morphing Attack. (2018), 1017–1021.

[6] R. Raghavendra, Kiran B. Raja, and Christoph Busch. 2016. Detecting morphed face images. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 1–7. https://doi.org/10.1109/BTAS.2016.7791169

[7] R Raghavendra, Kiran B Raja, Sushma Venkatesh, and Christoph Busch. 2017. Transferable Deep-CNN features for detecting digital and print-scanned morphed face images.

[8] Ulrich Scherhag, Luca Debiasi, Christian Rathgeb, Christoph Busch, and Andreas Uhl. 2019. Detection of Face Morphing Attacks Based on PRNU Analysis. *IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE* 1 (2019). Issue 4. https://doi.org/10.1109/TBIOM.2019.2942395

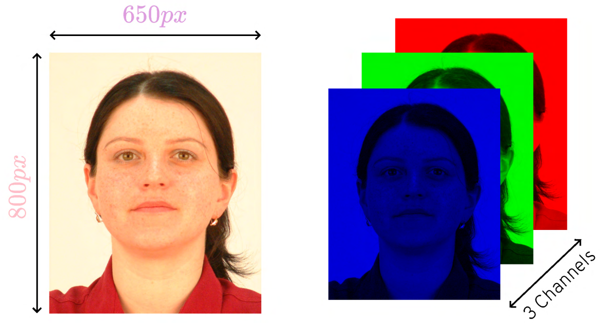
[9] Ulrich Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch. 2017. On the vulnerability of face recognition systems towards morphed face attacks. *Proceedings - 2017 5th International Workshop on Biometrics and Forensics, IWBF 2017* (5 2017). https://doi.org/10.1109/IWBF.2017.7935088

[10] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. 2020. Deep Face Representations for Differential Morphing Attack Detection. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY* 15 (2020), 3625–3639. https://doi.org/10.1109/TIFS.2020.2994750

[11] Le Bing Zhang, Fei Peng, and Min Long. 2018. Face Morphing Detection Using Fourier Spectrum of Sensor Pattern Noise. *Proceedings - IEEE International Conference on Multimedia and Expo* 2018-July (10 2018). https://doi.org/10.1109/ICME.2018.8486607

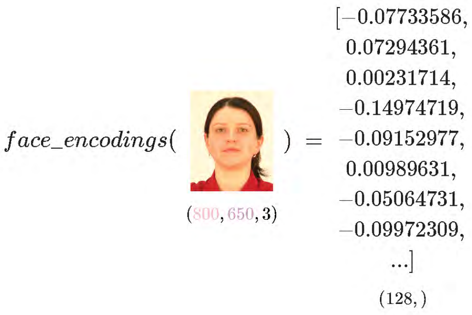
A FACE ENCODING

Face comparison is done via face encoding. As Fig. 16 shows, a picture consists of X width and and Y height, with 3 color channels. For an image with 800 height and 650 width this would mean one could compare 1.560.000 data points, so each pixel. Face encoding (which was used in this research) reduces this to a 128 dimension feature vector. The following implementation was used:
https://github.com/ageitgey/face_recognition/blob/master/examples/face_distance.py
It uses a ResNet behind the scenes, which was trained to give to faces which are similar encodings that are also similar.



$$shape : (800, 650, 3) \rightarrow 1.560.000\ datapoints$$

Fig. 16. Data shape



$$face\_encodings(\ \underset{(800, 650, 3)}{\square}\ ) = \underset{(128,)}{\begin{matrix}[-0.07733586,\\ 0.07294361,\\ 0.00231714,\\ -0.14974719,\\ -0.09152977,\\ 0.00989631,\\ -0.05064731,\\ -0.09972309,\\ \ldots]\end{matrix}}$$

Fig. 17. Face encoding

B FACE WARPING

Warping is the method of morphing a face to a different geometric shape. For warping, the two pictures have to be aligned, for example to the eyes. In Fig. 18 Person 1's face is warped to the face geometry of Person 39, as can be seen by the red landmark dots.

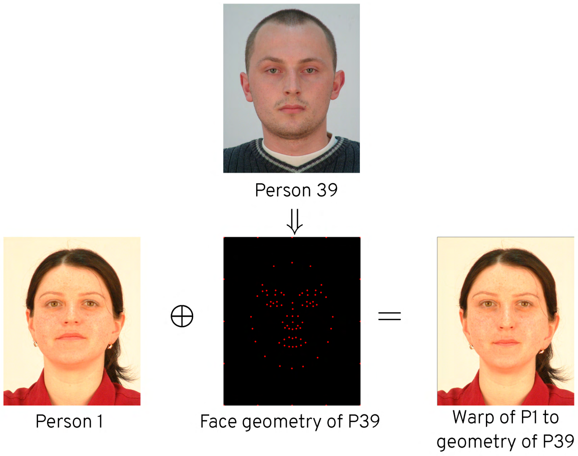


Fig. 18. Warping