Finding the Suitable Doppelgänger for a Face Morphing Attack

Alexander Röttcher*, Ulrich Scherhag†, Christoph Busch†
* Horst Görtz Institute for IT-Security (HGI), Ruhr-Universität Bochum, Germany
† da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
alexander.roettcher@rub.de, {ulrich.scherhag,christoph.busch}@h-da.de

Abstract

ID cards are uniquely linked to one individual via a printed or electronically provided facial image. Even though the face is usually treated as universal and distinctive characteristic, twins can weaken this distinctiveness because of their biological similarity. Also, one might sometimes falsely recognise an unknown person as a friend – colloquially named a Doppelgänger. Recently it was demonstrated that this biological effect of similar data subjects can be purposefully established between two individuals in order to improve the vulnerability of the so-called morphing attack. This image manipulation technique creates a melted facial image which is similar to two or more data subjects. If embedded into an ID card, the manipulated reference image can be used by all participating individuals and thus the concept of a unique link is broken. This work elaborates the rather neglected part of selecting morph pairs based on a similarity score instead of a simple random assignment. It discusses the applicability of different possible algorithms. The finally developed approach considers complex real-world constraints while being executable in a reasonable amount of time and producing acceptable large morph sets. It is shown that this algorithm greatly increases the vulnerability of automated face recognition systems. Surprisingly, it also proves that an effective pre-selection of pairs questions the need of in-depth optimized morphing algorithms.

1. Introduction

In the initial face morphing scenario of Ferrara *et al*. [6] a criminal and an accomplice succeed to deceive human officers and automated face recognition systems (FRS) with their *magic passport*. Such automated systems are for example installed at the airport in order to protect the border and to accelerate the passenger's passport inspection at the same time. Basically, a wanted criminal hides his identity by using someone's else passport. The *International Civil Aviation Organization* (ICAO) defined the facial image to be used to link a subject to his ID card [11]. If a manipulated image is embedded in a valid passport, this unique link is broken. But how is this possible? In compliance with ISO/IEC 19794-5 [13] and the ICAO specification [12], an applicant is allowed to bring a self printed version of a previously taken picture to the passport application office which is the common process in many countries [6]. One might argue that a human officer still verifies this image by comparing it to the applicants visual appearance. This is where *morphing*, an image manipulation which combines the biometric information of two or more inputs within one image, plays an important role. If applied, a passport image representing the facial information from both, the criminal and the accomplice, can be presented to the human officer, see Fig. 1. If the morph is of sufficient quality, various researchers have shown that the morphed image can match to different probe images from both subjects in the comparison done by a human or a machine [23]. Thus, the accomplice can apply for a passport under his name and the criminal crosses the border by using this valid passport.

(a) Subject A. (b) Morph. (c) Subject B.
Figure 1. Example of two morphed images.

The detection of these manipulated images is an active area of research with missing acknowledged robust countermeasures due to various challenges [23]. One of them, the automated generation of large sets of high quality face morphs, is very important for meaningful statistical analyses. While manual morphing of one image is very cost-efficient, today's researchers are faced with a need of automated morphing techniques. Amongst others, machine learning algorithms for morphing attack detection expect

large sets of training and test data sets. Morphing is, however, not limited to the image manipulation itself. Before images are morphed, the data subject pairs for the morphing process have to be selected in a preferably realistic way. Until today, no suitable and directly applicable database of mated images is publicly available [23]. Instead, today's works often simplify this step by randomly selecting partners of the same gender. The attacker's goal is, however, to create a reference image with a high similarity score when being compared to all contributing individuals. This questions if an attacker would ever choose his accomplice randomly. On the other hand, similarity based selection has been a little discussed topic, e.g. in [4, 22, 27]. A general empirical proof of its morph improvement is still missing and most current approaches are not hardened against unrealistically biased sets of morph pairs which, in turn, manipulates the reported detection performances.

The remainder is organized as follows: Sect. 2 provides an overview of related work. The complexity of a realistic morph candidate pair selection with acceptable large sets of morphs is discussed in Sect. 3, followed by the proposal of a newly developed algorithm. The conducted experiment to prove the algorithm's impact on FRS is described in Sect. 4 and a conclusion drawn in Sect. 5.

2. Related Work

Ferrara *et al*. [6] first discovered the effectiveness of morphing as passport image manipulation technique. Since then, different algorithms which morph two or more images have been evaluated and optimized [18, 23]. Most of them rely on so-called landmarks, defined points of the human face like eyes or nose. Generally, one can distinguish between methods which morph the full image and those that only consider some areas, *e.g*., the *splicing morphs* presented by Makrushin *et al*. [17]. To reduce the number of visible artefacts, this technique only morphs the face and splices the result into the background of one of the inputs.

Vyas *et al*. [27] developed a method to find the most suitable animal facial image to be morphed with a human facial image. For each class (human and different animals) the respective training image pixels are clustered by using k-means to derive so-called *textons*, the cluster centres, for each class. For a given test image, the distances to the textons are then compared to find the most suitable animal image. Recently, Damer *et al*. [4] analysed the influence of different accomplice selection approaches on morphing attack detection. Their proposed selection strategies are based on the most similar pairs with the same gender, randomly selected pairs with the same gender, and most dissimilar pairs without gender restriction. Here, similarity is derived by the euclidean distance of the respective OpenFace representations. OpenFace is an open-source FRS developed by Amos *et al*. [1]. A thorough search yielded that these are the only works which show an indication that the targeted selection of pairs for morphing can influence the human and machine ability to detect morphs. However, they include only holistic approaches which consider features extracted from the whole image: textons or OpenFace. Thus, it is unknown which attribute of the human face is the most relevant for the similarity based pairing process. Scherhag *et al*. [22] suppose that the consideration of soft-biometric attributes like gender or age creates more lookalike morphs.

Similarity based pairing simulates the process of multiple criminals searching for their lookalike accomplice. Damer *et al*. limit the number of partners per subject to two. Unfortunately, the authors do not report how this limitation is assured. It is assumed that a subject is discarded from a list of considered accomplices once it has reached this limit. In contrast, Vyas *et al*. concentrate their evaluation on the selection of one animal image which fits best to all other human images. Both approaches might not be optimal to simulate the criminal-accomplice-pairing: Teli *et al*. [26] showed that human faces can be classified according to their false-acceptance-rate of an FRS, i.e., some subjects are more often falsely accepted than others when presenting another's subject's passport image. Thus, the proposed pairing simulations could result in a set with a great variance in terms of similarity between selected pairs and a biased amount of pairs a respective subject is part of.

It is important that images used for morphing and those as probe for the vulnerability assessment are different and members of disjoint sets to avoid any unwanted correlation [22]. Standardized vulnerability metrics against the morphing attack are currently not available [23]. Makrushin *et al*. [17, 18] and Scherhag *et al*. [22] provide different but similar metrics. This work uses a variation of the *mated morph presentation match rate*, the *ProdAvg-MMPMR*, here abbreviated as MMPMR, as proposed by Scherhag*et al*. [22]. Basically, a morphing attack is considered successful if all contributing subjects can deceive a human examiner or FRS with the morphed image, *i.e*., all subjects are accepted. In addition, it takes into account that each subject might have multiple attempts to pass the border. Thus, the MMPMR lies between 0 and 1 where 1 implies that for all morphed images all probe images from all contributing subjects are successfully matched against the respective morph. For its interpretation the *false-non-match-rate* FNMR of an FRS is important to be known: Naturally, the MMPMR can never be higher than the *genuine-match-rate* $GMR = 1 - FNMR$ as otherwise morphing would improve the intra-subject acceptance.

3. Morph Candidate Pair Selection

Before morphing, an attacker searches obviously for his most similar accomplice, a Doppelgänger. Thus, the analysis of morphs from dissimilar inputs is out of scope. Sim-

ilarity between two images has to be expressed as numerical value to construct a ranking from all possible pairs. Also, images can be classified by categorical data, *e.g.*, gender. All subjects are allowed to provide different facial images.

How can the best partner be found? In real-world, the attacker could gather images from social media friends or, as recently done [20], join forces with others to find the most suitable accomplice. Within the experiment, a set of pairs P from the set of subjects S is sought such that only for a minimal number of pairs $p = (a \in S, b \in S) \in P$ another subject $x \in S$ with $a \neq b, a \neq x, b \neq x$ exists where (a, x) has a smaller distance D than (a, b):

$$min(|\{p \ : \ \exists x(D(a,b) > D(a,x))\}|). \tag{1}$$

The distance D of two subjects is derived by comparing the facial images. This is not a limitation as morphs are created from images. If just the top ranked pairs are selected, a dataset of morphs based on images from few subjects only might be created [26]. To avoid a subject-biased pairing, the number of subjects s within P needs to be maximized:

$$max(|\{s \ : \ \exists p(s = a \vee s = b)\}|). \tag{2}$$

With these two constraints a researcher might be tempted to simply go through the sorted list of similarity scores and select a partner as long as a given participation limit has not been reached. This is the approach proposed by Damer *et al*. While some subjects might get their most suitable partner, others have to accept one of their least preferred ones if no other subject is available anymore. Thus, also a minimized average distance is sought:

$$min(\tfrac{1}{|P|} \cdot \textstyle\sum_{p=1}^{P}(D(a_p, b_p))). \tag{3}$$

The following solution is inspired by the commonly known marriage problem [9] where pairs of males and females are sought such that no subject wants to change his partner. In the field of theoretical computer science different terms are used, which are sometimes opposite to the definition of the Harmonized Biometric Vocabulary [14]. Especially the term *matching* in graph theory refers to a subset of pairs where no participant is part of two pairs [3].

3.1. Data Preparation

In case multiple sets of subjects are needed, for example for training and testing of a morphing attack detector, the selection algorithm has to be executed on each set separately. In contrast to Damer *et al*., it is recommended to split the subjects before partners are selected. Otherwise a matching like (a, b), (b, c), (c, d), and (d, a) cannot be split into two disjoint sets with no subjects being part of multiple sets.

Only one image per subject is considered to reduce the algorithm's complexity. For each subject the image with the smallest average distance to the images of all other subjects is selected. The number of subjects within a face database of ICAO compliant images is often very small. Reducing this number weakens training and test results of machine learning algorithms. Also, limiting the number of accomplices could lead to very small groups. The reported results would then be influenced by the group's size. Thus, possible accomplices for each attacker are all other subjects or subjects of the same category: If enough subjects are available, the division by gender or ethnicity is more realistic than arbitrary splits.

Algorithm 1: Proposed Pairing Algorithm

```
Data: Subjects S with their respective facial images
Result: Pairs P with two images per pair
subject_sets = SplitSubjects(S);
forall subject_set in subject_sets do
    forall subject in subject_set do
        images += GetBestImage(subject);
    end
    categories = DeriveCategories(images);
    forall category in categories do
        images = GetImagesOfCategory(category);
        ranking = GetRanking(images);
        // Minimum Weight Matching
        cost_matrix = GetCostMatrix(ranking);
        P += HungarianAlgorithm(cost_matrix);
    end
end
P = DeleteForbiddedPairs(P);
```

Algorithm 1 depicts all conducted steps. For the creation of the ranking a distance function, *e.g.*, based on age, is used. Then, the *minimum weight matching* returns a set of pairs in compliance with the specified optimisation functions. Each subject is part of at most two pairs. This enables the creation of a *bipartite graph*. In such a graph the vertexes can be split into two disjoint sets with all edges connecting two vertexes of different sets [3]. This simplifies the pairing in terms of complexity and efficiency. Further, the dataset is not reduced as the final number of pairs is equal to the number of provided data subjects.

3.2. Minimum Weight Matching

A common example for this basic mathematical problem is the assignment of jobs to workers where each worker has a different cost for the respective job. To apply this idea to the morph candidate pair selection, the ranking is converted into a cost matrix, see Fig. 2a. By going through the ranking, the distance scores are successively inserted into the matrix where row and column are selected by the respective image's ID. If an unknown ID is found, a new row and column is appended to the matrix.

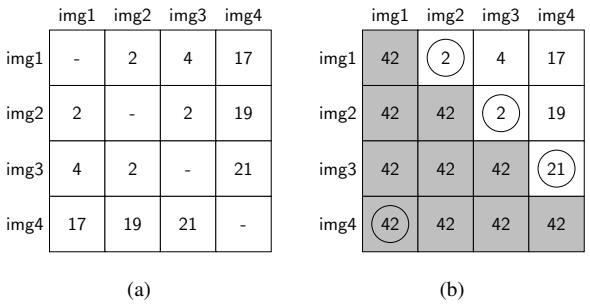


Figure 2. Exemplary cost matrix (a) and the respective matrix with blocked elements in grey as well as the encircled selected pairs (b).

Next, the probability for some combinations needs to be reduced. Images shall not be assigned to itself. Also, the calculated distances are independent of the pair's image order while the exact order might influence the later morphing process, *e.g*., for splicing morphs. The optimal order can, however, differ among morph algorithms. Thus, a pair shall not be present in both orders. To reduce the probability of both kind of pairs in the final result, the main diagonal as well as all entries below are filled with a score equal to two times the maximum score as shown in Fig. 2b. In case one of these combinations is nonetheless present, it is deleted.

To find the matching with the lowest average score, the Hungarian Algorithm by Kuhn [16] is performed on the cost matrix. It returns a matching where each row and column is selected exactly one time while minimizing the average cost. It is polynomial in computation time by the matrix's size. The maximum number of pairs is equal to the number of images and as such the number of subjects. In Fig. 2b an average score of 10.5 is reached (the 42 is replaced by its actual value 17) and no pair deleted.

3.3. Random Pair Matching

To compare the previous algorithm to a random selection, the same prerequisites are necessary to avoid any influence on the reported results. First, one image per subject is randomly selected. Each image is added to a list of available partners. Secondly, the images are paired randomly by ensuring that each image participates in exactly two pairs and that each pairing consists of two different images. Therefore, an image is deleted from the list of available partners once selected. In case the last image $imgN$ can only be paired with itself, the algorithm selects an already created pair $(other1,\ other2)$, removes this pair, and creates two new pairs $(other1,\ imgN)$ and $(imgN,\ other2)$.

4. Experiment

The experiment assesses the influence of pairing strategies on the vulnerability of automated FRS. For some steps a commercial off-the-shelf software (COTS) is used.

4.1. Experimental Setup

6,945 ICAO compliant images from 563 subjects of the FRGCv2 database [19] serve as input. Only subjects with at least two images are considered. The ICAO compliance is manually verified according to ISO 19794-5 [13]. To satisfy the geometric ranges, the minimum inter eye distance of 90 pixels is verified and the face rotated to align the eyes on a horizontal line. Finally, the images are cropped to centre the face while obtaining the required ratios of image width/height and head width/length. Missing pixels are inserted by repeating the pixels of the outer border. The images are not scaled to preserve the maximum quality.

Scherhag *et al*. [22] recommend the use of soft-biometric attributes as indication of similarity. Many morphing algorithms rely on the so-called Delaunay triangulation between averaged landmarks [23]. It is expected that smaller distances between landmarks of both input images result in better morphs. Due to the large number of images all data are derived automatically (except gender). Their general plausibility is manually verified. However, it is important that the final vulnerability influence is based on the tool's choice and not necessarily on the data type.

Age: COTS returns the age for an image. The absolute value of the age difference indicates the similarity. COTS' age calculation is independent of the provided FRS which can therefore still be used for the vulnerability assessment.

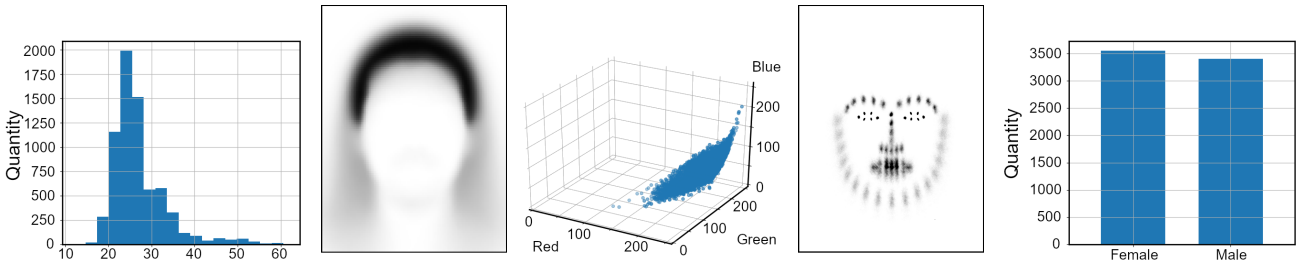
Shape of Hair: MobileUNet [25] returns a mask which values indicate the probability that a certain pixel represents human hair. The mask is binarized and the similarity between two masks derived by their hamming distance.

Skin colour: The RGB values of the nose's bridge without the tip are averaged. This area is found to be the only one which is never covered by human hair nor glasses and not biased by shadows, reflections or over-exposure. It is reliably detectable by dlib-ml [15]. The euclidean distance of two average RGB values represents their similarity.

Landmarks: dlib-ml [15] derives 68 special points of the human face. During comparison, both images are rescaled to a fixed size and the euclidean distance of each respective point calculated. Their median serves as similarity score.

OpenFace: Damer *et al*. [4] already showed that a preselection based on the euclidean distance of two representation matrices produced by OpenFace has an impact on the acceptance rate of a morphed image. This score is, in contrast to the other data, produced by the analysis of the whole face instead of hand-crafted features. By this decision OpenFace's vulnerability against morphs cannot be assessed: A usage of the same information during the pair selection, done by the attacker, and the verification, done by the defender, produces unrealistic results.

Gender: FRGCv2 provides the gender with two categories (male and female). Similarity is defined as the fact that two images belong to the same category.



(a) Age. (b) Shape of Hair. (c) Skin color. (d) Landmarks. (e) Gender.

Figure 3. Attribute distributions of the selected input images. A sufficiently large inter-subject variance is observed for all data.

All considered data are found to have a sufficient inter-subject variance, see Fig. 3. The OpenFace representation matrix is not shown due to its complexity. By the nature of this FRS the inter-subject matrices differ a lot while the intra-subject ones are as similar as possible. The variance is important as otherwise no difference between selection strategies would be observable. The *minimum weight matching* is performed for each numerical data type. One additional set of pairs is returned by a *random matching*. Both are repeated under the consideration of the assigned gender. The distance scores of the targeted selections are compared to the respective random selection and found to be considerably smaller. Thus, the proposed method produces more similar pairs than a random selection.

The applied morphing tools are selected because of their availability, the possibility of automation and their used techniques. Based on these criteria the *OpenCV library* [2] and *GIMP-GAP* [10] are chosen. These libraries are not especially created for morphing attacks but easily available for attackers. In addition, *Face Morpher* by Alyssa Quek [21] is selected as it provides an easy to use interface to quickly generate morphs without the need of programming skills as required for the other two tools. *UBO-Morpher* is applied in various publications of Ferrara *et al.* from the University of Bologna (UBO) [7]. It is designed to create morphs optimized for the morphing attack detection analysis. While this tool is not publicly available it shows what a more advanced attacker could produce to create better looking morphs. All morphing algorithms receive landmarks extracted by dlib-ml [15].

During a manual inspection of the produced morphs no visual difference due to the pairing algorithm could be observed. All kinds of visual artefacts are mostly influenced by the respective morph tool.

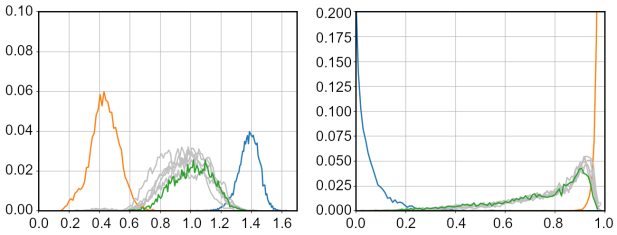
4.2. Face Recognition Vulnerability Assessment

The influence of pair selection strategies is evaluated on two pre-trained FRS: ArcFace [5] and the mentioned COTS, see Tab. 1. Both systems' thresholds are set up according to FRONTEX' requirement of a *false-match-rate* equal to or below 0.1% [8] by randomly selecting 6384 impostor and 6306 genuine attempts from the ICAO compliant image subset of FRGCv2. The GMR and therefore the maximum MMPMR is 99.87% for both FRS.

ArcFace: The MMPMR of all morph candidate sets and different morph tools resides between 61.82% and 98.26%. Thus, all sets and tools are applicable to perform a morphing attack on ArcFace with a high attack success rate. The morph tools OpenCV and GIMP-GAP which morph the whole image achieve always a higher MMPMR than UBO-Morpher and Face Morpher on the same set of morph candidate pairs. The latter tools use the splicing morph technique with the background of the first subject (UBO-Morpher) or an averaged background (Face Morpher). These morphs are thereby more similar to the first subject or have a blurred background which, as the experiment shows, affects the success rate. Almost all sets which are selected by the minimum weight matching have an about 0.64 to 28.87 percent points (pp) higher MMPMR than the respective random selection. Only one set performs 0.14 pp worse which can be neglected due to the small difference. With 98.26% the *Gender-OpenFace* set morphed with GIMP-GAP reaches almost the maximum possible MMPMR. For the given similarity score types and morphing algorithms it can be generally stated that OpenFace performs the best. For the categorical selection, gender performs slightly better than no category. However, this ranking has to be confirmed in future work: By using only one algorithm per similarity score type it cannot be stated if this is due to the algorithm's properties or the type. The presented pair selection algorithm increases the average MMPMR of all morph tools (95.74%) by up to 18.92 pp compared to the respective random selection (76.82%) and decreases the standard deviation between these tools. The worst performing tool, UBO-Morpher, benefits the most of the similarity based pairing. In Fig. 4a exemplary score distributions of ArcFace are depicted. Following the Doppelgänger approach it shows a clear shift of the preselected morph scores towards the genuine scores compared to the random selection. The same effect is observed for the other morph tools, similarity scores and categories. This emphasizes the conclusions drawn above. Despite those good results, a clear

		ArcFace						COTS					
		CV	GAP	UBO	FM	ø	σ	CV	GAP	UBO	FM	ø	σ
No category	Random	83,72%	83,21%	**61,82%**	78,55%	**76,82%**	**10,27%**	96,78%	96,54%	**70,64%**	94,08%	**89,51%**	**12,64%**
	Age	86,86%	87,01%	69,22%	81,96%	81,26%	8,37%	96,56%	96,07%	76,45%	94,02%	90,78%	9,61%
	Landmark	90,23%	90,10%	72,99%	85,68%	84,75%	8,12%	97,70%	97,68%	76,88%	95,94%	92,05%	10,15%
	Shape of Hair	87,19%	86,51%	66,44%	79,83%	79,99%	9,63%	97,37%	97,16%	75,98%	94,54%	91,26%	10,27%
	Skin colour	87,15%	86,17%	67,79%	82,65%	80,94%	8,98%	96,64%	96,38%	72,43%	93,57%	89,75%	11,63%
	OpenFace	**97,92%**	**97,92%**	90,69%	96,43%	**95,74%**	**3,44%**	**99,38%**	99,30%	87,92%	98,20%	**96,20%**	**5,55%**
Gender	Random	89,29%	88,11%	**69,36%**	82,82%	**82,40%**	**9,14%**	97,62%	96,90%	**74,87%**	95,30%	**91,17%**	**10,91%**
	Age	89,15%	88,75%	72,18%	84,25%	83,58%	7,92%	97,31%	97,05%	79,81%	94,60%	92,19%	8,34%
	Landmark	92,54%	92,61%	77,72%	90,13%	88,25%	7,11%	97,84%	97,71%	82,16%	96,68%	93,60%	7,64%
	Shape of Hair	91,92%	91,10%	72,72%	85,38%	85,28%	8,87%	97,88%	97,65%	81,39%	96,42%	93,33%	7,99%
	Skin colour	92,43%	91,93%	74,54%	88,56%	86,86%	8,39%	98,39%	98,21%	81,50%	96,52%	93,66%	8,15%
	OpenFace	97,96%	**98,26%**	91,80%	96,37%	**96,10%**	**2,98%**	**98,99%**	98,82%	89,99%	98,49%	**96,57%**	**4,39%**

Table 1. MMPMR of the attacked FRS with the morphs created by OpenCV (CV), GIMP-GAP (GAP), UBO-Morpher (UBO) and Face Morpher (FM) separated by used similarity score types and categorization criteria for the morph candidate pair selection process. The average MMPMR (ø) and standard deviation (σ) is calculated per set of pairs. Greatest and smallest values are marked in bold per area of interest. A significant influence of similarity based pairing compared to the respective random baseline is observed.



(a) ArcFace distance scores. (b) COTS similarity scores.

Figure 4. Probability density functions for impostor (blue), morph (grey and green), and genuine scores (orange). Similarity based selections are greyed compared to the random selection in green. Morphs are created by UBO-Morpher with no category class.

separation of impostor, morph, and genuine scores is visible even for the similarity based pairing. It confirms the recent findings of Scherhag *et al*. [24] for morphing attack detection using face representations.

COTS: Compared to the results produced by ArcFace, an even higher average MMPMR for all sets is observed. Multiple sets reach almost the maximum possible MMPMR. An FRS not only has to consider security aspects but also the overall throughput (low FNMR). Thus, a good performing real-world FRS does not only detect all impostors but generalizes also different images of the same subject to be robust against the natural intra-subject variance. These are indeed conflicting requirements and can result, as shown here, in an higher morph vulnerability if the generalization performs too well. A greater risk due to targeted selection is still observed: If the pairs are selected by OpenFace scores, the average MMPMR increases by 6.69 pp. Only some targeted selections reduce the MMPMR by up to 0.51 pp (*e.g.*, using skin colour and Face Morpher) which is therefore negligible. The actual score distributions prove this observation as well, see Fig. 4b. A general shift of morph scores towards genuine scores is clearly visible for similarity based pairing. It also shows that the MMPMR cannot be greatly improved as, compared to the observations on ArcFace, the morph scores are already very close to the genuine scores. A threshold closer to the peak of the morph scores could greatly decrease the MMPMR but might also increase the FNMR. So, an adjustment of COTS, to be more robust against morphs, is not as easy as for ArcFace.

All in all, both FRS are more vulnerable if the morph pairs are selected based on their similarity. Simple similarity indications as gender or skin colour are already effective. However, a holistic approach conducted with OpenFace has the overall highest impact. Further, more similar pairs minimize the standard deviation of the morph tools.

5. Conclusion and Future Work

To the authors' knowledge this work presents the first in-depth analysis on the impact of similarity based pair selection. By considering real-world constraints a new method is developed to efficiently select similar pairs. It is then compared to an adapted version of a random selection process which is often found in state-of-the-art morphing attack research. The conducted experiment proves that appropriate pair selection not only increases the morph quality (in terms of FRS vulnerability) but also substantially decreases the standard deviation between different morphing techniques. In other words, an effective preselection reduces the need for a perfect, low-artefact producing, morphing algorithm. This is very important as automated morphing is still error-prone with the difficulty to fully remove all artefacts. In

a nutshell, automated face recognition systems that operate on the purpose of determining the similarity between two facial images are not only vulnerable to morphed faces but can also contribute to a morphing attack by finding optimal pairs of data subjects in a sufficient manner. If these findings can be transferred to a manual inspection, *e.g.*, done by a border officer, is still an open research question.

In future work, the pairing algorithm is expected to be further optimized. Especially a random blocking of two pairs $(img1, img2)$ and $(img2, img1)$ instead of always blocking the second pair is expected to better utilize the Hungarian Algorithm's capabilities. Other holistic approaches to calculate the similarity of two images might be of interest as well. This leads to the final question: Can the vulnerability of a face recognition system be predicted by the average similarity score of the selected pairs?

6. Acknowledgement

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. The authors thank the Biometric System Laboratory of the University of Bologna for providing their UBO-Morpher.

References

[1] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[2] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 25:120–125, Nov. 2000.

[3] A. Chakrabarti. Maximum matching. [Online]. Available: https://www.cs.dartmouth.edu/~ac/Teach/CS105-Winter05/Notes/kavathekar-scribe.pdf.

[4] N. Damer, A. M. Saladié, S. Zienert, Y. Wainakh, P. Terhörst, F. Kirchbuchner, and A. Kuijper. To detect or not to detect: The right faces to morph. In *Proc. IAPR Int. Conf. Biometrics (ICB)*, June 2019.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2019.

[6] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sept. 2014.

[7] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Trans. Inf. Forensics Security*, 13:1008–1017, Apr. 2018.

[8] FRONTEX. *Best Practice Technical Guidelines for Automated Border Control ABC Systems*, 2015.

[9] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69:9, Jan. 1962.

[10] W. Hofer. Creation of morphing animations. [Online]. Available: https://github.com/GNOME/gimp-gap/blob/mainline/gap/gap_morph_main.c.

[11] Int. Civil Aviation Org. *Biometrics Deployment of Machine Readable Travel Documents. Version 1.9*, May 2003.

[12] Int. Civil Aviation Org. *ICAO Doc 9303, Machine Readable Travel Documents - Part 3: Specifications Common to all MRTDs*, 7 edition, 2015.

[13] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19794-5:2005. Information Technology - Biometric Data Interchange Formats - Part 5: Face Image Data*. International Organization for Standardization, 2005.

[14] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 2382-37:2012 Information Technology - Vocabulary - Part 37: Biometrics*. International Organization for Standardization, 2012.

[15] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, July 2009.

[16] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, Mar. 1955.

[17] A. Makrushin, T. Neubert, and J. Dittmann. Automatic generation and detection of visually faultless facial morphs. In *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISIGRAPP)*, Feb. 2017.

[18] A. Makrushin and A. Wolf. An overview of recent advances in assessing and mitigating the face morphing attack. In *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sept. 2018.

[19] NIST. Face recognition grand challenge. [Online]. Available: https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc.

[20] Peng! Kollektiv. Mask id. [Online]. Available: https://mask.id/en.

[21] A. Quek. Face morpher. [Online]. Available: https://github.com/alyssaq/face_morpher.

[22] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch. Biometric systems under morphing atacks: Assessment of morphing techniques and vulnerability reporting. In *Proc. IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sept. 2017.

[23] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019.

[24] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch. Deep face representations for differential morphing attack detection. *IEEE Trans. Inf. Forensics Security*, 2020.

[25] A. Sosa. Real-time semantic segmentation in mobile device. [Online]. Available: https://github.com/akirasosa/mobile-semantic-segmentation/blob/master/README.md.

[26] M. N. Teli, J. R. Beveridge, P. J. Phillips, G. H. Givens, D. S. Bolme, and B. A. Draper. Biometric zoos: Theory and experimental evidence. In *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011.

[27] J. P. Vyas, M. V. Joshi, and M. S. Raval. Automatic target image detection for morphing. *Journal of Visual Communication and Image Representation*, 27:28–43, Feb. 2015.