To Detect or not to Detect: The Right Faces to Morph

Conference Paper · June 2019

DOI: 10.1109/ICB45273.2019.8987316

7 authors, including:

Naser Damer
Fraunhofer Institute for Computer Graphics Research IGD
97 PUBLICATIONS **590** CITATIONS
SEE PROFILE

Alexa Moseguí Saladié
University of Barcelona
6 PUBLICATIONS **62** CITATIONS
SEE PROFILE

Philipp Terhörst
Fraunhofer Institute for Computer Graphics Research IGD
31 PUBLICATIONS **140** CITATIONS
SEE PROFILE

Florian Kirchbuchner
Fraunhofer Institute for Computer Graphics Research IGD
68 PUBLICATIONS **207** CITATIONS
SEE PROFILE

To Detect or not to Detect: The Right Faces to Morph

Naser Damer[12], Alexandra Moseguí Saladié[1], Steffen Zienert[1], Yaza Wainakh[1]
Philipp Terhörst[12], Florian Kirchbuchner[12], Arjan Kuijper[12]
[1]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[2]Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany
naser.damer@igd.fraunhofer.de

Abstract

Recent works have studied the face morphing attack detection performance generalization over variations in morphing approaches, image re-digitization, and image source variations. However, these works assumed a constant approach for selecting the images to be morphed (pairing) across their training and testing data. A realistic variation in the pairing protocol in the training data can result in challenges and opportunities for a stable attack detector. This work extensively study this issue by building a novel database with three different pairing protocols and two different morphing approaches. We study the detection generalization over these variations for single image and differential attack detection, along with handcrafted and CNN-based features. Our observations included that training an attack detection solution on attacks created from dissimilar face images, in contrary to the common practice, can result in an overall more generalized detection performance. Moreover, we found that differential attack detection is very sensitive to variations in morphing and pairing protocols.

1. Introduction

The recent spike in deep-learning based face recognition performance, along with the relatively high social acceptance, has pushed automatic face recognition to be a stable in security sensitive deployments such as identity management (e.g. travel documents) [15]. The critical nature of face biometric applications makes it a target of malicious attacks. One of such attacks is the presentation attack aiming at interfering with the operation of the biometric system, commonly by presenting a fake characteristic [4]. A relatively new sort of such attack is the morphing attack that aims at presenting one face reference image that is, automatically and by human experts, successfully matched to more than one person. These morphing attacks have been constructed by interpolating facial landmarks between the two face images creating the attack (LMA), or as was proposed very recently, by utilizing generative adversarial networks (MorGAN). Different solutions have been proposed to detect such attacks, weather by analyzing the features of the investigated image, or by including an assumed the existence of a bona fide live capture. Some works addressed the generalization issue of detecting approaches when dealing with different types of attacks (LMA or MorGAN) [5], re-digitization [20], or images from different original sources [24]. To create attacks, pairs of face images have to be chosen for morphing. Previous works used the same pairing criteria (usually choosing similar faces) for training and testing data and did not consider the generalization problems or gains that realistic variations in such protocol can produce.

In this work, we study the effect of varying the pairing protocol on the generalization of the detection performance over attacks with unknown pairing protocols and morphing types. To do that, we created a novel attack database with three subsets built with different pairing protocols. The attacks are created using both the LMA and MorGAN approaches. We analyzed the vulnerability of two face recognition approaches to the presented data subsets. We studied the known and unknown (different pairing protocol or attack type) attack detection performance under a single image attack detection and a differential attack detection scenarios, using both, handcrafted and CNN-based features. This is also the first work to test differential attack detection on GAN-based attacks. One of the major conclusion of this work is noticing a significant higher detection performance generalization when training the attack detector on attacks built from dissimilar face pairs. This is contrary to the common practice of assuming the need to train detectors on attacks created from similar faces.

2. Related work

The possibility of creating a morphed face image attack out of two images of two subjects was introduced by Ferrara et al. [7]. They compared morphed images with images of the original subjects using two face recognition solutions,

978-1-7281-3640-0/19/$31.00

and concluded with the high vulnerability of face recognition to such attacks. Other studies found that human experts fails most of the times in detecting morphing attacks [8, 21].

Different solutions were developed to detect face morphing attacks. Ramachandra et al. [18] were first to propose the automatic detection of morphed face images. They did that by extracting local image descriptors as the Binarised Statistical Image Features (BSIF) that tries to capture textural properties of the image, that are later classified using a Support Vector Machine (SVM). Later works looked into using CNN-based features [20], image quality measures [16], the effect of printing and re-scanning the images [20], and differences between triangulating and averaging the facial landmarks on the detection [19].

Other works considered the practical possibility of using a live probe image along the investigated image to detect morphing attacks. This was done either by looking at the differential vector between both images [25], analyzing the absolute distances and angels of the landmarks between images [22], or analyzing the directed distances between these landmarks [3]. Recently, Damer et al. [5] proposed a new possibility of morphing attacks. They built their solution on generative adversarial networks (GAN). They morphed the latent representation of the morphed images and generated the morphing attacks based on that morphed latent vector. These morphing attacks proved to be hard to detect in the cases where they were not considered in the training process of the morphing detector [5]. All previous works used the same protocol in selecting face image pairs (to create the attacks) in the training and testing data splits. This is unrealistic, as the providers of detection algorithms might use different protocols in comparison to the attackers.

3. Databases

This section discusses the presented database, the pairing protocols, the morphing approaches, and analyzes the face recognition vulnerability to the different database subsets.

3.1. Morphing and pairs selection protocols

The database is built on the the CelebA [13] database, composed of 202,599 face images of 10,177 identities and 40 attribute binary vectors. The image size is 178 x 218 pixels. To cover the frontal image condition in the International Civil Aviation Organisation (ICAO) travel document requirements [9], all non-frontal images are filtered out by detecting the central coordinate of the eyes and the upper coordinate of the nose. The two distances between each of the two eyes and the nose landmarks are calculated, and if the ratio of the difference between these distances to any of them was more than 0.05, the image is neglected. Further filtering was performed based on the provided attributes, images labeled as blurry, glasses, hat, were neglected.

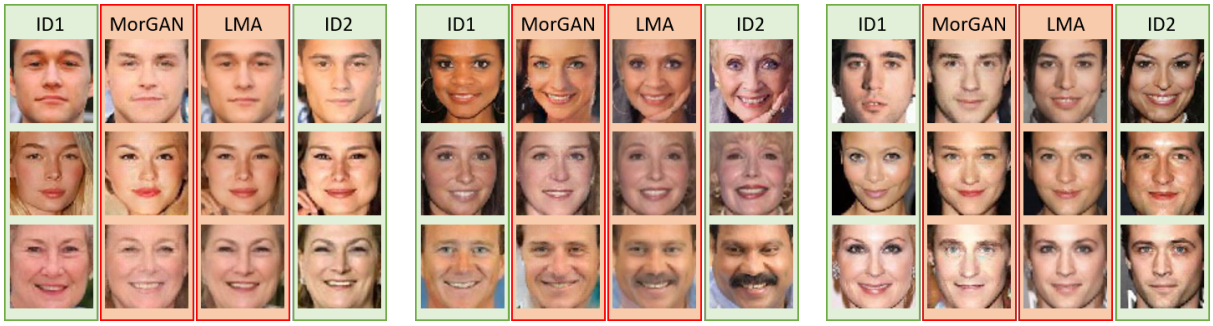
As a starting point, 200 images of 200 identities were manually chosen, split evenly between males and females. These 200 images were chosen to have neutral impression, good illumination quality, and no occlusion. Each of these images is matched twice, with two different images of different identities. The selection of these two other images followed one of three protocols. These protocols depended on the similarity between the key image and the selected paired images. The similarity was measured by the Euclidean distance between the OpenFace representations [2]. The pairing protocols are:

1. Similar (S): Each of the key images was paired with the two most similar faces in the database with restriction of the search within the same gender. This is the typical recommended protocol, which make sense if the goal is to create a confident attack. However, not certainly the protocol to create attack training data for a generalized solution.
2. Random (R): Here, each of the key images was paired randomly with two identities from the filtered database, with restriction within the same gender. This aims at having a more diverse nature of morphing attacks.
3. Dissimilar (D): In this protocol, each key face image was paired with the two most dissimilar images in the database without any gender restrictions.

For each of the protocols, each of the 200 key images was paired twice. This resulted in 400 morphing pairs and 600 bonafide references. For each of the 600 bona fide identities, a second bona fide image was chosen to be a bona fide probe, to be used in the differential morphing detection approach.

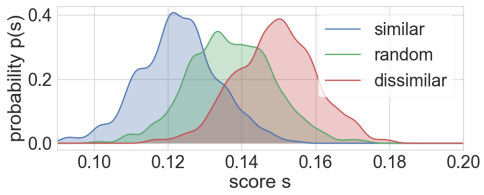
Each of the 400 morphing attacks pairs, in each protocol, is used to create an attack. These attacks are created by the landmark-based approach and using the MorGAN approach [5]. Morphing of the Landmark based attacks (LMA) is performed by detecting 68 landmarks on the face as proposed in [11]. The mean face points for each image are calculated and each image is then warped to sit on these coordinates after performing the Delaunay Triangulation [12]. Only the facial area is morphed and stitched into one of the original morphed images [14]. MorGAN attacks are created for the same image pairs using the approach and network trained in [5]. There, the face morphing is going to be performed in three steps given two face images x_1 and x_2. First, the two images are encoded into the latent space. Then, the resulting latent vectors are linearly interpolated with a factor $\beta = 0.5$. Finally, the interpolated latent vector is decoded into the image space.

The database contained three parts, each using a different pairing protocol. Each part contained 600 bona fide references, 600 bona fide probes, 400 LMA attacks, and 400 MorGAN attacks. The databases are split identically into a disjoint (identity and image) and equal train and test sets

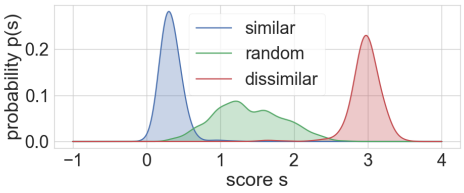


(a) Similar (b) Random (c) Dissimilar

Figure 1: Examples of the morphing attacks (MorGAN and LMA) created from different pairing protocols. In each protocol, Original reference images are on the right and left and attacks in the middel.



(a) VGG



(b) OpenFace

Figure 2: The dissimilarity (distance) scores distributions between the selected image pairs using the three different pairing protocols for both the VGG and OpenFace face recognition approaches.

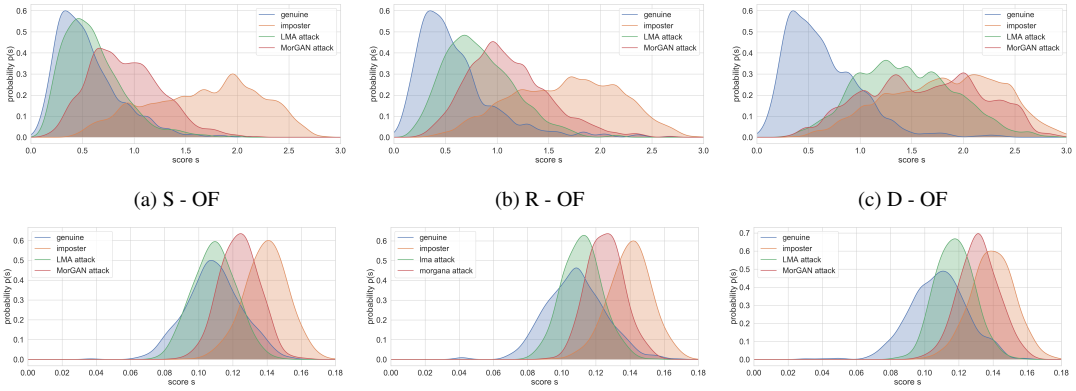
(300/300 bonafide, and 200/200 attacks). These sets are based on a random split of the initial 200 key reference images. Because of computational and structural limitations, the MorGAN attack images are of 64×64 pixels size (below the ICAO recommendations). To allow for a fair detection evaluation, bona fide and LMA attacks were also re-sized to the same size. Examples of the resulting image attacks and the original image pairs are presented in Figure 1 for different pairing protocols.

The distance (dis-similarity) score distributions of the selected morphing pairs (always imposter) under the three protocols are shown in Figure 2. In the figure, the scores produced by Euclidean distance between the VGG [17] and the OpenFace representation of the pairs [2] are shown. The score distributions clearly reflects the pairing protocol.

3.2. Vulnerability of face recognition

We investigate the vulnerability of face recognition algorithms to the produced attacks under different protocols. The vulnerability of two pre-trained face recognition systems are tested, the VGG-face as described in [17] and the OpenFace as described in [2], previously used for vulnerability assessment against morphing attacks [5] and other capture variations [6]. Given an aligned and cropped face image, this pre-trained network produces a discriminant representation of 128 elements. On the other hand, VGG-Face [17] is based on the VGG-Very-Deep-16 CNN. The feature representation is extracted from the last max pooling layer of of 7x7x512 dimensionality. Face representations, from VGG or OpenFace, are compared by calculating the Euclidean distance between two representation vectors.

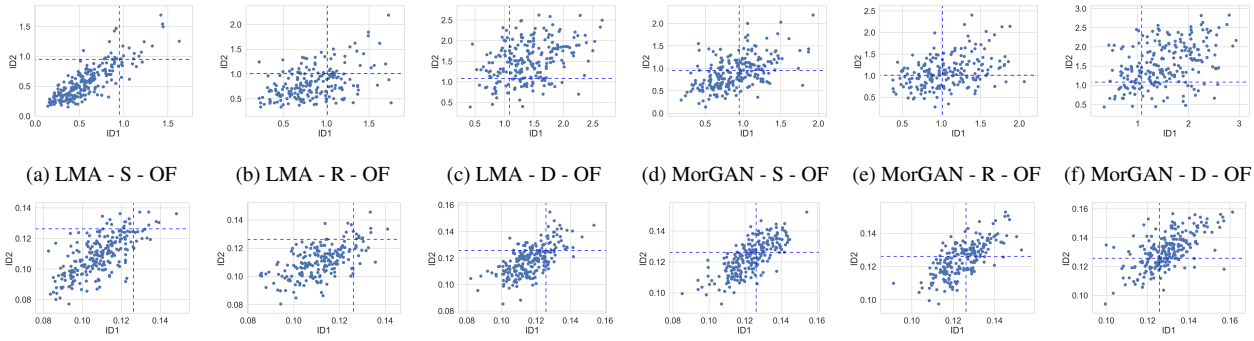
This vulnerability is discussed by showing the comparison score distributions of imposter, genuine, and morphed face attack comparisons to each of the probe original identities contained in the morph. To measure the attacks ability to simultaneously match both original identities, we plot the comparison scores between the morphing attacks and their two original identities images in the probe set. These comparisons scores are shown with respect to the threshold at the equal error rate (EER) operational point to get a relative measure of the attack success. The genuine and imposter comparisons (as well as the EER) are results of the 600 bona fide references cross-compared (NxN) with the 600 probes. The morphing attacks score distribution is based on comparing the 400 morphed images, each with their corresponding two identities in the probe set. The score distribution for these comparisons are shown in Figure 3. To demonstrate the preservation of both identities, we plot the comparison score between the attacks and the first involved identity vs. the one with the second identity in the bottom of Figure 4. The dotted lines in these plots represent the threshold value that achieves equal error rate. From both figures, 3 and 4, one can notice that the LMA attacks are more successful than MorGAN ones. However, LMA attacks becomes much worse when the pairing protocol select images randomly, while the MorGAN attacks remain relatively stable. Both attacks performs poorly, however with some success, when pairing is done between dissimilar images.



(a) S - OF (b) R - OF (c) D - OF

(d) S - VGG (e) R - VGG (f) D - VGG

Figure 3: Vulnerability of two face recognition approaches (OpenFace (OF) & VGG) to LMA and MorGAN attacks with pairing protocols (S, R, and D). The comparison score (dissimilarity) distributions of genuine (blue), imposter (orange), MorGAN (red), and LMA (green).



(a) LMA - S - OF (b) LMA - R - OF (c) LMA - D - OF (d) MorGAN - S - OF (e) MorGAN - R - OF (f) MorGAN - D - OF

(g) LMA - S - VGG (h) LMA - R - VGG (i) LMA - D - VGG (j) MorGAN - S - VGG (k) MorGAN - R - VGG (l) MorGAN - D - VGG

Figure 4: Vulnerability of two face recognition approaches (OpenFace (OF) & VGG) to LMA and MorGAN attacks with pairing protocols (S, R, and D). Morphing attacks comparison scores in comparison to the dotted line representing the threshold at EER.

4. Experiments

4.1. Detection methodologies

Single image attack detection: Our attack detection methodology aims at enabling a wider range of conceptual evaluation and more diverse coverage to the state-of-the-by considering image feature extraction methods of two different natures. One is the hand crafted classical image descriptors, the Local Binary Pattern Histogram (LBPH) [1]. The second is based on transferable deep-CNN features. Both types of features were previously utilized for the detection of face morphing attacks of similar nature to LMA [18][20]. The LBPH features are extracted from the cropped face image. A histogram is calculated for each block of an 8x8 grid of blocks in the face image. These histograms are concatenated to produce the final feature vector describing the image. Each LBP is extracted within a radius of one pixel and eight neighbor pixels. The transferable deep-CNN features are extracted using the well performing, and relatively small OpenFace NN4.SMALL2 model [2]. The extracted feature vector from an image, whether from CNN or LBPH, is classified by a support vector machine (SVM) classifier, to be originated from a morphed or a bona fide image. The approach based on direct single image feature vector classification is referred to as FV.

Feature differential pattern: The approach aims at capturing a pattern of change in the general appearance of the face image when it is morphed with another face image (a probe). This is based on the availability of a live capture image (known to be a bona fide) of the same subject, so that the pattern of morphing change can be measured. This appearance change pattern can be seen in the change of face image representation (feature vector) between the probe bona fide image of a certain subject and a morphed image of the same identity and a different one. This approach was implemented very recently using handcrafted features [25]. This approach will be referred to as differential feature vector DFV. This was implemented with both, the CNN and LBPH features, as used in the FV approach. The differential features are calculated as element wise feature vector subtraction between the feature vectors extracted from investigated image (attack or bona fide reference) and a bona fide probe.

The SVM used for the classification of both FV or DFV utilizes a Linear kernel given the database size. The SVM hyperparameters are found using Bayesian optimization. The SVM classifier (trained on the training data) produces a decision score that represent the confidence degree of the input image being a morphed one rather than a bona fide one. The training and evaluation were performed on the identity-disjoint training and testing sets subsequently.

The attack detection experiments are noted by the overall approach (FV or DFV), the feature type used (LBPH or CNN), by the type of attack used for training the detector (LMA or MorGAN) and the pairs selection protocol (S, R, D), and by the images used for evaluating the detector (LMA, MorGAN, bona fide (BF)) and the pairs selection protocol as well. When the same type of attack and pairs selection protocol is used for training and testing, we will refer to the attack as "known attack". Otherwise, we will refer to it as "unknown attack".

4.2. Experimental setup and metrics

The performance of the FV and DFV detection methodologies using both CNN-based and LBPH features are evaluated on known attacks and every type in unknown attacks. The unknown attacks are the ones created with a different pairing protocol (S, R, D) and the same morphing approach (LMA or MorGAN), or the ones created with a different morphing approach (LMA or MorGAN) and any pairing protocol. This aims at assessing the generalization ability of attack detectors trained on different attack types and pairing protocols to detect known and unknown attacks.

The performance of the morphed face detection is presented as a trade-off between two error rates, the Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER) as defined by the ISO/IEC 30107-3 [10] and advised by recent works [23]. Here, the APCER is the proportion of morphed face presentations incorrectly classified as bona fide presentations. The BPCER is the proportion of bona fide presentations incorrectly classified as morphed face attacks. The detection decision thresholds producing fixed APCER rates on known attacks are calculated. These threshold represent possible decision thresholds chosen for system deployment and they depend on the detection performance of known attacks. BPCER values achieved at fixed APCER rates are reported to enable direct comparison between different solutions at a wide range of operation points, these BPCER rates only depend on the bona fide images, and thus are the same for known and unknown attacks. APCER rates of unknown attacks are reported on the thresholds previously assigned for the fixed known attacks APCER rates, to measure the detectability of unknown attacks. Lower values of BPCER and APCER indicates higher detection performance.

		APCER[%]			
	Train/Test	1	10	20	30
CNN	LMA-S	99.7	98.3	95.7	94.3
	LMA-R	87.3	66.0	55.0	48.0
	LMA-D	11.0	2.0	1.0	0.3
	MorGAN-S	34.7	8.3	3.3	2.3
	MorGAN-R	0.0	0.0	0.0	0.0
	MorGAN-D	99.7	89.0	70.0	49.3
LBPH	LMA-S	45.7	19.7	8.3	3.0
	LMA-R	50.7	6.7	2.7	1.7
	LMA-D	16.0	0.3	0.3	0.3
	MorGAN-S	12.0	2.0	0.0	0.0
	MorGAN-R	3.3	0.0	0.0	0.0
	MorGAN-D	1.0	0.0	0.0	0.0

Table 1: BPCER rates achieved at fixed APCER rates of known attacks using the FV detection approach with both feature types.

5. Results

Starting by the **single image attack detection** approach (FV), and taking a look at the performance of known attacks detection in Table 1, one can notice that CNN features fails to detect known LMA attacks when they are paired using the S and R protocols. The detector performs better when trained and tested on the D protocol attacks. However, this solution is poorly generalizable on unknown attacks as seen in Table 2. Also using the CNN features, known MorGAN attacks are best detected when tested and trained on the R protocol, probably because of attack diversity (Table 1). However, this solution is also poorly generalizable to unknown attacks (both attack type and pairing protocol), as seen in Table 2. Using LBPH features also achieved best known LMA attack detection performance under the protocol D (Table 1). With relatively good generalization ability on both attack types and pairing protocols as seen in Table 3. Training on LBPH-LMA-S and LBPH-LMA-R achieved also good generalization (Table 3), however, with very high BPCER (Table 1). Detecting known MorGAN attacks using LBPH features performed quite well under all pairing protocols (Table 1). These also generalized good on MorGAN attacks produced by other pairing protocols, especially when trained on the pairing protocol D (Table 3). However, they performed poorly on unknown attacks created by LMA. FV detectors trained on CNN features with LMA-S, LMA-R, and MorGAN-D, as well as LBPH with LMA-S and LMA-R performed poorly even on known attacks. They produce relatively high BPCER values and therefore, analyzing their generalization on unknown attacks on these BPCERs is of low relevance. Generally, under the FV detection approach, LBPH features performed in a more stable manner over known and unknown attacks in comparison to CNN. Interestingly, training an attack detector on attacks of the pairing protocol D and LBPH features were able to be generalizable

Train	Test	APCER[%] (known attack) 1	10	20	30
CNN LMA-S	LMA-R	69.0	81.5	87.5	90.0
	LMA-D	28.5	51.5	74.5	80.5
	MorGAN-S	6.0	14.0	21.5	24.0
	MorGAN-R	14.0	23.0	34.5	39.5
	MorGAN-D	36.0	59.0	71.0	73.5
CNN LMA-R	LMA-S	0.5	12.0	22.0	30.0
	LMA-D	96.5	99.5	100.0	100.0
	MorGAN-S	41.5	74.5	86.5	88.5
	MorGAN-R	3.5	9.0	13.0	15.0
	MorGAN-D	8.0	19.5	28.5	32.0
CNN LMA-D	LMA-S	95.5	98.5	99.0	100.0
	LMA-R	98.5	99.5	100.0	100.0
	MorGAN-S	69.0	91.0	93.5	97.0
	MorGAN-R	94.5	99.5	100.0	100.0
	MorGAN-D	100.0	100.0	100.0	100.0
CNN MorGAN-S	MorGAN-R	0.0	0.0	0.0	0.0
	MorGAN-D	34.0	74.5	88.5	91.5
	LMA-S	19.0	57.5	70.5	76.0
	LMA-R	33.0	65.0	78.0	84.0
	LMA-D	10.5	31.0	46.5	57.5
CNN MorGAN-R	MorGAN-S	91.5	99.5	99.5	99.5
	MorGAN-D	40.0	61.0	71.5	80.5
	LMA-S	90.0	97.0	100.0	100.0
	LMA-R	75.5	93.0	95.5	98.0
	LMA-D	97.5	99.0	100.0	100.0
CNN MorGAN-D	MorGAN-S	0.0	1.0	2.5	4.0
	MorGAN-R	0.5	7.0	18.5	32.5
	LMA-S	8.0	38.5	59.5	75.0
	LMA-R	0.0	0.0	0.0	0.5
	LMA-D	0.0	0.0	0.0	0.0

Table 2: APCER values of unknown attacks achieved at the detection decision thresholds that produced certain fixed APCER of known attacks using FV approach with CNN-based features.

Train	Test	APCER[%] (known attack) 1	10	20	30
LBPH LMA-S	LMA-R	1.5	7.5	18.5	25.0
	LMA-D	3.5	7.0	17.5	26.0
	MorGAN-S	15.5	34.5	47.0	61.5
	MorGAN-R	10.5	33.5	47.5	56.5
	MorGAN-D	10.5	30.5	43.5	58.5
LBPH LMA-R	LMA-S	1.0	10.5	22.0	34.0
	LMA-D	0.5	5.5	12.0	18.0
	MorGAN-S	6.0	31.0	45.0	57.0
	MorGAN-R	1.0	28.5	40.0	53.0
	MorGAN-D	3.5	30.0	47.0	66.5
LBPH LMA-D	LMA-S	1.0	22.5	38.0	49.0
	LMA-R	1.5	14.5	26.0	33.5
	MorGAN-S	8.5	47.0	62.5	71.0
	MorGAN-R	2.0	30.5	45.0	53.0
	MorGAN-D	4.5	38.0	52.0	64.5
LBPH MorGAN-S	MorGAN-R	1.0	6.5	14.0	28.5
	MorGAN-D	2.5	17.5	28.5	40.0
	LMA-S	67.0	92.5	98.0	98.5
	LMA-R	55.5	92.0	98.0	99.0
	LMA-D	50.0	89.0	96.0	99.0
LBPH MorGAN-R	MorGAN-S	1.0	17.0	26.0	36.0
	MorGAN-D	2.5	24.0	33.0	44.0
	LMA-S	62.5	96.5	98.5	99.0
	LMA-R	62.0	94.5	99.5	99.5
	LMA-D	55.5	90.5	95.5	97.0
LBPH MorGAN-D	MorGAN-S	2.5	10.0	19.0	23.5
	MorGAN-R	2.5	5.5	15.5	20.5
	LMA-S	77.5	93.0	98.0	98.5
	LMA-R	79.5	92.5	97.5	98.5
	LMA-D	66.0	87.5	92.5	94.5

Table 3: APCER values of unknown attacks achieved at the detection decision thresholds that produced certain fixed APCER of known attacks using FV approach with LBPH-based features.

and thus perform better on unknown attacks of both attack types and pairing protocols when trained on LMA-D, and only on pairing protocols when trained on MorGAN-D.

We analyze the **feature differential pattern** (DFV) approach to detect known and unknown attacks. Using this approach along with the CNN-based features performed poorly on known attacks (Table 4), and therefore their generalization ability on very high BPCER is irrelevant. Only the CNN-based LMA-R setting achieve high detection rate on known attacks, probably because of the training data diversity. However, it again fails to generalize on attack type and pairing protocols variations (Table 5). The DFV approach performs almost perfectly on all known attack when using LBPH features (Table 4), this can be caused by the non-learned nature of these features and thus the focus on differences in general image properties rather than differences in face identities (what OpenFace is trained to do).

	Train/Test	APCER[%] 1	10	20	30
CNN	LMA-S	97.3	91.3	84.0	79.3
	LMA-R	0.3	0.0	0.0	0.0
	LMA-D	100.0	100.0	99.7	98.0
	MorGAN-S	100.0	100.0	100.0	100.0
	MorGAN-R	99.3	88.0	82.3	75.7
	MorGAN-D	99.7	89.3	74.6	59.2
LBPH	LMA-S	0.0	0.0	0.0	0.0
	LMA-R	0.0	0.0	0.0	0.0
	LMA-D	0.0	0.0	0.0	0.0
	MorGAN-S	0.0	0.0	0.0	0.0
	MorGAN-R	0.0	0.0	0.0	0.0
	MorGAN-D	0.0	0.0	0.0	0.0

Table 4: BPCER rates achieved at fixed APCER rates of known attacks using the DFV detection approach with both feature types.

In this DFV-LBPH setting, when the detector is trained on LMA attacks, the generalization ability on unknown Mor-

Train	Test	APCER[%] (known attack) 1	10	20	30
CNN LMA-S	LMA-R	71.8	90.5	97.0	98.5
	LMA-D	100.0	100.0	100.0	100.0
	MorGAN-S	100.0	100.0	100.0	100.0
	MorGAN-R	76.2	91.2	95.5	97.8
	MorGAN-D	82.7	98.2	100.0	100.0
CNN LMA-R	LMA-S	87.2	98.8	99.8	100.0
	LMA-D	85.0	98.2	99.5	100.0
	MorGAN-S	53.0	87.2	93.2	95.8
	MorGAN-R	98.8	100.0	100.0	100.0
	MorGAN-D	3.8	24.8	43.1	59.9
CNN LMA-D	LMA-S	27.8	79.8	87.8	90.8
	LMA-R	13.8	60.2	70.0	77.0
	MorGAN-S	28.0	76.8	83.0	88.2
	MorGAN-R	0.0	2.2	6.8	15.8
	MorGAN-D	0.0	0.0	0.0	0.0
CNN MorGAN-S	MorGAN-R	0.0	0.0	0.0	0.0
	MorGAN-D	0.0	0.8	4.3	7.3
	LMA-S	0.0	0.0	0.0	0.0
	LMA-R	1.5	16.0	26.0	37.2
	LMA-D	8.5	44.6	61.2	71.9
CNN MorGAN-R	MorGAN-S	0.0	4.5	10.5	15.8
	MorGAN-D	0.0	0.0	0.0	0.0
	LMA-S	0.2	11.2	21.0	30.5
	LMA-R	0.0	0.0	0.0	0.0
	LMA-D	0.0	7.3	13.5	20.6
CNN MorGAN-D	MorGAN-S	92.2	99.2	99.8	100.0
	MorGAN-R	0.0	8.2	24.2	38.8
	LMA-S	100.0	100.0	100.0	100.0
	LMA-R	94.2	99.2	99.8	99.8
	LMA-D	75.4	98.7	98.7	99.7

Table 5: APCER values of unknown attacks achieved at the detection decision thresholds that produced certain fixed APCER of known attacks using DFV approach with CNN-based features.

Train	Test	APCER[%] (known attack) 1	10	20	30
LBPH LMA-S	LMA-R	60.2	94.2	98.0	99.0
	LMA-D	1.3	6.8	14.3	21.3
	MorGAN-S	100.0	100.0	100.0	100.0
	MorGAN-R	100.0	100.0	100.0	100.0
	MorGAN-D	100.0	100.0	100.0	100.0
LBPH LMA-R	LMA-S	46.5	94.5	97.8	99.8
	LMA-D	11.3	64.9	77.7	87.5
	MorGAN-S	100.0	100.0	100.0	100.0
	MorGAN-R	100.0	100.0	100.0	100.0
	MorGAN-D	100.0	100.0	100.0	100.0
LBPH LMA-D	LMA-S	68.0	96.0	99.0	100.0
	LMA-R	4.8	28.5	46.2	63.5
	MorGAN-S	100.0	100.0	100.0	100.0
	MorGAN-R	100.0	100.0	100.0	100.0
	MorGAN-D	100.0	100.0	100.0	100.0
LBPH MorGAN-S	MorGAN-R	0.0	0.5	1.5	2.5
	MorGAN-D	0.0	0.0	0.0	0.0
	LMA-S	99.0	100.0	100.0	100.0
	LMA-R	100.0	100.0	100.0	100.0
	LMA-D	98.7	100.0	100.0	100.0
LBPH MorGAN-R	MorGAN-S	55.2	89.8	94.5	96.8
	MorGAN-D	93.2	99.7	100.0	100.0
	LMA-S	98.0	100.0	100.0	100.0
	LMA-R	21.0	61.8	81.5	92.8
	LMA-D	100.0	100.0	100.0	100.0
LBPH MorGAN-D	MorGAN-S	98.0	99.8	100.0	100.0
	MorGAN-R	0.0	0.0	0.0	0.0
	LMA-S	100.0	100.0	100.0	100.0
	LMA-R	0.0	0.0	0.2	0.2
	LMA-D	16.0	43.9	67.4	78.7

Table 6: APCER values of unknown attacks achieved at the detection decision thresholds that produced certain fixed APCER of known attacks using DFV approach with LBPH-based features.

GAN attacks is poor (Table 6). This is slightly better on LMA attacks of different pairing protocols, with relatively high errors on unknown LMA-S and LMA-R attacks (Table 6). Training the DFV-LBPH approach on MorGAN attacks produces more generalized solution in comparison to training on LMA attacks. Training on DFV-LBPH-MorGAN-S performs good on unknown MorGAN attacks of different pairing protocols, but fails with LMA attacks. The same approach trained on MorGAN-D performs good even on LMA-R, LMAD and MorGAN-R, but fails on the pairing protocol S in both attack types (Table 6). Generally, the DFV approach generalize poorly on unknown attacks of different types and protocols. Also here, although not optimal, choosing to train on attacks created by the pairing protocol D is relatively more generalizable on unknown attacks. However, in the DFV scenario, training on MorGAN attacks produces better overall results in comparison to LMA.

6. Conclusion

This work analyzed the effect of the pairing protocols on the generalization of morphing attack detection solutions. We presented a novel database with variations in pairing protocols and morphing techniques. We analyzed the known and unknown attack detection performance of both, single image and differential, detection approaches based on handcrafted and CNN features. We stated the generalization vulnerability of attack detection to the pairing protocols and stated recommendations for training such an attack detection system. For example, we found out that training a detector on dissimilar pairs can result in a more generalized solution over morphing and pairing protocols.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and

the Arts (HMWK) within CRISP.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Czech Republic, May, 2004. Proceedings, Part I*, volume 3021 of *LNCS*, pages 469–481. Springer, 2004.

[2] 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.

[3] N. Damer, V. Boller, Y. Wainakh, F. Boutros, P. Terhörst, A. Braun, and A. Kuijper. Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In *Pattern Recognition - 40th German Conference, GCPR 2018, Stuttgart, Germany, October 10-12, 2018, Proceedings*, LNCS. Springer, 2018.

[4] N. Damer and K. Dimitrov. Practical view on face presentation attack detection. In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*. BMVA Press, 2016.

[5] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper. MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Redondo Beach, CA, USA, October 22-25, 2018*. IEEE, 2018.

[6] N. Damer, Y. Wainakh, O. Henniger, C. Croll, B. Berthe, A. Braun, and A. Kuijper. Deep learning-based face recognition and the robustness to perspective distortion. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20-24, 2018*, pages 3445–3450. IEEE Computer Society, 2018.

[7] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics, Clearwater, IJCB 2014, FL, USA, September 29 - October 2, 2014*, pages 1–7. IEEE, 2014.

[8] M. Ferrara, A. Franco, and D. Maltoni. On the effects of image alterations on face recognition accuracy. In T. Bourlai, editor, *Face Recognition Across the Imaging Spectrum*, pages 195–222. Springer, 2016.

[9] International Civil Aviation Organisation (ICAO). ICAO Draft Technical Report: Portrait quality (reference facial images for MRTD). technical report, Version 0.9, 2017.

[10] International Organization for Standardization. ISO/IEC DIS 30107-3:2016: Information Technology Biometric presentation attack detection P. 3: Testing and reporting, 2017.

[11] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.

[12] D.-T. Lee and B. J. Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242, 1980.

[13] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[14] S. Mallick. Face morph using opencv c++ / python. https://www.learnopencv.com/face-morph-using-opencv-cpp-python/, 2016.

[15] Markets and Markets. Facial Recognition Market by Component (Software Tools and Services), Technology, Use Case (Emotion Recognition, Attendance Tracking and Monitoring, Access Control, Law Enforcement), End-User, and Region - Global Forecast to 2022. Report, November 2017.

[16] T. Neubert. Face morphing detection: An approach based on image degradation analysis. In *Digital Forensics and Watermarking - 16th International Workshop, IWDW 2017, Magdeburg, Germany, August 23-25, 2017, Proceedings*, volume 10431 of *LNCS*, pages 93–106. Springer, 2017.

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, September 7-10, 2015*, pages 41.1–41.12. BMVA Press, 2015.

[18] R. Ramachandra, K. B. Raja, and C. Busch. Detecting morphed face images. In *8th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2016, NY, USA, September, 2016*, pages 1–7. IEEE, 2016.

[19] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch. Face morphing versus face averaging: Vulnerability and detection. In *2017 IEEE International Joint Conference on Biometrics, IJCB 2017, Denver, CO, USA, October 1-4, 2017*, pages 555–563. IEEE, 2017.

[20] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21-26, 2017*, pages 1822–1830. IEEE Computer Society, 2017.

[21] D. J. Robertson, R. S. S. Kramer, and A. M. Burton. Fraudulent id using face morphs: Experiments on human and automatic recognition. *PLOS ONE*, 12(3):1–12, 03 2017.

[22] U. Scherhag, D. Budhrani, M. Gomez-Barrero, and C. Busch. Detecting morphed face images using facial landmarks. In *Image and Signal Processing - 8th International Conference, ICISP 2018, France, July, 2018, Proceedings*, volume 10884 of *LNCS*, pages 444–452. Springer, 2018.

[23] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. J. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *International Conference of the Biometrics Special Interest Group, BIOSIG 2017, Darmstadt, Germany, September 20-22, 2017*. GI / IEEE, 2017.

[24] U. Scherhag, C. Rathgeb, and C. Busch. Performance variation of morphed face image detection algorithms across different datasets. In *2018 I. W. on Biometrics and Forensics, IWBF 2018, Italy, June, 2018*, pages 1–6. IEEE, 2018.

[25] U. Scherhag, C. Rathgeb, and C. Busch. Towards detection of morphed face images in electronic travel documents. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 187–192. IEEE Computer Society, 2018.