A Principal Component Analysis-Based Approach for Single Morphing Attack Detection

Laurine Dargaud
Technical University of Denmark (DTU)
Kongens Lyngby, Denmark
s212803@dtu.dk, dargaud.laurine@gmail.com

Mathias Ibsen, Juan Tapia and Christoph Busch
Hochschule Darmstadt (hda), Germany
da/sec-Biometrics and Internet Security Research Group
mathias.ibsen,juan.tapia-farias,christoph.busch@h-da.de

Abstract

This paper proposes an explicit method for single face image morphing attack detection, using an RGB decomposition based on Principal Component Analysis from texture patterns. Handcrafted detection algorithms can be advantageous over deep learning-based methods as they constitute increased explainability, showcased in this work by visualizing relevant face areas for morphing attack detection. Such information can be relevant for deployed systems in real-world scenarios with humans in the loop. The morphing detection capability of the proposed method is evaluated extensively across three datasets and six morphing algorithms in single, cross-dataset and cross-morphed scenarios and compared to a fine-tuned MobileNetV2 architecture. The results show how single image morphing attack detection remains challenging, especially in cross-domain scenarios involving realistic diversity of morphing algorithms, including StyleGAN-based approaches. In such conditions, the proposed method can be as good or even better than the evaluated MobileNetV2 approach.

1. Introduction

Biometric systems observe biological or physiological characteristics to recognize individuals. Such systems are used in various security applications by governmental, industrial and private institutions. Face recognition systems are often deployed among the different types of biometric systems due to their high convenience and security. Many countries have, for instance, adopted Automatic Border Control (ABC) gates in combination with electronic Machine Readable Travel Documents (eMRTD), which use face recognition technologies for automatic face verification. However, despite the popularity of face recognition systems, it has been shown that they are vulnerable to digital manipulations and physical presentation attacks [20, 9, 28]. A specific type of digital attack is based on the use of morphed images, in which the facial images of two or more individuals are merged into a single image. In [7], the authors were the first to show that if a morphed image is stored in a passport, it can potentially be used by all the individuals contributing to the morphed image for circumventing the security of an automated face recognition system. To mitigate the vulnerability of face recognition systems with regard to morphing attacks, several algorithms for automated face Morphing Attack Detection (MAD) have been proposed [28]. However, it remains a challenging problem, as MAD has not yet achieved acceptable biometric performance at operationally-relevant false detection error rates [16].

There are two scenarios for detecting morphed images in an operational system: single image morphing attack detection (S-MAD) and differential morphing attack detection (D-MAD). In S-MAD, a single suspected image of the subject is analysed by the system to detect morphing attacks. In D-MAD, the system compares the subject's facial image with a reference image. This present work focuses on the more challenging S-MAD, which has the advantage over D-MAD of not requiring reliable live capture, and thus could serve also the needs of a forensic investigation.

The proposed method relies on Principal Component Analysis (PCA) performed on Local Binary Patterns (LPB) of images' individual RGB channels to extract features which are subsequently merged and used to train a classifier for detecting morphed images. The detection performance is evaluated on three different databases, in three experimental protocols, and compared to a fine-tuned MobileNetV2 architecture [24].

The main contributions of this proposal are:

- An explicit S-MAD method based on RGB decomposition, texture features, and PCA.
- An extensive benchmark of the proposed S-MAD method and MobileNetV2 in single, cross-dataset and cross-morphed evaluation protocols.
- A visualization method in a reduced vector space, to show the most relevant areas on faces for bona fide and morphed images, to get more explainable methods.

The rest of the paper is organized as follows: Section 2 describes relevant related work, Section 3 outlines the proposed method for detecting morphed images based on PCA features extracted from individual RGB channels. After that, the metrics and databases used in this work are described in Section 4 and 5, respectively. Subsequently, Section 6 describes the three sets of conducted experiments, corresponding results and relevant visualizations. Lastly, Section 7 concludes with a summary of the work carried out and the obtained results.

2. Related work

S-MAD can be performed on different kinds of image features that can be classified into five main categories: texture descriptors, gradient-based descriptors, key point descriptors, image forensics approaches, and the use of deep-learning [25].

The texture is considered one of the most important image characteristics, since the analysis of face textures can be used to support fundamental image processing tasks for morphing attack detection. According to a survey by Venkatesh et al., one of the first texture features-based approaches was presented by Raghavendra et al. [21], who worked with Binarized Statistical Image Feature (BSIF). Other examples include Local Binary Patterns (LPB) [30] and Local Phase Quantization (LPQ) [22].

Regarding gradient-based features, where focus is more on analysing the changes of information, Histogram of Oriented Gradients (HOG) was applied in [27] for S-MAD.

For key point descriptors, the main idea is to identify points of interest in the image and explicitly analyse the surrounding area [25]. Common key point descriptors are Scale-Invariant Feature Transform (SIFT) [14] and Speeded-Up Robust Features (SURF) [13].

Some studies suggest applying image forensics techniques to detect the origin of image manipulation. They focus on noise patterns by analysing pixel discontinuities that may be impacted by morphing algorithms – like Photo Response Non-Uniformity (PRNU) [26] and Sensor Pattern Noise (SPN) [34], or on image quality by quantifying image degradation of artefacts in morphed faces [13].

Tapia et al. [31] proposed to add an extra stage of feature selection after feature extraction of LBP, HOG and Raw images based on Mutual Information. Since high redundancy between features confuses the classifier, they identify the most relevant features, and remove the most redundant ones from the feature vector, to better separate bona fide and morphed images in an S-MAD scenario. The authors also conclude that eyes and nose are the most relevant facial areas.

Eventually, the emergence and the constant progress of deep learning methods have also been observed in S-MAD research: deep Convolutional Neural Networks (CNNs) such as VGG19, AlexNet, GoogLeNet or ResNet have been used in previous works to detect morphs [23] [29].

Although deep leaning-based methods usually outperform approaches based on handcrafted features when enough training data is available, the learned features and decision outcome is difficult to interpret. To advance more explainable methods, this work explores a method based on texture analysis, PCA and colour decomposition. In the early days, PCA has been used for face recognition by determining eigenfaces, e.g. principal image components that define the latent space of faces [15]. Even though PCA has been used for morphed face generation to improve GAN-based morphing [19], no application of PCA in S-MAD was found. Decomposing an image into separate colour channels and analysing each channel individually can help capture more relevant information. Venkatesh et al. [32] experimented with it by analysing the noise residuals of each Hue, Saturation and Value (HSV) colour channel. Raghavendra et al. [22] also proposed an S-MAD method, based on both HSV and YCrCb colour spaces decomposition: they extract LBP features for each colour channel independently, then they classify on the concatenated histogram of extracted features.

3. Proposed PCA-based method for S-MAD

The method proposed in this paper applies colour decomposition to an RGB image and, for each colour channel, extracts features using LBP. Subsequently, PCA is performed and data is projected in the PC-space of each colour channel. Finally, the three resulting projections are concatenated to train a classifier for distinguishing between morphed and bona fide images. Figure 1 shows a block diagram of the proposed method. In this section, each phase of the proposed method is detailed.

3.1. Preprocessing

In the preprocessing phase, detection, alignment, and cropping are applied to a facial image to obtain a 500×500 image focused around the facial region. In contrast with

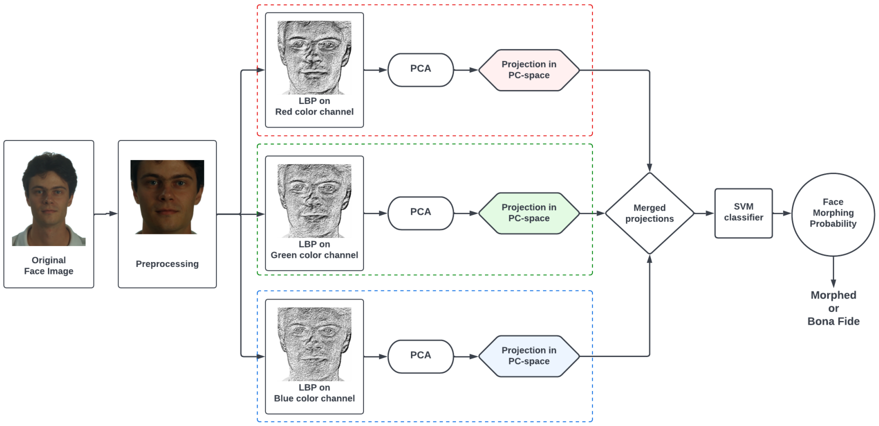

Figure 1. Block diagram of the proposed method for S-MAD.

several other S-MAD systems [21] [30] which use the Viola-Jones algorithm [33] for face detection, an implementation of Dlib [12] is used in this work. Indeed, it captures a larger area, by taking parts of forehead, chin and hair into account, which are likely to be affected by the morphing process. Additionally, each colour channel of the image is normalized to cover the full range of pixel values, from 0 to 255.

3.2. Feature extraction

To extract features from a face image, the image is first decomposed in the RGB colour space, forming a separate 2D array for each colour channel. The RGB colour space was selected in this work after analysing 3 common representations used in image analysis: Red-Green-Blue (RGB), Hue-Saturation-Value (HSV) and YCrCb. By default, a coloured image is defined in RGB. After comparing the efficiency of each configuration on individual dataset experiments, the RGB decomposition was found to be the best colour space configuration for the proposed method.

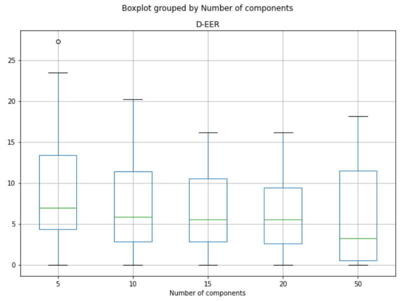

Figure 2. Box plot with the optimal number of PCA components. Results come from individual dataset experiments.

After the colour space decomposition, each colour channel can be independently analysed in terms of anomalies and artefacts, which can be used to detect the morphed images. First, LBP, using a 3×3 matrix without patch averaging, is applied to each colour channel independently to extract texture information. Then, images are flattened as 1D-vectors, and scaled by removing the mean and scaling to unit variance by using a scaler. Subsequently, PCA is performed, which reduces the feature dimensionality of the extracted LBP features. In this work, the PCA is applied resulting in 50 components, for each colour channel. The choice was made by analysing the Detection Equal Error Rate (D-EER) across a range of different numbers of components, as shown in Figure 2. Afterwards, the PCA features of the three colour channels are merged together to form the final feature vector.

3.3. Classification

To determine the probability that the input image is a morph, four different classifiers were tested: 1) Gaussian Process-based Bayesian classifier, with a radial basis function kernel of length scale 1.0. 2) K-Nearest Neighbours classifier, with $K = 5$ neighbours and uniform weights. 3) Logistic Regression, with a regularization strength of 1.0, a Limited-BFGS solver and a random state of 42 and 4) C-Support Vector Machine classifier, with a radial basis function kernel, and a regularization parameter of $C = 1.0$.

By performing individual dataset experiments, the C-Support Vector Machine classifier was found to provide the best performance.

4. Metrics

To evaluate the performance of the proposed method, metrics in compliance with the International Standard ISO/IEC 30107-1 [10] are used. The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of morphed faces incorrectly classified as bona fide faces, and the Bona fide Presentation Classification Error Rate (BPCER) as the proportion of bona fide face images

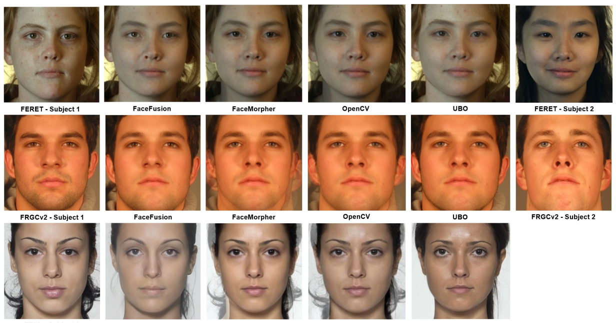

Figure 3. Examples of studied morphing algorithms for two subjects. Top: FERET. Middle: FRGCv2. Bottom: FRLL.

incorrectly classified as being morphed. In this paper, performance is evaluated by reporting the Detection Equal Error Rate (D-EER), which is the error rate when APCER and BPCER are equal. Additionally, plots of Detection Error Tradeoff (DET) curves are provided, which display the tradeoff between APCER and BPCER at different operating points.

5. Databases

In this study, three different databases of frontal faces are used: the Facial Recognition Technology database (FERET), the Face Recognition Grand Challenge database (FRGCv2) and the Face Research London Lab (FRLL). The morphed images in these datasets have been created using a morphing factor of 0.5, meaning both parent images contribute equally to the morphed image. FERET and FRGCv2 morphed images are generated with the four same morphing tools, i.e. UBO, FaceFusion, FaceMorpher and OpenCV. FRLL morphed images are generated with two other morphing tools i.e. StyleGAN and WebMorph, and another one in common with FERET and FRGC, i.e. FaceMorpher. A summary of the datasets is provided in Table 1. Bona fide subjects and morphs examples are displayed in Figure 3.

Table 1. Number of bona fide and morphed images used for FERET, FRGCv2 and FRLL databases. '-' refers to the case where a morphing tool is not applied to the given database.

Morphing Tool	Bona fide images (FERET/FRGC/FRLL)	Morph images (FERET/FRGCv2/FRLL)
UBO	622 / 1,440 / -	529 / 964 / -
FaceFusion	622 / 1,440 / -	529 / 964 / -
FaceMorpher	622 / 1,440 / 102	529 / 964 / 103
OpenCV	622 / 1,440 / -	529 / 964 / -
StyleGAN	- / - / 102	- / - / 270
WebMorph	- / - / 102	- / - / 212

5.1. FERET dataset

FERET dataset is a subset of the Colour FERET Database [18], generated in the context of the Facial Recognition Technology (FERET) program technically handled by the National Institute of Standards and Technology (NIST). It contains 622 bona fide and 529 morphs generated with four different morphing algorithms:

- **UBO** [1, 8]: developed at the University of Bologna, this morphing tool matches the facial landmarks of the two subjects, then it combines, averages and blends borders to generate a morph.
- **FaceFusion** [3] : this proprietary mobile application developed by MOMENT generates very realistic faces, since morphing artefacts are almost invisible.
- **FaceMorpher** [4]: this open-source Python implementation relies on STASM, a facial features finding package, for landmark detection, but generated morphs show many artefacts which make them more recognizable.
- **OpenCV** [6]: this open-source morphing algorithm is quite similar to FaceMorpher method, but it uses Dlib to detect face landmarks. Again, some artefacts remains in generated morphs.

5.2. FRGC dataset

The FRGC dataset used in this work is a constrained subset of the second version of the Face Recognition Grand Challenge (FRGCv2) dataset [17]. It contains 1,440 bona fide, and 964 morphs generated with the same four different morphing algorithms as used for the FERET dataset, i.e. UBO, FaceFusion, FaceMorpher, and OpenCV.

5.3. FRLL dataset

The FRLL dataset is a subset of the publicly available Face Research London Lab (FRLL) dataset [2]. It contains 102 bona fide. Three morphing algorithms were applied to obtain 103 morphs from FaceMorpher algorithm, 270 morphs from StyleGAN algorithm [11], and 212 morphs from WebMorph algorithm [5].

6. Experiments and Results

To evaluate the performance of the S-MAD method, three protocols are used: intra, cross-database and cross-morphed experiments. In experiment-A (Section 6.1), datasets are evaluated separately. In experiment-B (Section 6.2), morphing algorithms are evaluated in cross-database experiments, e.g. the applied morphing algorithm remains the same between the training and the test set, but initial database changes. Lastly, in experiment-C (Section 6.3), the performance is investigated when the morphing technique used for testing is unknown during training.

For each experiment, the performance of the proposed method is compared with a Deep Learning method: all parameters of the MobileNetV2 network [24], pre-trained on ImageNet, are fine-tuned. The model was trained on 100 epochs using a batch size of 32 elements. The Cross Entropy Loss is used as criterion, the Adam algorithm with a learning rate of $1e^{-4}$ as optimizer, and the cosine annealing schedule as learning rate scheduler.

The input for machine learning classifiers is a vector of size 1×150, representing the concatenation of the 50 PCA components for each colour channel (R, G, B). For MobileNetV2, the input is an RGB image of size $224 \times 224 \times 3$.

6.1. Experiment A: Individual dataset

For each experiment, datasets are created with all bona fide of a given database and a group of morphs based on the same database.

To build a dataset, a group of bona fide pictures is gathered with a group of morphed pictures, then randomly split into 3 sets: 60% as a training set, 20% as a validation set (used in the MobileNetV2 training phase) and 20% as a test set. To ensure a fair training, split is performed so that resulting datasets remain balanced between bona fide and morphed images.

D-EER performance results on the test set of each database are shown separately for the different morphing algorithms in Table 2.

Table 2. Performance results in terms of D-EER (in %) on experiment-A (individual datasets) for the proposed PCA-based method and the MobileNetV2 based method. The best result for each experiment is marked in bold.

Database	Morphs	PCA D-EER (%)	MobileNetV2 D-EER (%)
FERET	UBO	18.18%	**0.00%**
FERET	FaceFusion	17.12%	**1.85%**
FERET	FaceMorpher	11.32%	**1.00%**
FERET	OpenCV	11.71%	**0.93%**
FRGCv2	UBO	2.55%	**0.00%**
FRGCv2	FaceFusion	3.21%	**0.51%**
FRGCv2	FaceMorpher	**0.00%**	**0.00%**
FRGCv2	OpenCV	1.06%	**0.00%**
FRLL	StyleGAN	**0.00%**	28.00%
FRLL	WebMorph	**0.00%**	**0.00%**
FRLL	FaceMorpher	11.11%	**0.00%**

The results show that, generally, the CNN-based method performs better than the handcrafted PCA-based method, except for the recognition of StyleGAN morphs in the FRLL database. Indeed, in this case, the PCA-based method perfectly completes the classification task, while MobileNetV2 approach struggles clearly more. Nevertheless, this observation is inverted in FaceMorpher morphs detection for FRLL database. This shows that different approaches must be created to generalize to both, namely landmark-based methods and synthetic-based morph images. Also, the classification task performs better on FRGCv2 than on the FERET database for both methods. Besides, Scherhag [25] states that the comparison score distributions of the more complex morphing algorithms (FaceFusion and UBO) are consistently closer to the mated comparison score distributions than the comparison score distributions of the more basic morphing algorithms (FaceMorpher and OpenCV). This idea is confirmed since, in most cases, FaceFusion and UBO morphed faces lead to a larger detection error than OpenCV and FaceMorpher.

6.2. Experiment B: Cross-database

In cross-database experiments, training and test are performed on morphs generated with the same morphing algorithm but on different databases. This is a more challenging scenario than experiment-A, as the algorithms must learn to generalize to images acquired under different conditions.

In dataset construction, the whole data is considered as a test set, but training is still performed on the defined training dataset, e.g. 60% of the whole data.

The D-EER performance results are provided in Table 3.

Table 3. Performance results in terms of D-EER (in %) on experiment-B (cross-database) for the proposed PCA-based method and the MobileNetV2 based method. The best result for each experiment is marked in bold.

Training Database	Morphs	Test Database	PCA D-EER (%)	MobileNetv2 D-EER (%)
FERET	UBO	FRGCv2	9.65%	**3.42%**
FERET	FaceFusion	FRGCv2	12.24%	**4.56%**
FERET	FaceMorpher	FRGCv2	4.77%	**0.93%**
FERET	OpenCV	FRGCv2	6.43%	**0.73%**
FERET	FaceMorpher	FRLL	2.91%	**0.00%**
FRGCv2	UBO	FERET	24.20%	**15.12%**
FRGCv2	FaceFusion	FERET	25.14%	**17.58%**
FRGCv2	FaceMorpher	FERET	20.60%	**13.04%**
FRGCv2	OpenCV	FERET	22.68%	**14.74%**
FRGCv2	FaceMorpher	FRLL	7.77%	**0.00%**
FRLL	FaceMorpher	FERET	**20.42%**	21.17 %
FRLL	FaceMorpher	FRGCv2	**12.45%**	14.11%

The results show that in cross-database experiments, e.g. when training is performed on FERET or FRGCv2 database, the CNN-based approach performs better than the hand-crafted PCA-based method.

Nonetheless, when training is performed on the FRLL database, the PCA-based method is more performant than MobileNetV2. However, the experiments with FRLL as a training set need to be considered cautiously, since the corresponding training dataset is quite small. This is due to the few bona fide images for this dataset, which constrains the number of images in each class as the experiments are conducted on balanced classes. Hence, the FRLL dataset is smaller than FERET or FRGCv2 dataset.

6.3. Experiment C: Cross-morphed

In the cross-morphed experiments, morphs are generated with a different morphing algorithm between the training and test phase. As such, training and testing databases are not necessarily the same; making it a harder scenario than the previous single and cross-database experiments.

For the specific case where training and testing are performed on the same database but on different morphing algorithms, the test set has to be built in a specific manner since both share the same bona fide group. Let $D^{(1)}$ be

Table 4. Performance results in terms of D-EER (in %) on Experiment C (cross-morphed) for the proposed PCA-based method and the MobileNetV2 based method. The best result for each experiment is highlighted in bold.

Training		**Test**		**Performance (%)**	
Database	**Morphs**	**Database**	**Morphs**	**PCA D-EER**	**MobileNetV2 D-EER**
FERET	UBO	FERET	FaceFusion	16.22%	**0.90%**
FERET	UBO	FRGC	FaceFusion	12.34%	**6.95%**
FERET	UBO	FRLL	StyleGAN	34.07%	**18.89%**
FERET	FaceFusion	FERET	UBO	15.15%	**0.00%**
FERET	FaceFusion	FRGC	UBO	11.41%	**4.05%**
FERET	FaceFusion	FRLL	StyleGAN	38.52%	**24.44%**
FRGC	FaceMorpher	FRLL	StyleGAN	**33.33%**	50.37%
FRGC	UBO	FRLL	StyleGAN	**32.59%**	47.04%
FRGC	UBO	FRGC	FaceFusion	3.21%	**0.00%**
FRGC	UBO	FERET	FaceFusion	26.65%	**21.74%**
FRGC	FaceFusion	FRGC	UBO	1.53%	**0.00%**
FRGC	FaceFusion	FERET	UBO	25.52%	**16.63%**
FRGC	FaceFusion	FRLL	StyleGAN	**30.37%**	42.22%
FRLL	**StyleGAN**	FRLL	WebMorph	26.83%	**14.63%**
FRLL	**StyleGAN**	FERET	FaceFusion	**33.84%**	38.94%
FRLL	**StyleGAN**	FRGC	FaceFusion	**22.82%**	33.61%
FRLL	WebMorph	FRLL	StyleGAN	26.83%	**24.39%**

the dataset that gathers bona fide pictures of the database D and the morphed pictures generated with the morphing algorithm 1 on the database D. In the same way, let $D^{(2)}$ be the dataset that gathers bona fide pictures of the database D and the morphed pictures generated with the morphing algorithm 2 on the database D. Thus, $D^{(1)}$ and $D^{(2)}$ can be described as follows:

$$D^{(1)} = \begin{cases} D^{(1)}_{\text{train}} = \text{Bona fide}^{(1)}_{\text{train}} + \text{Morphs}^{(1)}_{\text{train}} \\ D^{(1)}_{\text{validation}} = \text{Bona fide}^{(1)}_{\text{validation}} + \text{Morphs}^{(1)}_{\text{validation}} \\ D^{(1)}_{\text{test}} = \text{Bona fide}^{(1)}_{\text{test}} + \text{Morphs}^{(1)}_{\text{test}} \end{cases}$$

$$D^{(2)} = \begin{cases} D^{(2)}_{\text{train}} = \text{Bona fide}^{(2)}_{\text{train}} + \text{Morphs}^{(2)}_{\text{train}} \\ D^{(2)}_{\text{validation}} = \text{Bona fide}^{(2)}_{\text{validation}} + \text{Morphs}^{(2)}_{\text{validation}} \\ D^{(2)}_{\text{test}} = \text{Bona fide}^{(2)}_{\text{test}} + \text{Morphs}^{(2)}_{\text{test}} \end{cases}$$

Therefore, the test set of a cross-morphed experiment for morphings algorithms 1 and 2, based on the same database D, can be built as follows:

$$D_{test} = \text{Bona fide}^{(1)}_{\text{test}} + \text{Morphs}^{(2)}_{\text{test}}$$

This way, bona fide pictures in the test set are definitely not present in the training set.

The D-EER performance results for the cross-morphed experiments are provided in Table 4.

The results show that both algorithms fail to achieve a robust detection performance across many of the experiments: it achieves unacceptable high D-EER $> 42\%$ across multiple experiments. The results show that especially the StyleGAN-generated morphs are difficult to detect when they are unseen during training.

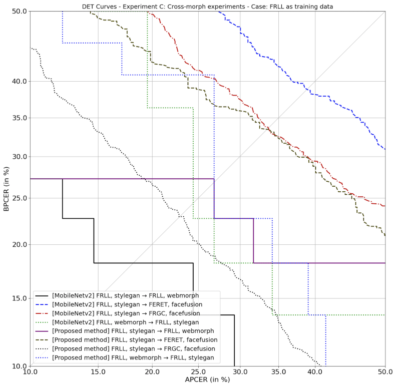

Figure 4. DET showing the performance for FRLL (StyleGAN, WebMorph) when used for training on cross-morphed datasets.

Figure 4 and Figure 5 show the performance in the most challenging scenarios, i.e. when training on FRLL and testing on FRGCv2 and FERET (Figure 4) and oppositely when testing on FRLL (Figure 5). For both approaches, PCA-based and MobileNetV2, the StyleGAN morphing tool reached higher values of EER. This validates the premise that methods developed with StyleGAN-generated images and used for training cannot classify with precision (low EER) landmark-based method in cross-morphed scenarios. The same conclusion can be observed according to Table 4 when the model was trained with a landmark-based morphing tool and tested with StyleGAN-generated images.

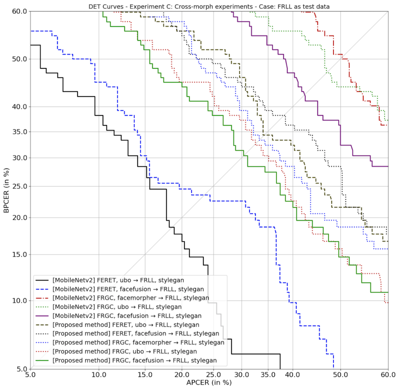

Figure 5. DET showing the performance to classify FRLL when used for test on cross-morphed datasets.

6.4. Visualizations

One strong advantage of the proposed PCA-based method is the capability to visualize and clarify facial regions containing important information. Hence, the proposed method makes it possible to visualize an image space where bona fide and morph can be better differentiated. This can be useful if applied in real-world forensic investigations, where algorithms usually perform in unison with humans.

To illustrate this, some visualizations are provided for Experiment-A, with FRGCv2 database and FaceMorpher morphs. In order to choose the most relevant PCA (Red, Green or Blue channel) and a suitable 2D PC-space for visualization, the most discriminative components are computed.

Let N be the size of the training set. Let $N_c = 50$ be the total number of components. P_n refers to the projection of the n-th image in a given PC-space, with $\{c_1, ..., c_{N_c}\}$ components. Therefore: $P_n = \sum_{i=1}^{N_c} \lambda_{ni} c_i$, with λ_{ni} the coefficients of the n-th image against component i.

To compute how principal components can discriminate bona fide from morphed images, a distance D_i between the two classes for a given component i was defined. Let I_{BF} be the set of the indices of bona fide images from training set, and I_{M} be the set of the indices of morph images from training set. N_{BF} refers to the size of I_{BF}, and N_{M} to the size of I_{M}. Thus, the distance D_i is defined as:

$$D_i = |\frac{1}{N_{\mathrm{BF}}} \sum_{n \in I_{\mathrm{BF}}} \lambda_{ni} - \frac{1}{N_{\mathrm{M}}} \sum_{n' \in I_{\mathrm{M}}} \lambda_{n'i}|$$

The table 5 provides the 5 most discriminative components of Red, Green and Blue colour channel's PCA on LPB, with corresponding distances, for the training set of Experiment A - FRGCv2 with FaceMorpher morphs, and the most discriminative component found is the first component in the Blue channel PCA. Therefore, the 2D PC-space with PC1 and PC2 of the Blue PCA is chosen.

Table 5. Distances between morphs and bona fide, for the 5 most discriminative components of R, G, B colour channel's PCA on LPB, for the training set of Experiment A - FRGCv2 with FaceMorpher morphs. The most discriminative component is marked in bold.

Ranking	**Colour Channel**		
	Red	**Green**	**Blue**
Top-1	PC2: 48.98	PC2: 74.58	**PC1: 96.65**
Top-2	PC5: 38.44	PC1: 50.90	PC2: 20.52
Top-3	PC3: 36.78	PC6: 12.42	PC4: 8.76
Top-4	PC1:35.52	PC7: 8.58	PC5: 5.13
Top-5	PC8:23.79	PC4: 8.34	PC14: 4.79

Figure 6 shows a projection of the two most discriminative principal components (PC1 and PC2) for the blue colour channel. As shown, the two classes can be distinguished, mostly thanks to the first component (PC1).

In order to better see how this first principal component helps to discriminate morphs from bona fide images, comparison score density plots are provided in Figure 7.

Also, it is possible to visualize this first principal component back into the scaled image space. Then the average of all projections of the training set is computed against this first principal component for each class, which is brought back into the scaled image space. This helps to better understand which parts of the LBP image are relevant to distinguish morphs from bona fide face images. In Figure 8, the first component is displayed as a 500 × 500 image in the scaled pixel range, with the average projection against this component, among all bona fide data. Similarly, the same visualization process is applied to all FaceMorpher data from the training set. To obtain these two average projections for the bona fide and morphed images, the values of the image reconstructed from PC1 are multiplied by the average coefficient for this component for the respective class.

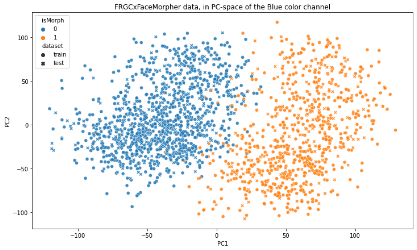

Figure 6. Scatter plot of FRGCv2 versus FaceMorpher individual experiment data, projected onto the two most discriminative components of the Blue colour channel

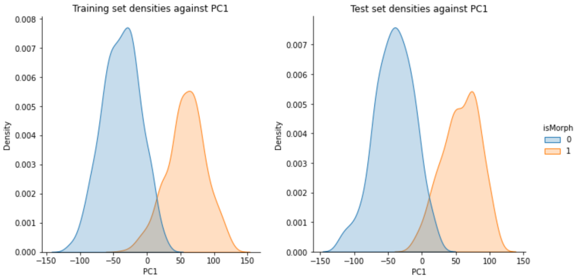

Figure 7. Comparison score density plots of training (left) and test (right) set of FRGCv2 versus FaceMorpher dataset for PC1 score of PCA on blue colour channel.

From this first principal component as an image, it was noticed that relevant areas, e.g. areas with extreme values (red or blue areas), to discriminate morphs from bona fide faces are mostly located around the eyes, nose, lips and cheeks. Red shows areas with a high LBP value, and blue, areas with a low LBP value. A morph face is supposed to be closer to this first component, since the associated coefficient tends to be more likely positive, according to the density plots in Figure 7. In contrast, a bona fide face is supposed to be the opposite of this first component, since the associated coefficient tends to be more likely negative.

These observations can be confirmed by plotting the average LPB face of the bona fide and morph class, in the scaled image space. In Figure 9, the average LBP face for the blue colour channel are shown for the bona fide and FaceMorpher images of the FRGCv2 training set. To obtain these two images, all images of the training set are scaled by applying the same scaler as used before performing PCA. Thereafter, it was computed the average image for all bona

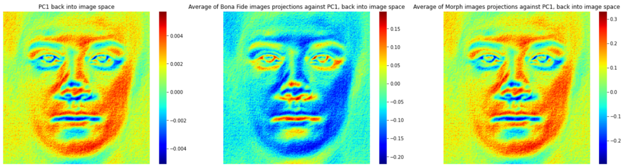

Figure 8. Left: first PC of the blue colour channel's PCA applied on FRGCv2 using FaceMorpher, back into the scaled image space. Middle: average of bona fide LBP projections into PC1 on the blue channel of FRGCv2. Right: average of FaceMorpher LBP projections into PC1 on the blue channel of FRGCv2.

fide faces, then for all FaceMorpher morphed faces. The same process is applied on the bona fide and FaceMorpher morphs images of the test data, as shown in Figure 10. For both training and test data, the pattern of PC1 can be recognized in the average morph face image, whereas it is opposite for the average bona fide face image.

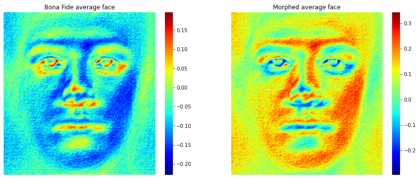

Figure 9. Left: average of scaled bona fide LBP training images on the blue channel of FRGCv2. Right: average of scaled FaceMorpher LBP training images on the blue channel of FRGCv2.

Finally, the comparison score density plots against this first component is plotted, with the average of scaled face images, for another dataset as test set. Thus, this makes it possible to better understand the performance of the PCA-based proposed method in cross-morphed scenario. As a study case, the dataset with StyleGAN morphing tool applied to FRLL database as a test set is studied. The comparison score density plots against PC1, and the two average faces - following the same protocol as for Figure 9 and 10, are presented in Figure 11. The scaler from the FRGCv2 versus FaceMorpher experiment previously studied is used to scale the FRLL versus StyleGAN dataset. Even though bona fide and StyleGAN morphs distributions are less discriminated than in the individual dataset scenario, the PC1 pattern as an image can still be recognized in the average StyleGAN morph LPB face, and its opposite in the average FRLL bona fide LPB face.

7. Conclusion

This work proposed a method for single image morphing attack detection based on features extracted by using Local Binary Patterns and Principal Component Analysis separately for each RGB channel. The generalizability of the PCA-based method was evaluated in intra, cross-database and cross-morphed scenarios and compared to a CNN-based method. The results show that even though MobileNetV2 outperforms the proposed method in most cases, the proposed method performs better on StyleGAN. Overall, both methods still need improvements to detect morphs efficiently, especially across challenging scenarios such as when the morphing technique and dataset are unseen during training. The proposed method has the advantage that it provides increased explainability compared to deep learning-based methods, as the principal components can be visualized back in the image space. However, future work should explore methods for explainable deep learning-based methods, which traditionally are seen as black box models. Additionally, synthetic data can be explored for improving performance in a cross-morphed evaluation scenario.

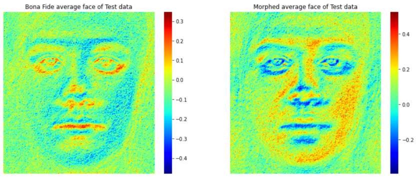

Figure 10. Left: average of scaled bona fide LBP test images for the blue channel of FRGCv2. Right: average of scaled FaceMorpher LBP images for the blue channel of FRGCv2.

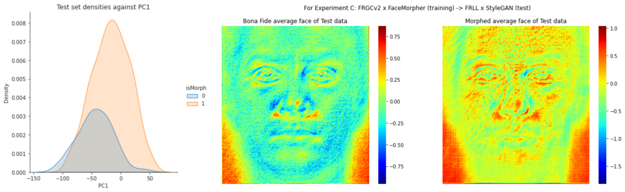

Figure 11. Left: Comparison score density plots of FRLL versus StyleGAN dataset as test set, for PC1 score of PCA on blue colour channel performed on FRGCv2 versus FaceMorpher training set. Middle: average of scaled bona fide LBP pictures on the blue channel of FRLL faces. Right: average of scaled StyleGAN morphs LBP test pictures on the blue channel of FRLL faces.

Acknowledgment

This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 883356 and under the Marie Skłodowska-Curie grant agreement No. 860813 - TReSPAsS-ETN. Additionally, it is supported 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.

References

[1] Biometric system lab - university of bologna. `http://biolab.csr.unibo.it`. Last accessed: 2022-10-05.

[2] Face research lab london set. http://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666/5. L. DeBruine and B. Jones, 2017.

[3] FaceFusion. www.wearemoment.com/FaceFusion. Last accessed: 2022-10-05.

[4] FaceMorpher. https://github.com/yaopang/FaceMorpher/tree/master/facemorpher. Last accessed: 2022-10-05.

[5] Webmorph morphing algorithm, implementation. https://github.com/debruine/webmorph. Last edit: 2021-12-07.

[6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[7] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE Intl. Joint Conf. on Biometrics (IJCB)*, pages 1–7, 2014.

[8] M. Ferrara, A. Franco, and D. Maltoni. Decoupling texture blending and shape warping in face morphing. In *Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*, September 2019.

[9] M. Ibsen, C. Rathgeb, D. Fischer, P. Drozdowski, and C. Busch. Digital face manipulation in biometric systems. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, Advances in Computer Vision and Pattern Recognition, pages 27–43. Springer Verlag, 2022.

[10] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Information Technology - Biometric presentation attack detection - Part 1: Framework*. International Organization for Standardization, 2016.

[11] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. 2020.

[12] D. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 2009.

[13] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann. Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing. In *Proc. Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, pages 21–32, 2017.

[14] D. G. Lowe. Object recognition from local scale-invariant features. In *IEEE Intl. Conf. on Computer Vision (ICCV 1999)*, volume 2, pages 1150–1157, 1999.

[15] Z. Lu, Y. Fu, Y. Qiu, and B. Lu. A new algorithm of improved two-dimensional principal component analysis face recognition. In *33rd Youth Academic Annual Conf. of Chinese Association of Automation (YAC)*, pages 106–111, 2018.

[16] M. Ngan, P. Grother, K. Hanaoka, and J. Kuo. Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection. *National Institute of Standards and Technology (NIST IR 8292)*, September 2022.

[17] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, et al. Overview of the Face Recognition Grand Challenge. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 947–954, 2005.

[18] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[19] S. Price. *Landmark Enforcement and Principal Component Analysis for Improving GAN-Based Morphing*. PhD dissertation, Benjamin M. Statler College of Engineering and Mineral Resources, 2022.

[20] R. Raghavendra and C. Busch. Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Comput. Surv.*, 50(1):1–37, 2017.

[21] R. Raghavendra, K. Raja, and C. Busch. Detecting morphed face images. In *IEEE 8th Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2016.

[22] R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch. Face morphing versus face averaging: Vulnerability and detection. In *Intl. Joint Conf. on Biometrics (IJCB)*, pages 555–563, 2017.

[23] R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch. Transferable deep-CNN features for detecting digital and print-scanned morphed face images. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1822–1830, 2017.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[25] U. Scherhag. *Face Morphing and Morphing Attack Detection*. PhD thesis, Technische Universität Darmstadt, 2020.

[26] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl. Detection of face morphing attacks based on PRNU analysis. *IEEE Trans. on Biometrics, Behavior, and Identity Science (TBIOM)*, 2019.

[27] U. Scherhag, C. Rathgeb, and C. Busch. Performance variation of morphed face image detection algorithms across different datasets. In *6th Intl. Workshop on Biometrics and Forensics*, pages 1–6, 2018.

[28] 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.

[29] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert. Accurate and robust neural networks for face morphing attack detection. *Journal of Information Security and Applications*, Volume 53, 2020.

[30] L. Spreeuwers, M. Schils, and R. Veldhuis. Towards robust evaluation of face morphing detection. In *26th European Signal Processing Conf. (EUSIPCO)*, 2018.

[31] J. E. Tapia and C. Busch. Single morphing attack detection using feature selection and visualization based on mutual information. *IEEE Access*, 9:167628–167641, 2021.

[32] S. Venkatesh, R. Raghavendra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch. Morphed face detection based on deep color residual noise. In *9th Intl. Conf. on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2019.

[33] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, 2001.

[34] L.-B. Zhang, F. Peng, and M. Long. Face morphing detection using fourier spectrum of sensor pattern noise. In *IEEE Intl. Conf. on Multimedia and Expo (ICME)*, July 2018.