Chapter 15
Face Morphing Attack Detection Methods

Ulrich Scherhag, Christian Rathgeb, and Christoph Busch

Abstract Morphing attacks pose a serious threat to face recognition systems, especially in the border control scenario. In order to guarantee a secure operation of face recognition algorithms in the future, it is necessary to be able to reliably detect morphed facial images and thus be able to reject them during enrolment or verification. This chapter provides an overview of morphing attack detection algorithms and metrics to measure and compare their performance. Different concepts of morphing attack detection are introduced and state-of-the-art detection methods are evaluated in a comprehensive cross-database experiments considering various realistic image post-processings.

15.1 Introduction

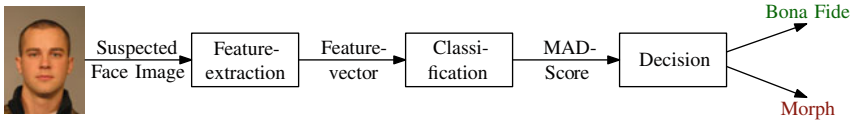
Facial recognition systems have been found vulnerable to Morphing Attacks (MAs). In these attacks, the facial images of two (or more) individuals are combined (morphed) and the resulting morphed facial image is then presented during registration as a biometric reference. If the morphed image is accepted, it is likely that all individuals that contributed to the morphed facial image can be successfully authenticated against it. Morphing attacks thus pose a serious threat to facial recognition systems, in particular in scenarios where the reference image is often provided in printed form by the applicant. The vulnerability of facial recognition systems to face MAs is already well known [5, 29]. Many different approaches for Morphing Attack Detection (MAD)

U. Scherhag
iCOGNIZE GmbH, Dietzenbach, Germany
e-mail: ulrich.scherhag@icognize.de

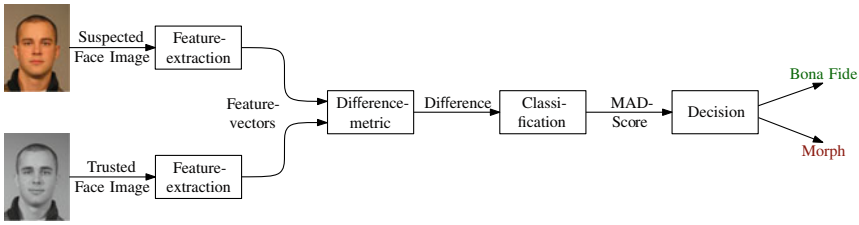
C. Rathgeb (✉)
secunet Security Networks, Essen, Germany
e-mail: christian.rathgeb@secunet.com

C. Busch
Hochschule Darmstadt, Darmstadt, Germany
e-mail: christoph.busch@h-da.de

C. Rathgeb et al. (eds.), *Handbook of Digital Face Manipulation and Detection*,
Advances in Computer Vision and Pattern Recognition,
https://doi.org/10.1007/978-3-030-87664-7_15



(a) no-reference morphing detection scheme



(b) differential morphing detection scheme

Fig. 15.1 Categorisation to no-reference and differential morphing attack detection scheme

have been proposed in the scientific literature. For a comprehensive survey on published morphing attack detection methods the interested reader is referred to [29, 31]. An automated detection of morphed face images is vital to retain the security of operational face recognition systems. According to [25], MAD systems can be divided into two categories: no-reference or single image MAD and reference-based or differential MAD. The corresponding scheme for single image MAD is shown in Fig. 15.1a.

The image to be analysed is passed to the MAD system. First, features are extracted, based on which a classifier decides whether the presented image is a morph or bona fide. The single image MAD scheme can be used during enrolment as well as during verification.

Differential MAD can be used in scenarios where another image, a Trusted Live Capture (TLC), is available in addition to the suspected morph. For example, during verification in an Automated Border Control (ABC) gate, when the probe image is acquired in addition to the extracted reference image from the electronic travel document (suspected morph). The schematic process of differential MAD is depicted in Fig. 15.1b. In general, the same features are extracted from both provided images. These are compared according to a fixed metric and the classifier uses this difference to decide if the suspected morph is a morph or bona fide. This method has the advantage that the additional information of the TLC is used for the decision. However, it should be noted that in real scenarios TLCs are usually acquired in semi-supervised environments, e.g. border gate, and therefore may show a lower quality and higher variance compared to the suspected images.

This bookchapter is organised as follows: Sect. 15.2 briefly discusses related works on MAD. Section 15.3 describes the considered MAD pipeline. The used database is described in Sect. 15.4. MAD methods are presented in Sect. 15.5 and evaluated in Sect. 15.6. Finally, a summary is given in Sect. 15.7.

15.2 Related Works

In recent years, numerous approaches for the automated detection of MAs have been presented. The majority of works is based on the single image scenario. The single image MAD approaches can be categorised into three classes: texture descriptors, e.g. in [20, 24, 26], forensic image analysis, e.g. in [23, 32], and methods based on deep neural networks, e.g. in [7, 21]. These differ in the artefacts they can potentially detect. A brief overview is given in Table 15.1.

Differential MAD can be categorised into approaches that perform a biometric comparison directly with the two facial images, e.g. in [30], and algorithms that attempt to reverse the (potential) morphing process, e.g. in [6, 16]. In the former category, features from both face images, the potentially morphed facial image and the probe image, are extracted and then compared. The comparison of the two feature vectors and the classification as bona fide comparison or MA is usually done using machine learning techniques. By specifically training these procedures for the recognition of MAs, they can—in contrast to facial recognition algorithms—learn to recognise specific patterns within the differences between the two feature vectors for these attacks. This has already been demonstrated for features derived from general purpose texture descriptors. While training a deep neural network from scratch in order to learn discriminative features for MAD requires a high amount of training data, pre-trained deep networks can be employed. The second type of differential MAD procedure aims at reversing the morphing process in the reference image ("demorphing") by using a probe image. If the reference image was morphed from two images and the probe image shows a person contributing to the morph (the attacker), the face of the accomplice would ideally be reconstructed, which would be rejected in a subsequent comparison with the probe image using biometric face recognition; if, on the other hand, a bona fide reference image is available, the same subject should still be recognisable after the reversal of a presumed morph process with the probe image, and thus the subsequent comparison of the facial recognition process should be successful.

Despite promising results reported in many studies, the reliable detection of morphed facial images is still an open research task [14]. In particular, the generalis-

Table 15.1 Categories of singe image MAD approaches and analysed artefacts

Category	Analysed artefacts
Texture descriptors	Smoothened skin texture, ghost artefacts/ half-shade effects (e.g. on pupils, nostrils), distorted edges, offset image areas
Forensic image analysis	Sensor pattern noise, compression artefacts, inconsistent illumination or colour values
Deep-learning approaches	All possible artefacts learned from a training dataset

ability and robustness of the published approaches could not yet be proven while some results are hardly comparable and comprehensible. The vast majority of publications use internal databases of the respective research groups for training and testing [27]. In addition, different evaluation metrics are used in the publications. Since most implemented MAD procedures are not made publicly accessible, comparative independent evaluation of the detection performance is difficult. First efforts towards benchmarking MAD algorithms have been made in [15, 22]. Furthermore, most publications only use images from a single database and morphs generated with a single algorithm for training and testing, so that the generalisation capability of the methods cannot be assessed across different databases and morphing methods. In publications on differential MAD, the comparison images used often show a low variance with respect to poses, facial expressions and illumination and are usually produced shortly after the reference image—in real scenarios such as border control, a much higher variance is to be expected. In addition, many studies neglect the probable application of image post-processing techniques by an attacker, such as subsequent image sharpening, and the print-scan transformation [14].

15.3 Morphing Attack Detection Pipeline

The individual modules of the pipeline considered for MAD algorithms are illustrated in Fig. 15.2. The pipeline consists of the following 4 steps: data preparation, feature extraction, feature preparation, and classifier training.

15.3.1 Data Preparation and Feature Extraction

For most feature extractors it is necessary to pre-process the face image beforehand. The result of feature extractors depends on the resolution of the analysed image, requiring a normalisation of the image size. Especially with the TLCs, variances in position and pose may occur, which can be corrected by the data preparation. In addition, it is useful, for example, for texture-based feature extractors, to crop the image to the relevant facial area, ensuring that no information from the background influences the feature vector.

Depending on the feature extractor selected and the configuration, the obtained feature vector will contain different information, information not contained in the

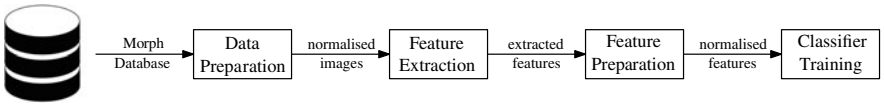


Fig. 15.2 Design of MAD pipeline

feature vector is not available to the algorithms in the further process. For example, if a basic Local Binary Patterns (LBP) histogram is calculated, the feature vector will not contain any spatial information. If, despite the use of LBP histograms, spatial information is to be included in the feature vector, the image to be analysed can be divided into cells, a histogram can be calculated for each cell and the resulting histograms can be concatenated. Thus, spatial information in resolution of the cells can be preserved, however, the length of the feature vector increases accordingly.

15.3.2 Feature Preparation and Classifier Training

Once the feature vectors have been created, they have to be prepared for the training of the classifier. For example, many classifiers only accept one-dimensional input data, requiring multi-dimensional characteristics to be prepared accordingly. Further, for differential MAD algorithms, this module combines the feature vectors of the suspected morph and the TLC. The choice of the combination method is arbitrary but determines the length of the resulting feature vector as well as the contained information. Most classifiers require normalised data for optimal training, thus feature normalisation may be required.

In the last module classifiers are trained on basis of the previously prepared feature vectors. In order to achieve the best possible separation of the feature vectors into classes, appropriate classifiers and parameters have to be chosen. The optimal classifier and parameters depend on the information in the respective feature vectors.

15.4 Database

The face image database used in this work is based on the publicly available FERET [19] and FRGC [18] face image databases. The creation of the database requires 3 categories of images: bona fide reference images, morph input images, and TLC images. The bona fide reference images correspond to an unaltered passport image and should meet the corresponding quality criteria. The morph input images are used in pairs for the morphing process. These should be of passport image quality as well. For the selection of the images in passport image quality, the guidelines standardised in ISO/IEC 19794-5 [8] were followed. Consequently, only images with a closed or minimally opened mouth and a neutral facial expression or a slight smile were included. Images with reflecting glasses were discarded. The class of TLC images corresponds to live recordings, for example, at an ABC gate. Therefore, the images should not be of a controlled, high quality, as this cannot be expected from semi-supervised capturing. For this class, all images not classified as suitable for passport photos in the above pre-selection can be considered. Thus, these images may contain variations in sharpness, lighting, facial expressions, pose, etc.

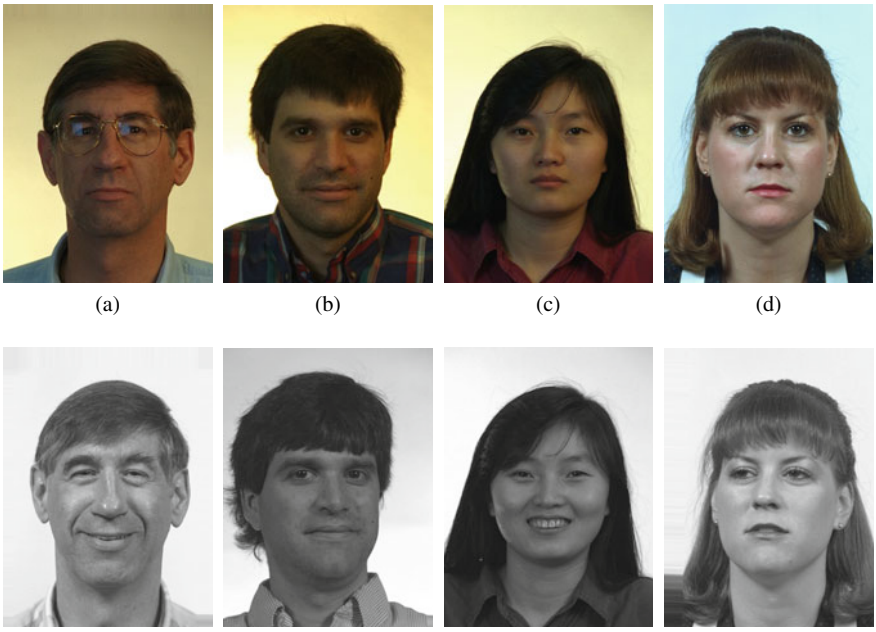


Fig. 15.3 Examples of reference and grey scale TLC images for FERET

Table 15.2 Composition of the database resulting from the image pre-selection

Database	Subjects	Male	Female	Bona fide	Morph input	TLC
FERET	530	330	200	530	530	791
FRGC	533	231	302	984	964	1726

The partitioning of the images into the classes *passport image quality* and *TLC quality* was carried out manually. In the FERET subset, mainly different facial expressions and slight rotations in the pose are included, examples are given in Fig. 15.3. In the FRGC subset, the variances are more significant. In addition to different facial expressions, different backgrounds, illuminations and focuses of the images can be observed, examples are shown in Fig. 15.4.

Based on the two pre-sorted classes, the images are divided into three categories: bona fide reference images, morph input images and TLC images. In order to create realistic scenarios, the time of capture between the passport images and the probe images is maximised as far as possible on the basis of the databases. Due to the large differences in the number of images per subject between the databases, different protocols are used for both databases. The composition of the resulting database is listed in Table 15.2.

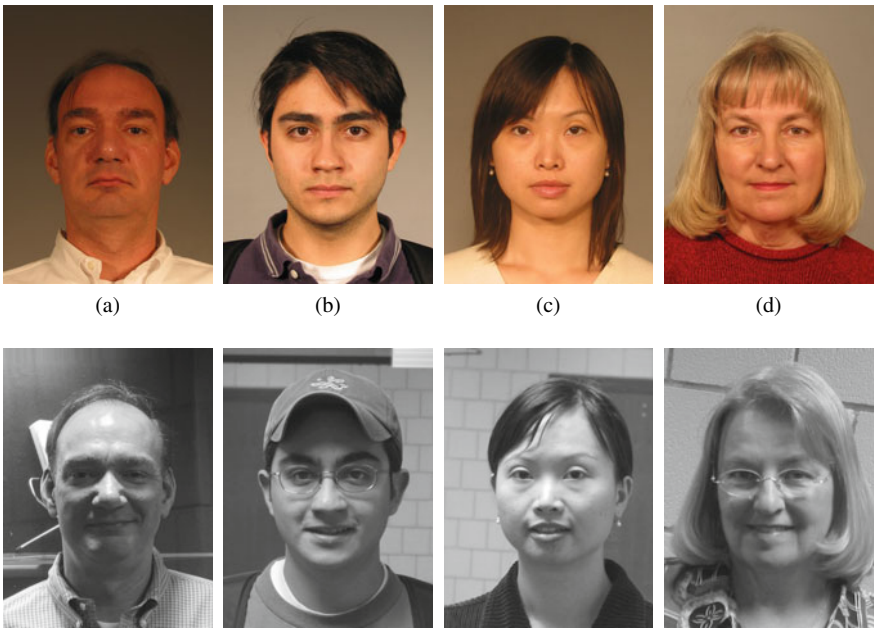


Fig. 15.4 Examples of reference and grey scale TLC images for FRGC

15.4.1 Image Morphing

In order to enable the database to be used for evaluating the generalisability of MAD algorithms towards differing morphing algorithms, four different morphing algorithms are applied to construct the database, hereafter referred to as FaceFusion,[1] FaceMorpher,[2] OpenCV and UBO Morpher:

- **FaceFusion** is a proprietary morphing algorithm. Originally being an iOS app, an adaptation for Windows which uses the 68 landmarks of Dlib and Delaunay triangles was applied. After the morphing process, certain regions (eyes, nostrils, hair) of the first face image are blended over the morph to hide artefacts. Optionally, the corresponding landmarks of upper and lower lips can be reduced as described in [12] to avoid artefacts at closed mouths. The created morphs have a high quality and low to no visible artefacts. An example is shown in Fig. 15.5b.
- **FaceMorpher** is an open-source implementation using Python. In the version applied for this work, the algorithm uses STASM for landmark localisation. Delaunay triangles, which are formed from the landmarks, are warped and blended. The

[1] www.wearemoment.com/FaceFusion.

[2] https://github.com/alyssaq/face_morpher.

(a) Subject 1 (b) FaceFusion (c) FaceMorpher

(d) OpenCV (e) UBO Morpher (f) Subject 2

Fig. 15.5 Examples of morphing face images **a** and **f** using all four algorithms (**b**)–(**e**)

area outside the landmarks is averaged. The generated morphs show strong artefacts in particular in the area of neck and hair. An example is shown in Fig. 15.5c.

- **OpenCV** is a self implemented morphing algorithm derived from "Face Morph Using OpenCV".[3] This algorithm works similar to FaceMorpher. Important differences between the algorithms are that for landmark detection Dlib is used instead of STASM and that for this algorithm landmarks are positioned at the edge of the image, which are also used to create morphs. Thus, in contrast to FaceMorpher, the edge does not consist of an averaged image, but like the rest of the image, of morphed triangles. However, strong artefacts outside the face area can be observed, which is mainly due to missing landmarks. An example is shown in Fig. 15.5d.
- **UBO Morpher** is the morphing tool of University of Bologna, as used, e.g. in [6]. This algorithm receives two input images as well as the corresponding landmarks. Dlib landmarks were used in this work. The morphs are generated by triangulation,

[3] www.learnopencv.com/face-morph-using-opencv-cpp-python.

Table 15.3 Number of comparisons per post-processing in the resulting database

Database	Genuine Comp. Bona Fide Comp.	Impostor comparisons	Morph comparisons	Bona fide samples	Morph samples
FERET	791	418,966	791	530	529
FRGC	3,298	1,695,086	3,246	984	964

warping and blending. To avoid artefacts in the area outside the face, the morphed face is copied to the background of one of the original images. Even if the colours are adjusted, visible edges may appear at borderline of the blended areas. An example is shown in Fig. 15.5e.

The morph input images are used to create the morphs. Morph pairs were formed in a way to keep the ratio between morphs and bona fide images in balance. Two parameters, namely, sex and whether the subject wears glasses, are taken into account for the construction of the morph pairs. Morphing subjects of different sexes usually results in morphs with unnatural appearance. The creation of morphs with subjects of different sex are not to be expected in the real scenario, thus they are excluded from the database. Furthermore, it has been found, that if two subjects wearing glasses are morphed, the resulting morph contains double glasses. To avoid this kind of artefacts, morph pairs are formed with at most one subject wearing glasses.

The morph pairs are formed within one face database, in order to enable a clear separation of datasets during training and evaluation. Due to the different number of morph input images per subject in both databases, different protocols are defined. With each morphing tool morphs were created from all available morph pairs. The morphs were created with a blending and warping factor of 0.5. However, due to the automatic improvement processes of FaceFusion and UBO Morpher, the morphs created by these algorithms are not symmetrical.

The properties of the resulting database are listed in Table 15.3. For the evaluation of differential MAD algorithms the number of bona fide comparisons and morph comparisons is relevant, for single image MAD algorithms the number of bona fide samples and morph samples, respectively. The values given are per post-processing, quadrupling the actual number of passport images contained in the database.

15.4.2 Image Post-Processing

The passport images (morph and bona fide) and the TLC images are post-processed in a different way. The TLC images are converted to greyscale, as some camera systems used at border control are only providing monochrome images. Since the morphing algorithms produce different, and sometimes recognisable, outputs, for example, by partially normalising the images, all passport images (including the bona fides) are

(a) original (b) resized (c) JPEG2000 (d) Print/Scan

Fig. 15.6 Examples of an original image and the three post-processing types

normalised. This also prevents from over-fitting to artefacts not present in a real scenario, such as different image sizes between morphs and bona fides. During the normalisation process, images are scaled to 960×720 pixels, resulting in a face region of 320×320 pixels.

Depending on the process by which the facial image is inserted into the passport, various post-processing steps are performed on the image. To reflect the realistic scenarios, the database contains four different post-processing chains for all passport photographs (Fig. 15.6):

- **Unprocessed**: The images are not further processed. In the text below referred to as *NPP* (no post-processing). This serves as baseline.
- **Resized**: The resolution of the images is reduced by half, reflecting the average size of a passport image. This pre-processing corresponds to the scenario that an image is submitted digitally by the applicant.
- **JPEG2000**: The images are resized by half and then compressed using JPEG2000, a wavelet-based image compression method that is recommended for EU passports [4]. The setting is selected in a way that a target file size of 15KB is achieved. This scenario reflects the post-processing path of passport images if handed over digitally at the application desk.
- **Print/Scan–JPEG2000** The original images (uncompressed and not resized) are first printed with a high quality laser printer (*Fujifilm Frontier 5700R Minlab* on *Fujicolor Crystal Archive Paper Supreme HD Lustre photo paper*) and then scanned with a premium flatbed scanner (*Epson DS-50000*) with 300 dpi. A dust and scratch filter is then applied in order to reduce image noise. Subsequently, the images are resized by half and then compressed to 15 KB using JPEG2000.[4] This scenario reflects the post-processing path of passport images if handed over at the application desk as a printed photograph.

[4] Due to the glossy print, the scans exhibit a visible pattern of the paper surface, which is only partly removed by the dust and scratch filter and results in stronger compression artefacts than for scans of glossy prints.

15.5 Morphing Attack Detection Methods

Different types of MAD methods are considered in a single image and differential scenario. According to the previously described MAD pipeline, these use a similar pre-processing and the same classification. For the feature extraction step different types of texture descriptors are employed, including traditional algorithms as well as gradient-based methods. In addition, deeply learned features are used.

15.5.1 Pre-Processing

In the pre-processing, face images are normalised by applying suitable scaling, rotation and padding/cropping to ensure alignment with respect to the eyes' positions. Precisely, facial landmarks are detected applying the *dlib* algorithm [11] and alignment is performed with respect to the detected eye coordinates with a fixed position and an intra-eye distance of 180 pixels. Subsequently, the normalised images are cropped to regions of 160×160 pixels centred around the tip of the nose.

15.5.2 Feature Extraction

For the feature extraction step, three types of descriptors are considered: texture descriptors, gradient-based descriptors, as well as descriptors learned by a deep neural network.

Texture Descriptors: During the creation of morphed facial images, the morphing process introduces changes into the image that can be used to detect said images. In particular, these changes are reflected by faulty regions, such as overlapping landmarks, which result in incorrectly distorted triangles, as shown in Fig. 15.7a. Another error common to automated morphing algorithms is artefacts in the eye region, which is particularly prone to errors due to the high contrast provided by shadows and wrinkles, and the difficult detection of the iris. An example of artefacts in the eye region is given in Fig. 15.7b. Furthermore, ghost artefacts can be caused by landmarks that are too few or too poorly positioned. This happens frequently in the area of the neck or hair, as visualised in Fig. 15.7c. In order to be able to map this kind of image changes in feature vectors, texture descriptors can be used. In this work, the suitability of LBP [1] and Binarized Statistical Image Features (BSIF) [10] for detecting these artefacts is investigated.

By calculating the classical LBP histogram obtained from 3 LBP patches, any local information contained in the image is discarded. To preserve local information, the LBP image can be divided into cells, subsequently a histogram is calculated for each cell. As a result, the length of the feature vector multiplies by

the number of cells, but spatial information is obtained in resolution of the cell division. An inevitable correlation exists between cell division, patch size, image size and the resulting histogram. The finer the cell division and the larger the patch, the fewer values can be calculated per cell and the sparser the histogram. As the resolution increases, the number of values per cell increases as well. For the applied patch sizes and the region of 160×160 pixels, a subdivision into 4×4 cells has shown to be appropriate, thus it is implemented in addition to the LBP calculation without cell division.

As a further texture descriptor, BSIF is used. As for LBP, it has been shown that the use of larger BSIF patches results in more robust systems, but using smaller BSIF patches results in significantly higher performance [28]. In order to allow a better comparison to BSIF with a patch size of 3×3 pixels with 8 filters are used. The resulting feature vector is of length 256. Also, to ensure comparability, the same configuration as for LBP of division into 4×4 cells is implemented.

Gradient-based Descriptors: Histograms of Oriented Gradients (HOG) [2, 13] represents a gradient-based descriptor. For HOG, the definition of the parameters influences the result of the histogram calculation, as well as the length and content of the feature vector. In order to achieve a robust and general applicable HOG extraction, recommended standard parameters[5] are applied, namely 9 orientations, 8×8 pixels per cell (which corresponds to 20×20 cells for regions of 160×160 pixels), and 3×3 cells per block, resulting in a feature vector of length 26,244.

Deep Features: Machine learning algorithms, especially Deep Convolutional Neural Networks (D-CNN), can be used to extract statistically significant features from images in addition to hand-crafted feature extractors. The difficulty of this approach is the dependence of the information represented in the extracted features on the nature of the training data used to train the feature extractor. If the wrong training data is chosen, this might cause an over-fitting of the feature extractor, resulting in very good results on known data, which, however, cannot be reproduced in a real use case on unknown data. In order to avoid this effect, only D-CNN pre-trained for face recognition are applied in this thesis. These networks have been trained to extract representative features from facial images, without containing morphed facial images in the training process, thus implicitly preventing an over-fitting to artefacts of a specific morphing algorithm. In the implemented MAD pipeline the feature extractors of three different face recognition systems are used, which are described in more detail in the following sections.

In the MAD pipeline the existing implementation[6] of the authors of [3] is utilised. In contrast to the previously mentioned methods, here the images are

[5] The standard parameters are derived from the documentation of the used HOG implementation: https://scikit-image.org/docs/dev/api/skimage.feature.html.

[6] The corresponding source code can be found at: https://github.com/deepinsight/insightface.

(a) Example of errors introduced by incorrectly distorted triangles

(b) Example of errors in eye region

(c) Example of errors in hair region

Fig. 15.7 Example of errors introduced by incorrect morphing

normalised using MTCNN and scaled to 112×112 pixels, prior to training or feature extraction. The authors offer several pre-trained models, in this pipeline the model *LResNet50E-IR,ArcFace@ms1m-refine-v1* is chosen, since, according to the authors, it achieves the most stable performance on the tested databases. The architecture of the selected network is, as the name suggests, a residual network comprised of 50 layers. A residual network is characterised by shortcut connections between different layers, allowing the output of a previous layer (residuals) to be processed as input on subsequent layers, simplifying the computationally expensive training of very deep CNN.

15.5.3 Classification

In a single image MAD system, the detector processes only the suspected reference image. For this detection approach, the extracted feature vectors are directly analysed. In contrast, in the differential detection systems, a trusted live capture from an authentication attempt serves as additional source of information for the detector. This information is utilised by estimating the vector *differences* between feature vectors extracted from processed pairs of images. Specifically, an element-wise subtraction of feature vectors is performed.

Support Vector Machines (SVMs) with Radial Basis Function (RBF) kernels are used to distinguish between bona fide and retouched face images. In order to train SVMs, the *scikit-learn* library [17] is applied. Since the feature elements of extracted feature vectors are expected to have different ranges, data-normalisation is employed. Data-normalisation turned out to be of high importance in cross-database experiments. It aims to rescale the feature elements to exhibit a mean of 0 and a standard deviation of 1. At the time of training, a regularisation parameter of $C = 1$ and a kernel coefficient Gamma of $1/n$ is used, where n represents the number of feature elements.

15.6 Experiments

To compare different MAD algorithms with each other, uniform evaluation methods and metrics are essential. For the evaluation of the vulnerability of face recognition systems against MAs, different metrics have been introduced in previous publications, e.g. [25], which will not be described further. To evaluate the performance of MAD algorithms, each comparison is considered individually, since each morph has to be detected separately. For this reason, the metrics defined in ISO/IEC 30107-3 [9] for the performance reporting of presentation attacks can be used, namely, Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER), which are defined as follows [9]:

- **APCER**: proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario.
- **BPCER**: proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario.

In an effective MAD system, the resulting MAD scores of MA and bona fide samples should be clearly separable. For overlapping MAD score distributions, a trade-off between security (low APCER) and high throughput (low BPCER) has to be found by setting a corresponding decision threshold. The Detection Equal Error Rate (D-EER) reflects the error rates in a single operating point where the APCER is equal to the BPCER. Hereafter the D-EER will be used for measuring the performance of MAD methods.

15.6.1 Generalisability

In the first experiment, the generalisability of MAD methods across heterogeneous data sources in analysed. To this end, the MAD methods based on LBP and BSIF texture descriptors are evaluated in a single image and a differential scenario. On the one hand, this is done for a split of the FRGC dataset into a training and test set. On the other hand, the entire FRGC dataset is used for training while testing is performed on the FERET dataset. Obtained results are summarised in Table 15.4. It can be observed that D-EER values significantly increase in case the data source is unknown. This holds for both, the single image and differential scenario when using LBP and BSIF for the feature extraction. That is, MAD algorithms may overfit to certain data sources which underlines the importance of evaluating MAD methods in cross-database experiments. In all of the following experiments, the FRGC database will be used during the training stage and testing is performed on the FERET dataset.

Table 15.4 Influence of unknown data sources on MAD methods

Training		Test		Single image		Differential	
Database	Morphing algorithm	Database	Morphing algorithm	LBP (%)	BSIF (%)	LBP (%)	BSIF (%)
FRGC-Train	OpenCV	FRGC-Test	OpenCV	5.2	3.5	3.9	4.7
FRGC	OpenCV	FERET	OpenCV	22.4	20.1	28.8	18.1

15.6.2 Detection Performance

In the next experiment, the suitability of all feature extractors for MAD is investigated. Here, training is conducted on low quality morphs (FaceMorpher and OpenCV) while the testing is done on high quality morphs (FaceFusion, UBO Morpher) in order to obtain a more challenging scenario. Table 15.5 summarised the three best performing MAD methods in the single image and differential scenario (best results marked bold). For the single image scenario, the most competitive results are achieved when using HOG for feature extraction. However, obtained D-EERs are still rather high, i.e. reliable MAD appears more challenging in the single image scenario. In contrast, for the differential MAD methods significantly lower D-EER can be obtained. In particular, for the use of deep features D-EER values below 3% are achieved. Note, that deep features have not been found suitable for the single image MAD. Hence, it can be concluded that deep features are highly suitable for differential MAD which has also been reported in [15, 30]. Focusing on single image MAD more elaborated feature extractors are required to better distinguish between bona fide and morphed face images.

Table 15.5 Performance of MAD algorithms

Training	Test	Single image			Differential		
Morphing algorithm	Morphing algorithm	LBP (%)	BSIF (%)	HOG (%)	LBP (%)	HOG (%)	Deep features (%)
FaceMorpher	FaceFusion	31.01	30.76	**24.05**	24.30	19.37	**2.71**
	UBO morpher	26.71	28.99	**19.75**	19.62	15.70	**2.58**
OpenCV	FaceFusion	26.20	31.01	**23.92**	22.41	18.73	**2.71**
	UBO morpher	24.05	28.61	**20.63**	19.11	15.70	**2.71**

15.6.3 Post-Processing

Eventually, the influence of considered image post-processings on the used MAD methods is estimated. Here, training is performed on the original images and testing on post-processed ones. It was found that resizing has negligible impact on MAD performance of the considered methods. Table 15.6 summarises the impact of image compression using JPEG2000 for the best performing single image and differential MAD approach. Focusing on the best single image MAD based on HOG, a significant increase of D-EER values can be observed. This means image compression negatively impacts this single image MAD algorithm. Due to the compression, artefacts which have been learned to distinguish morphed images from bona fide images might vanish which is particularly the case for the used JPEG2000 algorithm. In contrast, deep features turn out to be robust to image compression. This is the case since these are extracted by a face recognition model which has been trained to extract discriminative face representations which are highly robust to such post-processings.

Finally, the impact of printing and scanning on the MAD performance is evaluated. Corresponding results are summarised in Table 15.7. Again, a significant drop in the detection performance can be observed for the single image MAD method based on HOG. The artefacts introduced by the printing and scanning process increase the D-EER to a large extent. However, the differential MAD algorithm based on deep features maintains detection performance for printed and scanned images.

Table 15.6 Influence of image compression on MAD methods

Morphing algorithm		Single image	Differential
Training	Test	HOG	Deep features
FaceMorpher	FaceFusion	28.2% (+4.1)	3.0% (+0.3)
	UBO morpher	27.3% (+7.6)	3.1% (+0.5)
OpenCV	FaceFusion	31.9% (+8.0)	2.7% (+-0)
	UBO morpher	31.0% (+10.4)	2.7% (+-0)

Table 15.7 Influence of printing and scanning on MAD methods

Morphing algorithm		Single image	Differential
Training	Test	HOG	Deep features
FaceMorpher	FaceFusion	34.1% (+10)	1.3% (-1.4)
	UBO Morpher	36.6% (+16.8)	3.2% (+0.6)
OpenCV	FaceFusion	53.4% (+29.5)	1.4% (-1.3)
	UBO Morpher	37.1% (+19.5)	3.1% (+0.4)

15.7 Summary

MAs pose a high security risk to modern facial recognition systems in particular for border control. To counteract this, reliable methods for MAD must be developed. Various research groups from the fields of image processing and biometrics have recently published scientific papers on this topic, and several publicly funded research projects are currently dealing with this problem. However, research in this field is still in its infancy and does typically not address the variance of the image data available in border control scenarios. The development of MAD approaches that are effective and robust in real-world scenarios will require a considerable amount of future research as well as close collaborations with border guard agencies.

The majority of the MAD methods published so far—in particular the single image MAD methods—aim at the detection of artefacts that can easily be avoided, e.g. clearly visible ghost artefacts, double compression artefacts and changed image noise patterns. Further, usually face images are taken from a single data source, i.e. face image database. Hence, reported detection rates tend to be over-optimistic. MAD approaches are, like any classification task, susceptible to over-fitting to training data. When evaluating MAD approaches, images of which source and properties differ from those of the training data, i.e. images from other databases and morphs created with other techniques should be employed. In case of unknown MAs, i.e. face images stem from different data sources and were created with unknown morphing algorithms, the detection performance of MAD methods may significantly drop, as shown in this work. Further, it was shown that post-processing steps applied to reference images like printing/scanning and strong image compression may cause drastic drops in the detection performance at least for single image MAD, since artefacts caused by morphing vanish in the post-processed reference. In contrast to many published works on MAD (see [29, 31]), the results reported in this work are supported by external evaluations conducted in [15, 22].

In contrast, research should focus on the development of MAD methods that detect artefacts that are difficult to avoid. While the detection performance for differential MAD based on deep features showed promising results in the experiments of this work, the used datasets might not fully reflect real-world scenarios. For border control scenarios, MAD techniques need to be robust against print-scan transformations, resizing and strong compression of reference images. Similarly, in the case of differential MAD, considerable variance of illumination, background, pose, appearance (hair, beard, glasses, etc.) and ageing (up to 10 years for passports) can be expected in probe images. In order to be applicable to these scenarios, MAD approaches should be trained and evaluated on images exhibiting these characteristics.

Acknowledgements This research work has been partially funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Centre for Applied Cybersecurity ATHENE.

References

1. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. Springer, Berlin, pp 469–481
2. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the 2005 computer society conference on computer vision and pattern recognition (CVPR)
3. Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the 2019 computer vision and pattern recognition (CVPR)
4. Commission European (2018) EU-eMRTD specification. Technical report, European Commission
5. Ferrara M, Franco A, Maltoni D (2014) The magic passport. In: Proceedings of the 2014 international joint conference on biometrics (IJCB). IEEE
6. Ferrara M, Franco A, Maltoni D (2018) Face demorphing. IEEE Trans Inf Forensics Secur 13(4):1008–1017
7. Ferrara M, Franco A, Maltoni D (2021) Face morphing detection in the presence of printing/scanning and heterogeneous image sources. IET-Biometrics, pp 1–13
8. ISO/IEC JTC1 SC37 Biometrics (2011) Information technology–Biometric data interchange formats–Part 5: face image data
9. ISO/IEC JTC1 SC37 Biometrics (2017) Information technology–biometric presentation attack detection–part 3: testing and reporting. ISO ISO/IEC IS 30107-3:2017, International Organization for Standardization, Geneva, Switzerland
10. Kannala J, Rahtu E (2012) BSIF: binarized statistical image features. In: Proceedings of the 21st international conference on pattern recognition (ICPR), pp 1363–1366
11. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
12. Makrushin A, Neubert T, Dittmann J (2017) Automatic generation and detection of visually faultless facial morphs. In: Proceedings of the 12th international joint conference on computer vision, imaging and computer graphics theory and applications (VISIGRAPP). SCITEPRESS - Science and Technology Publications
13. McConnell RK (1986) Method of and apparatus for pattern recognition
14. Merkle J, Rathgeb C, Scherhag U, Busch C, Breithaupt R (2019) Face morphing detection: issues and challenges. In: Proceedings of the international conference on biometrics for borders (ICBB)
15. Ngan M, Grother P, Hanaoka K, Kuo J (2020) Face recognition vendor test (FRVT) part 4: Morph-performance of automated face morph detection. Technical report, National Institute of Standards and Technology (NIST)
16. Ortega-Delcampo D, Conde C, Palacios-Alonso D, Cabello E (2020) Border control morphing attack detection with a convolutional neural network de-morphing approach. IEEE Access 8:92301–92313
17. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res (JMLR) 12:2825–2830
18. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: Proceedings of the 2005 computer society conference on computer vision and pattern recognition (CVPR). IEEE
19. Phillips PJ, Wechsler H, Huang J, Rauss PJ (1998) The FERET database and evaluation procedure for face-recognition algorithms. Image Vision Comput 16(5):295–306
20. Raghavendra R, Raja KB, Busch C (2016) Detecting morphed face images. In: Proceedings of the 8th international conference on biometrics theory, applications and systems (BTAS). IEEE
21. Raghavendra R, Raja KB, Venkatesh S, Busch C (2017) Transferable deep-CNN features for detecting digital and print-scanned morphed face images. In: Proceedings of the 2017 conference on computer vision and pattern recognition workshops (CVPRW). IEEE
22. Raja K, Ferrara M, Franco A, Spreeuwers L, Batskos I et al (2020) Morphing attack detection - database, evaluation platform and benchmarking. IEEE Trans Inf Forensics Secur
23. Scherhag U, Debiasi L, Rathgeb C, Busch C, Uhl A (2019) Detection of face morphing attacks based on PRNU analysis. IEEE Trans Biom Behav Identity Sci (T-BIOM), pp 1–16

24. Scherhag U, Kunze J, Rathgeb C, Busch C (2020) Face morph detection for unknown morphing algorithms and image sources: a multi-scale block local binary pattern fusion approach. IET-Biomet pp 1–11
25. Scherhag U, Nautsch A, Rathgeb C, Gomez-Barrero M, Veldhuis RMJ, Spreeuwers L, Schils M, Maltoni D, Grother P, Marcel S, Breithaupt R, Raghavendra R, Busch C (2017) Biometric systems under morphing attacks: assessment of morphing techniques and vulnerability reporting. In: Proceedings of the 2017 international conference of the biometrics special interest group (BIOSIG). IEEE
26. Scherhag U, Ramachandra R, Raja KB, Gomez-Barrero M, Rathgeb C, Busch C (2017) On the vulnerability of face recognition systems towards morphed face attacks. In: Proceedings of the 5th international workshop on biometrics and forensics (IWBF). IEEE
27. Scherhag U, Rathgeb C, Busch C (2018) Performance variation of morphed face image detection algorithms across different datasets. In: Proceedings of the 6th international workshop on biometrics and forensics (IWBF). IEEE
28. Scherhag U, Rathgeb C, Busch C (2018) Towards detection of morphed face images in electronic travel documents. In: Proceedings of the 13th workshop on document analysis systems (DAS). IAPR
29. Scherhag U, Rathgeb C, Merkle J, Breithaupt R, Busch C (2019) Face recognition systems under morphing attacks: a survey. IEEE Access 7:23012–23026
30. Scherhag U, Rathgeb C, Merkle J, Busch C (2020) Deep face representations for differential morphing attack detection. IEEE Trans Inf Forensics Secur (TIFS), pp 3625–3639
31. Venkatesh S, Ramachandra R, Raja K, Busch C (2021) Face morphing attack generation detection: a comprehensive survey. IEEE Trans Technol Soc, pp 1–1
32. Zhang LB, Peng F, Long M (2018) Face morphing detection using fourier spectrum of sensor pattern noise. In: Proceedings of the 2018 international conference on multimedia and expo (ICME). IEEE