Journal of
Imaging

Article

Detecting Morphing Attacks through Face Geometry Features

Stephanie Autherith [1] and Cecilia Pasquini [2,*]

[1] Department of Computer Science, University of Innsbruck , Technikerstraße 21A, 6020 Innsbruck, Austria; stephanie.autherith@uibk.ac.at

[2] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123 Trento, Italy

* Correspondence: cecilia.pasquini@unitn.it

Received: 19 September 2020; Accepted: 21 October 2020; Published: 29 October 2020

Abstract: Face-morphing operations allow for the generation of digital faces that simultaneously carry the characteristics of two different subjects. It has been demonstrated that morphed faces strongly challenge face-verification systems, as they typically match two different identities. This poses serious security issues in machine-assisted border control applications and calls for techniques to automatically detect whether morphing operations have been previously applied on passport photos. While many proposed approaches analyze the suspect passport photo only, our work operates in a differential scenario, i.e., when the passport photo is analyzed in conjunction with the probe image of the subject acquired at border control to verify that they correspond to the same identity. To this purpose, in this study, we analyze the locations of biologically meaningful facial landmarks identified in the two images, with the goal of capturing inconsistencies in the facial geometry introduced by the morphing process. We report the results of extensive experiments performed on images of various sources and under different experimental settings showing that landmark locations detected through automated algorithms contain discriminative information for identifying pairs with morphed passport photos. Sensitivity of supervised classifiers to different compositions on the training and testing sets are also explored, together with the performance of different derived feature transformations.

Keywords: face morphing; forensics detection; face landmarks; automatic border control

1. Introduction

Automated face recognition and verification are widely studied problems in computer vision, for which accurate solutions have been developed and commercialized [1,2]. As a result, they are used in security contexts as means for person authentication, thus representing an alternative to more traditional schemes based on passwords and PINs (Personal Identification Number) and to other biometric traits like fingerprints. This includes applications such as face-based authentication in mobile devices and automated border controls (ABC) through passport photos [3].

In the ABC scenario, face information is used for identity verification starting from electronic Machine Readable Travel Documents (eMRTD). To this end, a live probe image of the subject physically present at border control is acquired and compared with the image stored in his/her eMRTD via face verification (FV) algorithms, which provide a binary output indicating whether the two images depict the same subject. In order to aid both algorithmic and human FV, photos in eMRTD must fulfil restrictive quality standards, as specified by the International Standard Organization (ISO) and the International Civil Aviation Organization (ICAO) guidelines. In particular, the face must be straight looking, acquired in frontal position, and not covered by hair or clothes.

Facilitated by these requirements, advanced FV algorithms can typically perform identity verification rapidly and accurately, but their effectiveness can be compromised if the images stored in eMRTDs contain alterations. A relevant case is represented by face images resulting from morphing operations [4], i.e., when two images of different subjects are blended together through geometric operations. In this case, FV algorithms are led to detect a match between the morphed image of the eMRTD and probe images from both subjects, as we illustrate in Figure 1.

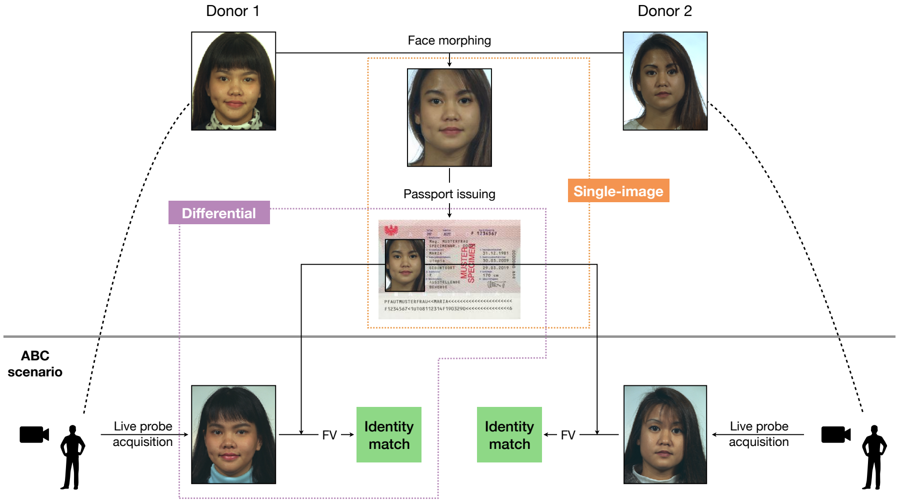


Figure 1. Illustration of a morphing attack against face verification (FV)-based automated border control (ABC) systems (examples are taken from the dataset in [5]): the area of analysis of single-image and differential approaches is highlighted.

In many countries, the image stored in the eMRTD is provided by the citizen during the passport application process, either in printed format or via web-platforms. This offers opportunities for an attacker to introduce altered visual information to be used to their advantage. In this context, a *morphing attack* would allow the same passport containing a morphed photo to be used by two different subjects, potentially including citizens with known criminal records for which border crossing would be forbidden. This kind of attack is particularly insidious as humans can be deceived as well with good probability, as it is shown in [6]. Moreover, it does not require physical forgeries of passports.

In order to contrast possible frauds exploiting these vulnerabilities, techniques for the detection of morphing attacks have been proposed in recent years.

The majority of them focuses on a *single-image* scenario, i.e., they analyse the photo in the eMRTD looking for traces of morphing operations. This includes inhomogeneities in texture patterns, camera fingerprints and compression traces, or visual artefacts like ghost shadows or illumination patterns. An advantage of this class of techniques is that they operate on eMRTD information only and could in principle reveal anomalies before the actual ABC context or even directly during the passport application process, thus enabling an early prevention of morphing attacks. However, they typically suffer from generalization issues due to the high variability of pre- and postprocessing operations which should be expected in real world scenarios [7]. In fact, as widely investigated in the field of image forensics, steps like compression [8], printing/scanning operations [9], resizing [10], and aspect ratio correction might be applied to the photo under investigation with highly diverse parameters and in turn introduce further subtle distortions and artifacts, which can have a strong impact on the (typically weak) morphing traces in the image signal [11,12].

Another interesting yet less explored approach is to consider a *differential* scenario, where the morphing detection is performed with the identity verification process at border control. In this case, the eMRTD photo and the live probe image can be jointly analyzed; thus, the decision is based on an *image pair*. While less timely than the single-image case in detecting anomalies, differential detection can leverage the additional information given by the acquired probe image.

In our work, we address this differential scenario and focus on the use of geometric face features to determine whether the image pair actually contains photos of the same subject or the reference eMRTD image depicts a morphed face. The rationale behind this choice is to capture the geometric inconsistencies between the morphed face and the genuine subject's face that are unavoidably introduced in the morphing process. In fact, the morphing operation impacts the 2D face geometry, while its role has been only marginally investigated in the literature for morphing detection [13,14]. We fill this gap by developing and assessing the effectiveness of binary detectors based on the location of facial landmarks detected in both faces, the eMRTD photo, and the live probe. Those detectors are intended to be applied at ABC on top of the FV algorithm in cases where it detects an identity match between the two faces, since morphing attacks steer the FV decisions towards a positive match.

We can summarize our contributions as follows:

- We conduct an extensive experimental campaign to assess the effectiveness of landmark-based geometric features for the pairs. This includes adopting different training/testing conditions to encourage a sufficiently high variability between training and testing sets in terms of source datasets and subject characteristics and to better assess the generalization abilities of the detectors. A corpus of images belonging to different source datasets has been constructed, which represents a wider and more diverse benchmark with respect to previous studies in this direction [13,14].
- We identify the more relevant face areas for morphing detection through an ablation study on semantically related groups of landmarks, thus gaining insights on the face locations where more discriminative patterns can be found.
- We compare the performance of different transformations of the full set of facial landmarks, including feature representations previously proposed the literature [13,14] and geometric features stemming from findings in facial anthropometry.
- We evaluate the effect of noise sources that can typically affect the image pairs in realistic scenarios, revealing that the performance of the proposed detectors against unseen processing in the training tests are largely preserved. This confirms the advantage of geometric-based method of being stable against common image alterations, as opposed to texture-based approaches.

The manuscript is organized as follows: Section 2 reports an overview of existing approaches for face morphing creation and detection; in Section 3, we illustrate the detection framework and feature representations adopted; Section 4 fully reports the outcomes of the experimental tests we conducted; and Section 5 concludes the paper.

2. Related Work

We illustrate how morphed faces are created (Section 2.1) and then give an overview of the detection techniques proposed in the literature, differentiating between single-image (Section 2.2) and differential (Section 2.3) approaches.

2.1. Creation of Morphed Faces

Face morphing consists of merging together two images depicting two different subjects (called *donors*) into one *morphed* face image, which contains characteristics of both subjects. This process generally involves several rule-based procedures and, although variants can be devised [15], we refer to the work in [6] and visually summarize the main steps in Figure 2.

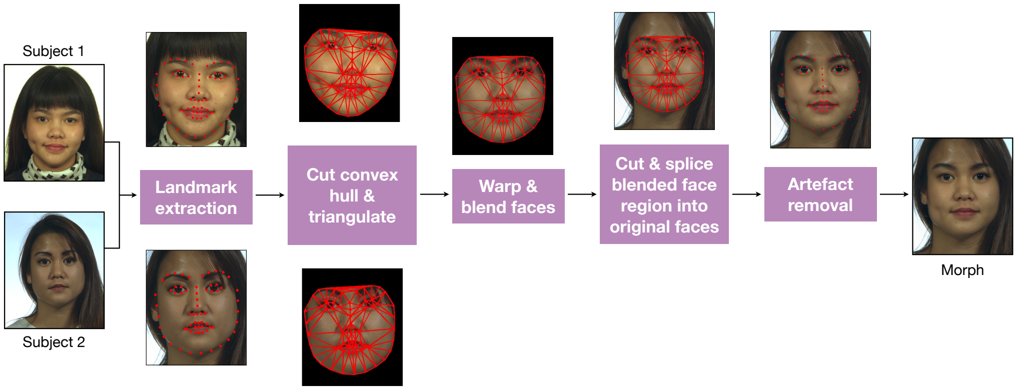

Subject 1
Subject 2
Landmark extraction
Cut convex hull & triangulate
Warp & blend faces
Cut & splice blended face region into original faces
Artefact removal
Morph

Figure 2. Visualization of the morphing process.

Firstly, facial landmarks are detected in both images and linearly blended with a factor which is commonly set to 0.5 [6], so to obtain intermediate landmarks, which are subsequently triangulated. Then, both images are warped to be aligned to the intermediate landmarks and joined together again through cross-dissolving. This can be done on the entire image or by operating only on the convex hull of the landmark set to ease seamless alterations. Additional manual operations can then be applied to remove visual artifacts. Also, visually plausible morphs are generally possible provided that the subjects are depicted in a frontal pose and share similar characteristics, including the same gender.

While some tools are available online [16], obtaining high-quality full-face morphs that do not contain evident visual artefacts and that could then be used for potential attacks is highly time-consuming or requires specific software, generally proprietary [17] or not publicly available [6].

Given the impressive results obtained for other visual tasks, in [18], the authors attempt to use Generative Adversarial Models (GAN) to systematically create morphed faces, although generated images have a fairly low resolution. A follow-up study has been reported in [19], where a higher quality is reached, thus highlighting the potential advantages and promising outcomes of this approach.

2.2. Single-Image Detectors

The methods developed to detect morphing attacks on the reference eMRTD photo mostly rely on pattern recognition techniques used in image processing and image forensics. In fact, the key idea is to detect traces in the image signal of the operations involved in the morphing creation process.

Several approaches explored the effectiveness of texture and keypoint descriptors in detecting anomalies within the passport photo [20–22]. This includes Local Binary Patterns (LBP) [23], Binarized Statistical Image Features (BSIF), and Weighted Local Magnitude Patterns, also combined with other handcrafted features used in computer vision such as Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF) [24], and Histogram of Oriented Gradients (HOG) [20].

Other methodologies resort to techniques originating from image forensics for the detection of local image modifications. To this purpose, a possible approach is to analyse the Photo Response Non-Uniformity (PRNU), which is an imperceptible spatial noise pattern caused by inaccuracies in the sensor manufacturing process. Every acquiring sensor has a characteristic PRNU, and alterations due to morphing can be revealed through its estimation [25,26]. Similarly, local modifications imply diversified compression histories within the same picture, which can be captured by analyzing proper statistical artifacts [15,27]. Also, traces of alterations can be found through modeling light reflection and light sources in different faces areas, observing whether they are physically consistent [28].

Recently, deep features have also been used for morphing detection, either by training or fine-tuning known architectures [29,30] or by using pretrained models as feature extractors . The advantage of neural networks is that they can in principle detect different kind of artifacts, although large datasets with high variance are necessary for training them successfully.

2.3. Differential Detectors

Differential detectors are less explored with respect to single-image methods, and few approaches appear so far in the literature.

One direction is explored by the work in [31,32], where the authors develop a pipeline to reverse the morphing process and to retrieve two face images starting from the one stored in the eMRTD. A morphing attack is detected if one of the two resulting face strongly matches the probe image.

Then, the works in [13,14] firstly combine information from facial landmarks detected in both images, and are further defined in Section 3.2, as they are considered as baselines in our tests. Therefore, the *directed distances* proposed in [13] constitute a transformation aimed at exposing shifting patterns in the landmark geometry. Those geometric artefacts are introduced by the warping step specifically in the morphing process. The features in [14] instead comprise *distances and angle differences* computed between landmarks of two face images. Herein, the angle differences are calculated between neighboring landmarks, while the distance features consider combinations of all the available landmarks. Finally, a solution building on deep face representations has been described in the recently published work [33].

3. Detection Framework

The analyzed geometry-based detectors operate in the presence of the eMRTD and the probe live image depicting the physical subject. As explained in Section 1, the detection is intended to be applied after the FV outcome if an identity match is detected.

In fact, advanced FV algorithms for ABC are designed and calibrated to robustly link faces belonging to the same subject, which are generally in frontal position with close-to-neutral expression but also contain common disturbance factors (such as differences in pose, illumination, and subject's age/haircut). On the other hand, morphing attacks specifically challenge the FV's ability to differentiate very similar yet strategically altered face geometries and thus to reject image pairs containing this kind of inconsistency. For this reason, the geometry-based detectors act as specialized modules based on facial geometry for the detection of potential morphing attacks among image pairs where an identity match results from the FV system, as depicted in Figure 3. Thus, the following classes of image pairs are used for training and testing:

- *Bona fide pairs:* the eMRTD contains a genuine face image of the physical subject.
- *Attacked pairs:* the eMRTD contains a morphed face image of which physical subject is a donor.

The geometry-based detector is a machine learning model that classifies the pair as either bona fide or attacked, based on the facial landmark information extracted from the two images. In Section 3.1, we describe the workflow adopted for the extraction and processing of the landmarks. Moreover, the extracted landmark vector $\mathbf{L}$ can be further combined and transformed through a function Φ to obtain derived feature representations $\Phi(\mathbf{L})$. This can be done in order to reduce the feature dimensionality (and thus to facilitate training also in the case of scarce training data) or to provide more interpretable outcomes, which is typically an advantage of handcrafted features. Thus, in addition to the full set of landmarks, we define different transformations of **L** inspired by studies in craniofacial anthropometry [34], the discipline that analyzes measurements and proportions of human faces.

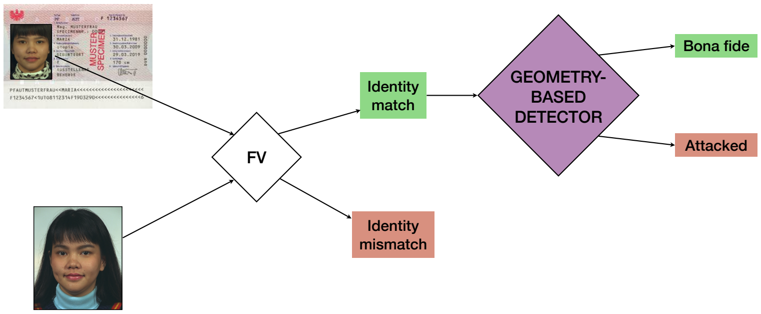


Figure 3. Detection framework.

3.1. Landmark Extraction

Facial landmarks are biologically meaningful keypoints of human faces, widely used for many tasks in computer vision. Several algorithms have been proposed for the automatic detection and localization of these keypoints, and in our work, we use the `dlib` library, which outputs the coordinates of 68 landmarks as depicted in Figure 4. The eye centers are computed starting from the 6 landmarks of each eye, and the landmark coordinates are rotated so that the eye centers lie on the same horizontal line. After being mapped into the interval $[0, 1]$ through a min-max normalization, they are scaled in such a way that the two eye centers of each face are aligned.

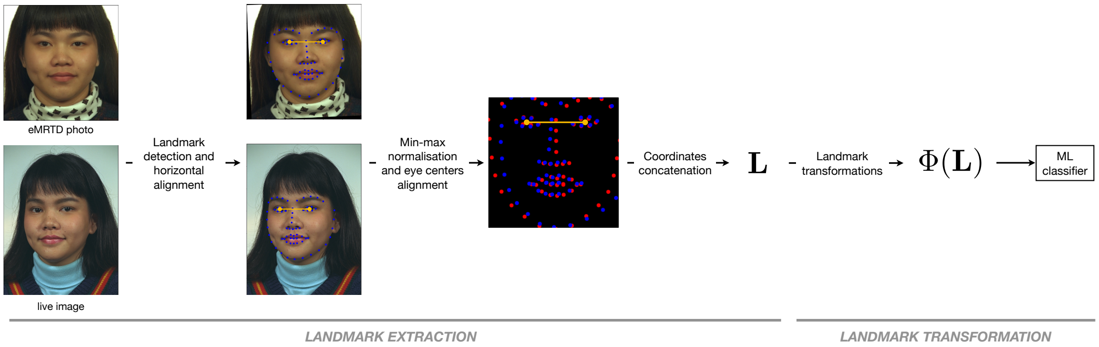

eMRTD photo
live image
Landmark detection and horizontal alignment
Min-max normalisation and eye centers alignment
Coordinates concatenation
L
Landmark transformations
Φ(L)
ML classifier
LANDMARK EXTRACTION
LANDMARK TRANSFORMATION

Figure 4. Landmark extraction and transformation.

The resulting vectors containing the bidimensional coordinates of the face in the passport photo and in the live image, respectively, are then concatenated together into a $68 \times 2 \times 2 = 272$-dimensional vector **L**.

3.2. Landmark Transformations

In order to better encode in the feature vectors geometric characteristics of the two compared faces, handcrafted feature transformations can be applied to **L**. Here, we introduce for comparative testing (see Section 4) two different transformations inspired by anthropometric studies Φ_R and Φ_A (and their union), and we recall previously proposed landmark-based feature representations.

3.2.1. Anthropometry-Based Features

Anthropometric craniofacial proportions [34] are characteristic ratios of distances between specific cranial and facial keypoints. They have been widely studied by anthropologists and used in different domains (ranging from art to medicine and from computer graphics to forensic sciences), and they have also been explored for 2D and 3D face recognition purposes [35,36]. We define the following transformations, yielding different features vectors:

- **Ratios (Φ_R):** for each face, we consider 47 pairs of landmarks and compute the distance between them, as depicted in (Figure 5, left). Those landmarks are selected as highly involved in the morphing process and less sensitive to slight expression variations. Then, those distances are divided individually by the two benchmark distances depicted in red in (Figure 5, middle) and chosen so that they are reliably detected and relatively stable through the morphing process, according to the approach proposed in [36]. Those 94 ratio values from each face are then concatenated, resulting in a feature vector $\Phi_R(\mathbf{L})$ of size 188.
- **Angles (Φ_A):** we take the 47 distances and the 2 benchmark distances used for Φ_R transformation. The angle between each of these distances and the horizontal line are then computed for the two faces (see Figure 5, right) and stored in a vector, resulting into a feature vector $\Phi_A(\mathbf{L})$ of size $49 \times 2 = 98$.
- **Ratios+Angles ($\Phi_R + \Phi_A$):** in this case, $\Phi_R(\mathbf{L})$ and $\Phi_A(\mathbf{L})$ are simply concatenated, the size of the feature vector being $188 + 98 = 286$.

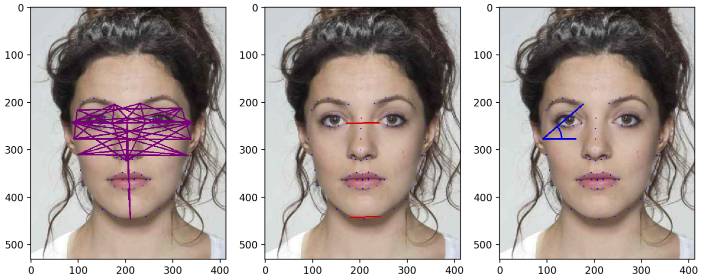


Figure 5. (**Left**) Forty-seven distances used in Φ_R. (**Middle**) Two benchmark distances used in Φ_R. (**Right**) Angle calculation as in Φ_A.

3.2.2. Previously Proposed Landmark-Based Features

As mentioned in Section 2.3, previous approaches in morphing detection have utilized facial landmarks which consist of transformations of the vector **L**:

- **Directed Distances (Φ_{DD}):** proposed in [13], the transformation yields a 136-dimensional vector containing shifting patterns between corresponding landmarks in the two faces.
- **All Distances and Neighbour Angles (Φ_{AD}, Φ_{NA}):** the approach in [14] leads to two transformations: Φ_{AD} calculates a 2278-dimensional feature vector based on distances between all extracted landmarks of a face image; Φ_{NA} only considers angle differences between neighbouring landmarks and yields a 68-dimensional feature vector.

A common trait of these two landmark transformations is that they perform a one-to-one comparison of differente landmarks among the two faces, thus heavily relying on an accurate alignment of the two landmark sets. Instead, Φ_A and Φ_R process the landmark vectors separately for each face (ratios and angles are always computed within the same face) and then concatenate the two feature vectors of every pair. This mitigates potential inaccuracies of the alignment process, for instance, caused by slight pose variations.

4. Experimental Results

We now report the results of our experimental campaign, where the effectiveness of landmark-based geometric detectors is assessed. In Section 4.1, we describe the experimental setup adopted for our tests, including the datasets used, the machine learning classifier, and the evaluation metrics. Section 4.2 reports the results of our approach when the feature vector **L** containing all landmark locations is used for discrimination in different training and testing scenarios. An ablation study on different face areas is performed in Section 4.3, while in Section 4.4, we compare the different landmark transformation approaches described in Section 3.2. Finally, the robustness of the developed detectors in the presence of unknown processing in the testing phase is assessed in Section 4.5.

4.1. Experimental Setup

We used different datasets to create bona fide and attacked image pairs. Since most of the datasets were created for different tasks, in each case, we have selected images with frontal facing subjects exhibiting neutral expressions, according to the structure of each dataset. For the sake of clarity, in the following, we define multiple pair sets.

- Bona-fide pairs:
  - **AR**: 472 pairs formed starting from images in the AR dataset [37]. For every subject, pictures taken in two different acquisitions and distinct poses are available. We selected the 2 available frontal facing images where the face shows neutral expressions from both sessions and paired them with each other.
  - **REPLAY**: 140 pairs formed from frames extracted from the Replay dataset [38], which was originally proposed to benchmark detectors of face spoofing attacks.
  - **MISC**: a collection of 1000 pairs extracted from different datasets, including the Radboud Faces Dataset [39], the CVL Face Database [40], PUT Face Database [41], the FEI Face Database [42], and the Chicago Face Database [43].
- Attacked pairs:
  - **AMSL**: a total of 8700 pairs built from the publicly available AMSL Face Morph Image Dataset [44] used in [11]. A subset $\mathbf{AMSL}_{1000}$ is also determined by randomly selecting 1000 pairs from **AMSL**.
  - **FERET**: 4306 pairs composed from a dataset of morphed images released by Biometix [5]. The morphs have been created starting from images of the Feret database [45], which includes multiple acquisitions of the same subject.

Those sets will be differently combined for creating the training set $\mathcal{TR}$ (i.e., the union of bona fide and attacked training pair sets $\mathcal{TR}_{\text{BF}}$ and $\mathcal{TR}_{\text{A}}$)and the testing set $\mathcal{TS}$ (i.e., the union of bona fide and attacked testing pair sets $\mathcal{TS}_{\text{BF}}$ and $\mathcal{TS}_{\text{A}}$) for supervised machine learning models, as described in the following subsections. The operator $|\cdot|$ will indicate the number of pairs contained in each set.

In each test, an SVM classifier with radial basis function (RBF) kernel has been used for classification. The parameters *gamma* and C of the SVM have been selected via grid-search over a logarithmic grid ranging from 10^{-4} to 10^{1} for each dataset composition. Note that we have focused on the RBF kernel as it always outperformed linear and polynomial kernels in our tests. All the experiments have been performed in Python 3 and the `scikit-learn`, `OpenCV`, and `dlib` packages.

Consistently with other works in this domain, we adopt the metrics defined for the detection of presentation attacks in biometrics to measure the performance of the classification (i.e., thresholding the SVM score at 0):

- *APCER* (Attack Presentation Classification Error Rate): ratio of attacked pairs erroneously classified as bona fide pairs;
- *BPCER* (Bona fide Presentation Classification Error Rate): ratio of bona fide pairs erroneously classified as attacked pairs;
- *ACC* (Accuracy): fraction of image pairs that are correctly classified (either as bona-fide or attacked)

In addition, for selected cases, we show the Detection Error Tradeoff (DET) curve plotting *APCER* vs. *BPCER* obtained by varying the decision threshold on the output score of the SVM, and we will report the *EER* (Equal Error Rate), i.e., the error rate at the operating point where *APCER* = *BPCER*.

4.2. Full Landmark Set

We first test the effectiveness of the feature representation given by the full set of facial landmarks extracted from both images, i.e., the vector $\mathbf{L}$.

We consider different experimental scenarios, always arranging the pairs in such a way that no subject appearing in the testing set is part of any pair used in the training set, not even as a donor of one morphed face. This is in fact the case for real-world applications where we cannot expect the identities in the testing phase to be present in the training set. To this purpose, we define a splitting procedure to form training and testing groups, where we select a part of the subjects appearing in a certain pair set and isolate all the pairs that contain those subjects. Note that the attack pairs consist of a morph and a probe image of one of its donors which was preferably not used during the morphing process. Thus, each morph yields at least 2 attack pairs with 2 images of its different donors. Given $p \in [0,1]$ and a set **SET**, the following steps are performed:

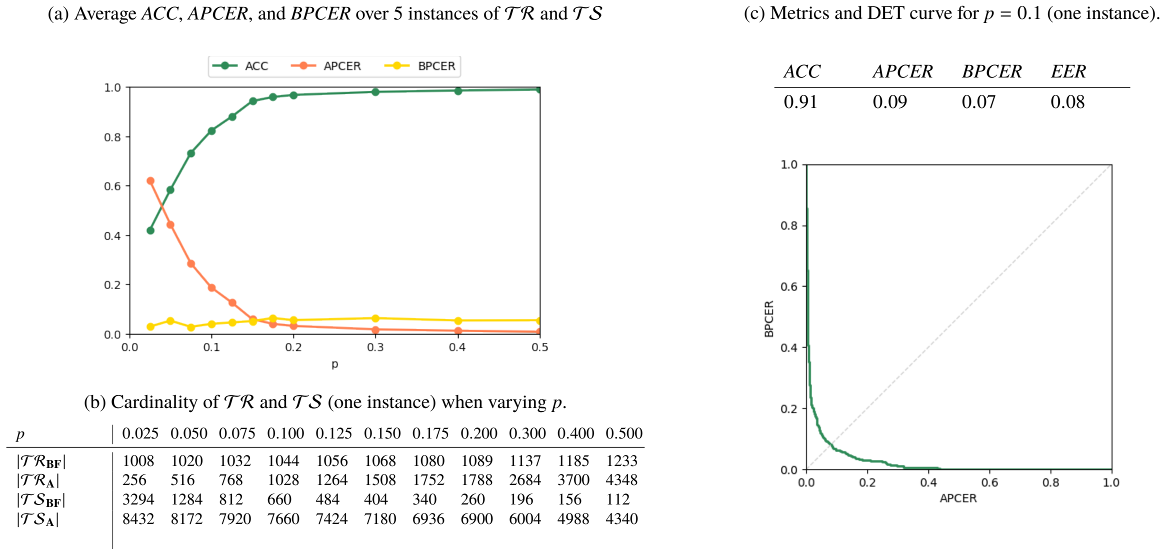
1. a fraction p of the subjects appearing in **SET** are randomly chosen;
2. all the pairs in **SET** which depict any of these subjects in one or both images or as donors of a morphed fac, are stored in $\mathbf{SET}(p)$
3. the remaining pairs in **SET** are stored in $\overline{\mathbf{SET}(p)}$

This procedure has been used to create $\mathcal{TR}$ and $\mathcal{TS}$ by varying p. In particular, we consider three scenarios differing for the composition of $\mathcal{TS}$, as described in Table 1. The bona fide pairs are the same for each row, and the share of **AR** between training and testing varies with p. In the first two scenarios, the attacked pairs in $\mathcal{TR}_\text{A}$ and $\mathcal{TS}_\text{A}$ are drawn from the same pair set. In the third more challenging scenario, $\mathcal{TS}_\text{A}$ is composed by 1000 **AMSL** pairs plus a number of **FERET** pairs (depending on p), while only **FERET** pairs are tested; thus, only a fraction of training samples are from the same set as the testing samples. By doing so, we can observe how performance are affected by the numerosity and composition of $\mathcal{TR}$ and $\mathcal{TS}$.

Table 1. Training/testing scenarios adopted in Section 4.2.

	$\mathcal{TR}_{\mathbf{BF}}$	$\mathcal{TR}_{\mathbf{A}}$	$\mathcal{TS}_{\mathbf{BF}}$	$\mathcal{TS}_{\mathbf{A}}$
AMSL-*only*	$\mathbf{MISC} \cup \mathbf{AR}(p)$	$\mathbf{AMSL}(p)$	$\overline{\mathbf{AR}(p)} \cup \mathbf{REPLAY}$	$\overline{\mathbf{AMSL}(p)}$
FERET-*only*		$\mathbf{FERET}(p)$		$\overline{\mathbf{FERET}(p)}$
Mixed		$\mathbf{AMSL}_{1000} \cup \mathbf{FERET}(p)$		$\overline{\mathbf{FERET}(p)}$

Results for the **AMSL**-*only*, **FERET**-*only*, and *Mixed* scenarios are reported in Figures 6–8, respectively. In Figures 6a–8a, we plot the metrics *ACC*, *APCER*, and *BPCER* for different values of p. Since step 1 of the splitting procedure involves a random choice of a fraction p of subjects (for which the images are then included in $\mathcal{TR}$), the metrics are averaged over 5 different splitting instances for validation. Figures 6b–8b report the cardinality of the resulting training and testing groups for each class on a single splitting instance, for which we report in Figures 6c–8c the DET curve and the performance metrics at $p = 0.1$.



(a) Average *ACC*, *APCER*, and *BPCER* over 5 instances of $\mathcal{TR}$ and $\mathcal{TS}$

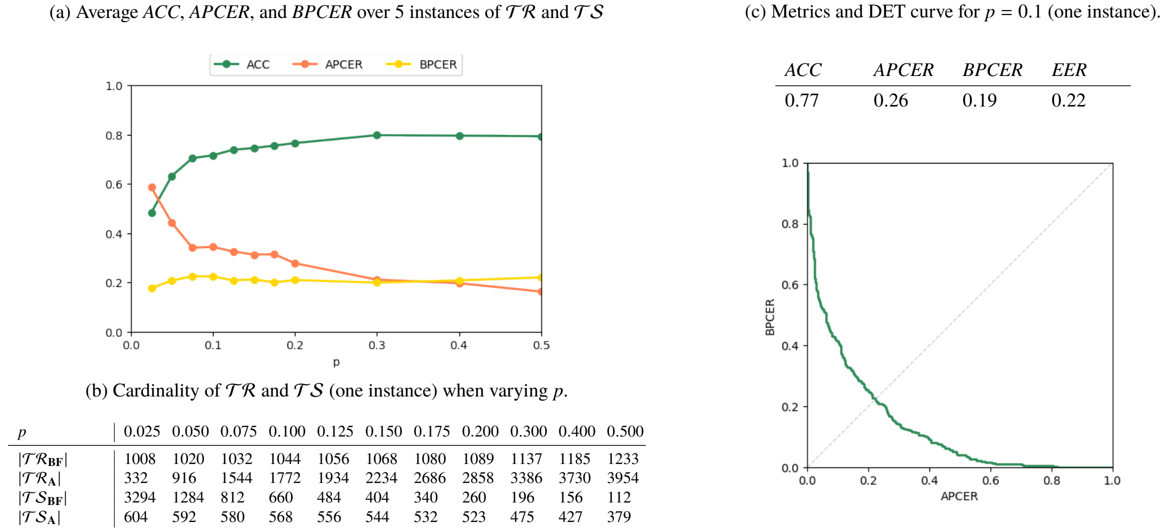
(b) Cardinality of $\mathcal{TR}$ and $\mathcal{TS}$ (one instance) when varying p.

p	0.025	0.050	0.075	0.100	0.125	0.150	0.175	0.200	0.300	0.400	0.500
$\lvert\mathcal{TR}_{\mathbf{BF}}\rvert$	1008	1020	1032	1044	1056	1068	1080	1089	1137	1185	1233
$\lvert\mathcal{TR}_{\mathbf{A}}\rvert$	256	516	768	1028	1264	1508	1752	1788	2684	3700	4348
$\lvert\mathcal{TS}_{\mathbf{BF}}\rvert$	3294	1284	812	660	484	404	340	260	196	156	112
$\lvert\mathcal{TS}_{\mathbf{A}}\rvert$	8432	8172	7920	7660	7424	7180	6936	6900	6004	4988	4340

(c) Metrics and DET curve for p = 0.1 (one instance).

ACC	*APCER*	*BPCER*	*EER*
0.91	0.09	0.07	0.08

Figure 6. Results for the **AMSL**-*only* scenario.



(a) Average *ACC*, *APCER*, and *BPCER* over 5 instances of $\mathcal{TR}$ and $\mathcal{TS}$

(b) Cardinality of $\mathcal{TR}$ and $\mathcal{TS}$ (one instance) when varying p.

p	0.025	0.050	0.075	0.100	0.125	0.150	0.175	0.200	0.300	0.400	0.500
$\lvert\mathcal{TR}_{\mathbf{BF}}\rvert$	1008	1020	1032	1044	1056	1068	1080	1089	1137	1185	1233
$\lvert\mathcal{TR}_{\mathbf{A}}\rvert$	332	916	1544	1772	1934	2234	2686	2858	3386	3730	3954
$\lvert\mathcal{TS}_{\mathbf{BF}}\rvert$	3294	1284	812	660	484	404	340	260	196	156	112
$\lvert\mathcal{TS}_{\mathbf{A}}\rvert$	604	592	580	568	556	544	532	523	475	427	379

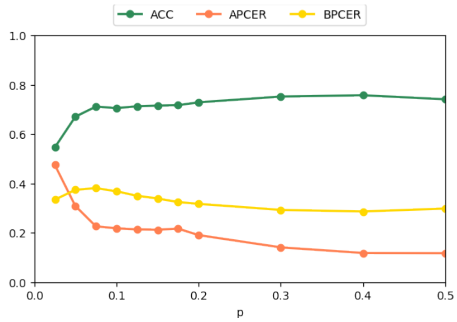
(c) Metrics and DET curve for p = 0.1 (one instance).

ACC	*APCER*	*BPCER*	*EER*
0.77	0.26	0.19	0.22

Figure 7. Results for the **FERET**-*only* scenario.

As expected, the performance increases with p, i.e., when more numerous and representative training samples are available. Overall, the best results are obtained in the **AMSL**-*only* scenario, with a global accuracy approaching 1 for $p > 0.2$. The **FERET**-*only* scenario instead shows a lower accuracy, which stabilizes at around 0.77 even when p increases. This is explained by the higher variability of acquisition conditions of the probe images in the **FERET** pairs, which makes it harder to discriminate face geometry anomalies due to variabilities in the probe images or due to morphing operations, thus causing increased *APCER*, *BPCER*, and *EER*.

(a) Average *ACC*, *APCER*, and *BPCER* over 5 instances of $\mathcal{TR}$ and $\mathcal{TS}$



(b) Cardinality of $\mathcal{TR}$ and $\mathcal{TS}$ (one instance) when varying *p*.

p	0.025	0.050	0.075	0.100	0.125	0.150	0.175	0.200	0.300	0.400	0.500
$\lvert\mathcal{TR}_{\mathbf{BF}}\rvert$	1008	1020	1032	1044	1056	1068	1080	1089	1137	1185	1233
$\lvert\mathcal{TR}_{\mathbf{A}}\rvert$	1332	1916	2544	2772	2934	3234	3686	3858	4386	4730	4954
$\lvert\mathcal{TS}_{\mathbf{BF}}\rvert$	3294	1284	812	660	484	404	340	260	196	156	112
$\lvert\mathcal{TS}_{\mathbf{A}}\rvert$	604	592	580	568	556	544	532	523	475	427	379

(c) Metrics and DET curve for p = 0.1 (one instance).

ACC	*APCER*	*BPCER*	*EER*
0.76	0.17	0.32	0.22

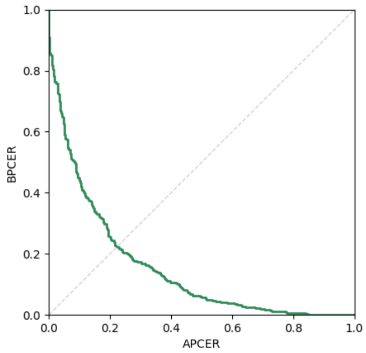


Figure 8. Results for the *Mixed* scenario.

Finally, the *Mixed* scenario results show that, when trained mostly on the 1000 **AMSL** pairs, the detector struggles in recognizing **FERET** attacked pairs and performs well when samples in $\mathcal{TR}_{\text{A}}$ are roughly equally splitted in **AMSL** and **FERET** pairs. This suggests that different datasets carry peculiar characteristics and, as it typically happens for supervised machine learning solutions, there exists a risk of overfitting on specific sources.

For the sake of completeness, we have also investigated the use of a 1D convolutional neural network (CNN) classifier to better process the information contained in the landmark vector. We considered an architecture with 4 1-dimensional convolutional layers with a kernel size of 3 and $[64, 128, 256, 256]$ filters each. Succeeding the second convolution, we apply instance normalisation after every feature extraction layer. Before the classification, we apply a dense layer with 128 neurones.

In Table 2, we provide a comparison between the RBF SVM and the 1D CNN classifier for the *Mixed* scenario and $p = 0.1$ (one instance) both in terms of performance and training/testing time. In fact, we report the average training time over different values of p and the average prediction time for the two classifiers; for the SVM, tests have been conducted on a 2.3 GHz 8-core Intel Core i9, while the CNN was trained and tested on an NVIDIA GTX 1080Ti GPU.

As it can be observed, the gain in performance with respect to the RBF SVM is rather marginal, in front of a much higher computational effort. We therefore employ the RBF SVM for the following analyses. Moreover, for the sake of brevity, we will stick to the *Mixed* scenario and the case $p = 0.1$.

Table 2. Performance metrics of different classifiers: the average training time is computed as the mean of training times over distinct values of p. We define the average prediction time as the mean of the time (measured over 100 examples) that it takes for our models to classify one image pair.

Model	***ACC***	***APCER***	***BPCER***	**Average Training Time per *p***	**Average Prediction Time per Pair**
RBF SVM	0.76	0.17	0.32	0.69 min	0.0031 s
1D CNN	0.77	0.19	0.29	36.08 min	0.1768 s

4.3. Ablation Study

In order to determine the importance of different landmarks, we group them into distinct semantic groups, as shown in Figure 9, and observe their detection results. These groups correspond to facial attributes and are inspired by the semantic landmark groupings in [46].

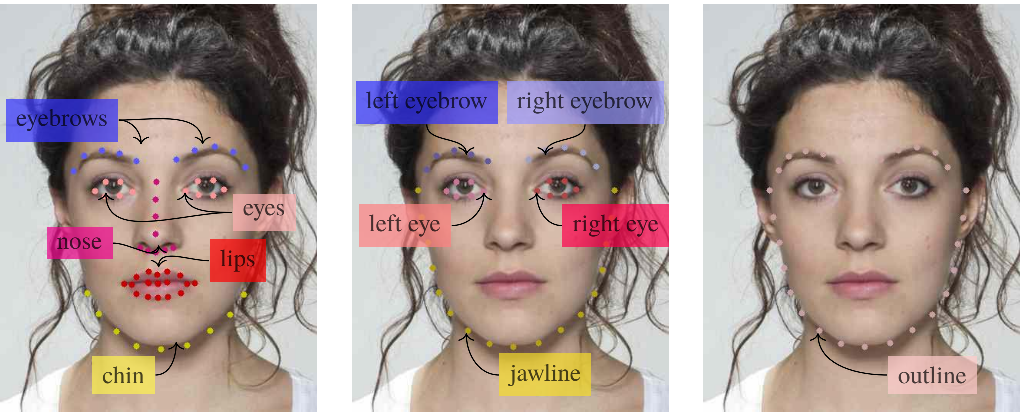


Figure 9. Landmark groups for our ablation study.

We separately test landmarks corresponding to different semantic groups for each image pair and concatenate them to obtain a feature vector $\mathbf{sL}_g$, where g indicates a single semantic group or a combination of them. We then feed $\mathbf{sL}_g$ to a RBF SVM, just like we did for $\mathbf{L}$.

The results are reported in Table 3. It can be observed that $\mathbf{L}$ is generally better performing. However, $\mathbf{sL}_{outline}$ achieves comparable results, thus suggesting that most of the relevant geometric information resides in the relative position of the face line and the eyes.

Moreover, it is worth noticing that the accuracy drop of different variants of $\mathbf{sL}_g$ is mostly due an increase in *BPCER* while the *APCER* remains rather low. This bias towards false alarms might be due to the selected features being less distinctive and the training set not containing enough information for characterizing non-attacked samples, so that bona fide pairs likely exhibit unseen patterns at testing time and are classified as attacked.

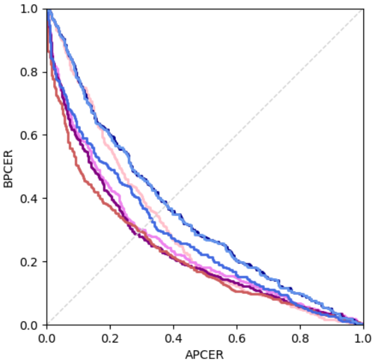
Table 3. Results for the ablation tests.

Feature Representation	*ACC*	*APCER*	*BPCER*	*EER*
$\mathbf{L}$	0.67	0.17	0.32	0.22
$\mathbf{sL}_{\text{eyes}}$	0.37	0.07	0.77	0.39
$\mathbf{sL}_{\text{left eye}}$	0.11	0.06	0.93	0.40
$\mathbf{sL}_{\text{right eye}}$	0.25	0.05	0.85	0.42
$\mathbf{sL}_{\text{eyebrows}}$	0.51	0.07	0.73	0.38
$\mathbf{sL}_{\text{left eyebrow}}$	0.32	0.01	0.89	0.35
$\mathbf{sL}_{\text{right eyebrow}}$	0.36	0.02	0.85	0.41
$\mathbf{sL}_{\text{eyebrows + eyes}}$	0.52	0.16	0.57	0.34
$\mathbf{sL}_{\text{left eyebrow + eye}}$	0.37	0.08	0.77	0.36
$\mathbf{sL}_{\text{right eyebrow + eye}}$	0.46	0.09	0.75	0.39
$\mathbf{sL}_{\text{nose}}$	0.21	0.11	0.83	0.43
$\mathbf{sL}_{\text{lips}}$	0.44	0.15	0.48	0.30
$\mathbf{sL}_{\text{chin}}$	0.20	0.05	0.79	0.39
$\mathbf{sL}_{\text{jawline}}$	0.41	0.06	0.69	0.34
$\mathbf{sL}_{\text{outline}}$	0.64	0.09	0.38	0.23

4.4. Comparison of Landmark Transformations

We now compare the performance of the different feature representations derived from the landmark location vector $\mathbf{L}$ introduced in Section 3.2, comparing them with other transformations proposed in the literature.

We report the results for the case of $p = 0.1$ in Figure 10. On the left, the DET curves for the different transformations are reported, and for each of them, the performance metrics are reported on the right. Higher *ACC* and *EER* are obtained by the anthropometry-based transformations $\Phi_R + \Phi_A$ and Φ_A. In general, the angle-based features in $\Phi_A(\mathbf{L})$ are more informative than the ratio-based ones in $\Phi_R(\mathbf{L})$, although their dimensionality is lower than all the other considered transformations. Φ_{DD} also has competitive performance, but its *APCER* and *BPCER* are strongly unbalanced. Φ_{AD}, Φ_{NA}, and their combinations yield less accurate classifications.



Color	Feature representation	*ACC*	*APCER*	*BPCER*	*EER*
▬	$\Phi_R(\mathbf{L})$	0.64	0.21	0.54	0.35
▬	$\Phi_A(\mathbf{L})$	0.70	0.27	0.33	0.30
▬	$\Phi_R(\mathbf{L}) + \Phi_A(\mathbf{L})$	**0.72**	0.26	0.31	**0.29**
▬	$\Phi_{DD}(\mathbf{L})$ [9]	0.68	0.06	0.62	0.30
▬	$\Phi_{AD}(\mathbf{L})$ [42]	0.63	0.06	0.72	0.34
▬	$\Phi_{NA}(\mathbf{L})$ [42]	0.59	0.11	0.75	0.38
▬	$\Phi_{AD}(\mathbf{L}) + \Phi_{NA}(\mathbf{L})$ [42]	0.59	0.11	0.75	0.38

Figure 10. Results for different feature representations.

However, note that all feature representations underperform with respect to $\mathbf{L}$. This suggests that, in the considered experimental scenario, the SVM model is powerful enough to learn effective classification boundaries directly in the feature space of $\mathbf{L}$, and further, handcrafted transformations are not beneficial in terms of global accuracy.

4.5. Robustness to Processing Operations

We assess the robustness of our detector in the case of diverse processing operations applied to the eMRTD photo. This is in fact known to be a typical issue of for previous detectors, especially single-image ones, as passport pictures can undergo several operations in its digital history (e.g., printing/scanning and compression). To this purpose, we run our models also on different variants of the testing set, where selected postprocessing operations are applied to the passport photos as listed and described in Table 4.

Examples of the different processing operations are reported in Figure 11, where a portion of the image is magnified.

In each case, we measure the performance loss with respect to the baseline case, where neither training nor testing underwent any processing. If *ACC* is the accuracy in the baseline case and ACC_P is the accuracy when a certain processing P is applied to the testing set, we calculate the accuracy loss as

$$ACC_{\text{Loss}} = ACC - ACC_{\text{P}}. \quad (1)$$

Figure 12 reports ACC_{Loss} for each processing operation and for the feature representations $\mathbf{L}$, $\Phi_R(\mathbf{L})$, $\Phi_A(\mathbf{L})$, and $\Phi_R(\mathbf{L}) + \Phi_A(\mathbf{L})$.

Table 4. Manipulations applied for the robustness test.

Name	**Description**
Noise	Additive Gaussian noise with $\sigma = 0.5$
Blur	Blurring with normalized box filter
Scaling V	Downscaling the vertical dimension by 1–2%
Scaling H	Downscaling the horizontal dimension by 1–2%
Affine 1	Applying small offsets to three selected landmarks and the corresponding affine transform to the whole image
Affine 2	Applying a small offset to one selected landmark and the corresponding affine transform to the whole image
Rotation	Rotating the image by ±3% degrees
Speckle	Multiplicative noise
Salt and pepper	Punctual noise on 4% of pixels

We can see that the accuracy loss is always below 5% and involves mostly angle-based feature representations. The loss for the full landmark feature vector $\mathbf{L}$ is however very small (always below 2%) and essentially oscillates around 0. We can then conclude that the trained models generally preserve their effectiveness also in the presence of these unseen processing operations appearing in the testing set.

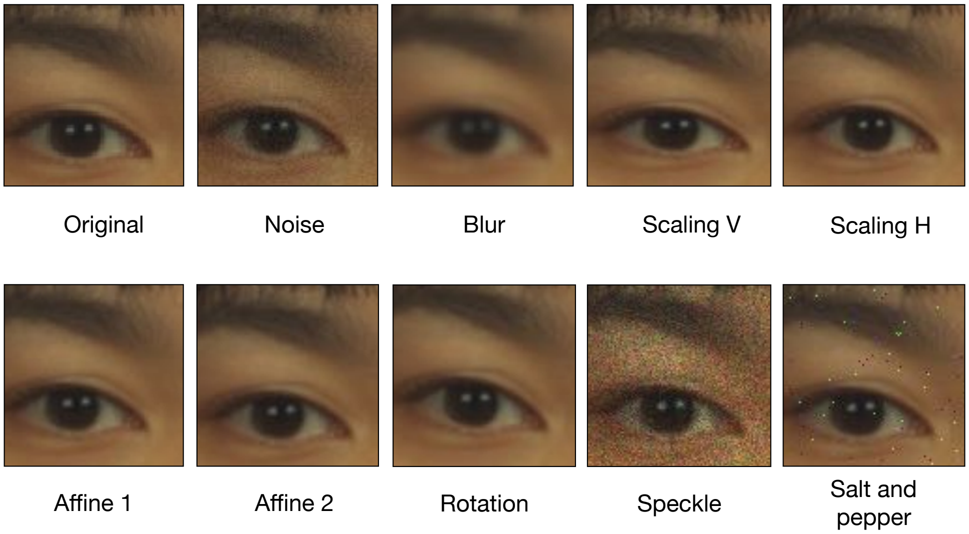


Figure 11. Example of processed electronic Machine Readable Travel Documents (eMRTD) pictures with different manipulations.

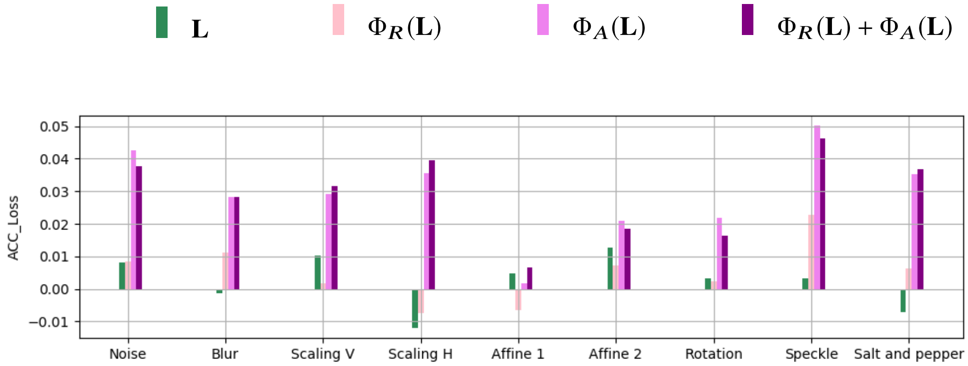


Figure 12. Values of ACC_P for different processing operations and feature representation.

5. Conclusions

We have addressed the problem of detecting morphed faces in electronic passports at border control in a differential scenario, i.e., by jointly analyzing the photo contained in the electronic passport and the live probe image acquired on site. In doing so, we have performed a comparative analysis of geometric face features by developing detectors based on the facial landmarks and by exploring their effectiveness in different directions.

In different scenarios, best results are obtained by operating directly in the feature space of the 2D coordinates of the 68 facial landmarks extracted from the two face images of the pair under investigation. The performance remains essentially stable even when the testing samples are modified via processing operations that are unseen in the training phase. This confirms the advantage of relying on geometric cues like landmarks, for which extraction is generally reliable even after visual alterations that are not too impactful.

Moreover, ablation tests suggests that a non-processed full set of landmark coordinates provides more discriminative information in every case. Among the compared handcrafted features, the ones based on facial anthropometry concepts are generally more effective with respect to approaches previously proposed in the literature.

The obtained results confirm the potential of a geometric differential analysis leveraging also the probe image for detecting morphing attacks. The extracted features are indeed limited in dimensionality (thus are lighter to process with respect to more computationally expensive techniques [32]), while offering fair detection performance and high interpretability of the detector's outcome. This is an advantage with respect to other differential detection approaches based on deep networks [33], which do not explicitly look for geometric distortions that are inherent to morphing attacks but rather rely on the distribution of deep features used for general face-recognition problems. However, our study also exposes typical issues affecting supervised machine learning techniques, namely the risk of overfitting training data and reduced generalization abilities when different data sources are tested. Multi-clue detectors would in fact be recommended for improved performance in realistic scenarios. In fact, a promising direction for future work would be to analyze geometric cues in conjuction with richer representations like the ones based on deep networks [33], which has brought a significant performance boost in many related tasks.

Author Contributions: Conceptualization, C.P. and S.A.; methodology, C.P.; software, S.A.; validation, S.A.; data curation, S.A.; writing–original draft preparation, S.A.; writing–review and editing, C.P. All authors have read and agreed to the published version of the manuscript.

Funding: The second author was partially supported by the project PREMIER (PREserving Media trustworthiness in the artificial Intelligence ERa), funded by the Italian Ministry of Education, University, and Research (MIUR) within the PRIN 2017 program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; BMVA Press: Durham, UK, 2015; pp. 41.1–41.12.
2. Balaban, S. Deep learning and face recognition: The state of the art. In *Proceedings Volume 9457 Biometric and Surveillance Technology for Human and Activity Identification XII*; SPIE: Bellingham, WA, USA, 2015. [CrossRef]
3. Neubert, T.; Kraetzer, C.; Dittmann, J. A Face Morphing Detection Concept with a Frequency and a Spatial Domain Feature Space for Images on eMRTD. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 3–5 July 2019; pp. 95–100.
4. Ferrara, M.; Franco, A.; Maltoni, D. The magic passport. In Proceedings of the IEEE International Joint Conference on Biometrics, 2014, Clearwater, FL, USA, 29 September–2 October 2014; pp. 1–7. [CrossRef]
5. Biometix Pty Ltd. New Face Morphing Dataset (for Vulnerability Research). 2018. Available online: http://www.biometix.com/2017/09/18/new-face-morphing-dataset-for-vulnerability-research/ (accessed on 1 December 2019.)
6. Makrushin, A.; Neubert, T.; Dittmann, J. Automatic Generation and Detection of Visually Faultless Facial Morphs. In Proceedings of the Internation Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Porto, Portugal, 27 February–1 March 2017; pp. 39–50. [CrossRef]
7. Scherhag, U.; Rathgeb, C.; Busch, C. Performance variation of morphed face image detection algorithms across different datasets. In Proceedings of the 2018 International Workshop on Biometrics and Forensics (IWBF), Sassari, Italy, 7–8 June 2018; pp. 1–6. [CrossRef]
8. Pasquini, C.; Schöttle, P.; Böhme, R.; Boato, G.; Pèrez-Gonzàlez, F. Forensics of High Quality and Nearly Identical JPEG Image Recompression. In Proceedings of the ACM Information Hiding and Multimedia Security Workshop, Vigo, Spain, 20–22 June 2016; pp. 11–21.
9. Shang, S.; Kong, X. Printer and Scanner Forensics. In *Handbook of Digital Forensics of Multimedia Data and Devices*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015; Chapter 10; pp. 375–410.
10. Pasquini, C.; Böhme, R. Information-Theoretic Bounds for the Forensic Detection of Downscaled Signals. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1928–1943. [CrossRef]
11. Neubert, T.; Makrushin, A.; Hildebrandt, M.; Krätzer, C.; Dittmann, J. Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biom.* **2018**, *7*, 325–332. [CrossRef]
12. Scherhag, U.; Rathgeb, C.; Busch, C. Morph Detection from Single Face Image: A Multi-Algorithm Fusion Approach. In *Proceedings of the 2018 2Nd International Conference on Biometric Engineering and Applications*; ICBEA '18; ACM: New York, NY, USA, 2018; pp. 6–12. [CrossRef]
13. Damer, N.; Boller, V.; Wainakh, Y.; Boutros, F.; Terhörst, P.; Braun, A.; Kuijper, A. Detecting Face Morphing Attacks by Analyzing the Directed Distances of Facial Landmarks Shifts. In *Pattern Recognition*; Brox, T., Bruhn, A., Fritz, M., Eds.; Springer International Publishing: Berlin, Germany, 2019; pp. 518–534.
14. Scherhag, U.; Budhrani, D.; Gomez-Barrero, M.; Busch, C. Detecting Morphed Face Images Using Facial Landmarks. In *Image and Signal Processing*; Springer International Publishing: Cham, Switzerland, 2018; pp. 444–452. [CrossRef]
15. Scherhag, U.; Rathgeb, C.; Merkle, J.; Breithaupt, R.; Busch, C. Face Recognition Systems Under Morphing Attacks: A Survey. *IEEE Access* **2019**, *7*, 23012–23026. [CrossRef]
16. FaceMorpher. Available online: https://github.com/stheakanath/facemorpher (accessed on 27 July 2020).
17. FaceMorpher Luxand. Available online: https://www.luxand.com/facemorpher/ (accessed on 27 July 2020).
18. Damer, N.; Saladié, A.M.; Braun, A.; Kuijper, A. MorGAN: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Generative Adversarial Network. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Los Angeles, CA, USA, 22–25 October 2018; pp. 1–10. [CrossRef]

19. Venkatesh, S.; Zhang, H.; Ramachandra, R.; Raja, K.B.; Damer, N.; Busch, C. Can GAN Generated Morphs Threaten Face Recognition Systems Equally as Landmark Based Morphs?-Vulnerability and Detection. In Proceedings of the IEEE International Workshop on Biometrics and Forensics (IWBF), Porto, Portugal, 29–30 April 2020; pp. 1–6.
20. Scherhag, U.; Rathgeb, C.; Busch, C. Towards Detection of Morphed Face Images in Electronic Travel Documents. In Proceedings of the 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 24–27 April 2018; pp. 187–192. [CrossRef]
21. Wandzik, L.; Kaeding, G.; Garcia, R.V. Morphing Detection Using a General- Purpose Face Recognition System. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1012–1016.
22. Jassim, S.; Asaad, A. Automatic Detection of Image Morphing by Topology-based Analysis. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1007–1011.
23. Rashid, R.D.; Asaad, A.; Jassim, S. Topological data analysis as image steganalysis technique. In Proceedings of the Mobile Multimedia/Image Processing, Security, and Applications, Bellingham, WA, USA, 16–17 April 2018; Volume 10668, pp. 103–111.
24. Kraetzer, C.; Makrushin, A.; Neubert, T.; Hildebrandt, M.; Dittmann, J. Modeling Attacks on Photo-ID Documents and Applying Media Forensics for the Detection of Facial Morphing. In Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, Philadelphia, PA, USA, 20–21 June 2017; IH&MMSec '17; ACM: New York, NY, USA, 2017; pp. 21–32. [CrossRef]
25. Debiasi, L.; Rathgeb, C.; Scherhag, U.; Uhl, A.; Busch, C. PRNU Variance Analysis for Morphed Face Image Detection. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Los Angeles, CA, USA, 22–25 October 2018; pp. 1–9. [CrossRef]
26. Scherhag, U.; Debiasi, L.; Rathgeb, C.; Busch, C.; Uhl, A. Detection of Face Morphing Attacks Based on PRNU Analysis. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *1*, 302–317. [CrossRef]
27. Makrushin, A.; Kraetzer, C.; Neubert, T.; Dittmann, J. Generalized Benford's Law for Blind Detection of Morphed Face Images. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Innsbruck, Austria, 20–22 June 2018; pp. 49–54. [CrossRef]
28. Seibold, C.; Hilsmann, A.; Eisert, P. Reflection Analysis for Face Morphing Attack Detection. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1022–1026.
29. Seibold, C.; Samek, W.; Hilsmann, A.; Eisert, P. Detection of Face Morphing Attacks by Deep Learning. In *Digital Forensics and Watermarking*; Kraetzer, C., Shi, Y.Q., Dittmann, J., Kim, H.J., Eds.; Springer International Publishing: Berlin, Germany, 2017; pp. 107–120.
30. Raghavendra, R.; Raja, K.B.; Venkatesh, S.; Busch, C. Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1822–1830. [CrossRef]
31. Ferrara, M.; Franco, A.; Maltoni, D. Face Demorphing. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1008–1017. [CrossRef]
32. Peng, F.; Zhang, L.; Long, M. FD-GAN: Face-demorphing generative adversarial network for restoring accomplice's facial image. *IEEE Access* **2019**, *7*, 75122–75131. [CrossRef]
33. Scherhag, U.; Rathgeb, C.; Merkle, J.; Busch, C. Deep Face Representations for Differential Morphing Attack Detection. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3625–3639. [CrossRef]
34. Farkas, L.G.; Munro, I.R. Anthropometric facial proportions in medicine. *Am. J. Orthod. Dentofac. Orthop.* **1987**, *92*, 522.
35. Gupta, S.; Markey, M.; Bovik, A.C. Anthropometric 3D Face Recognition. *Int. J. Comput. Vis.* **2010**, *90*, 331–349. [CrossRef]
36. Shi, J.; Samal, A.; Marx, D. How Effective Are Landmarks and Their Geometry for Face Recognition? *Comput. Vis. Image Underst.* **2006**, *102*, 117–133. [CrossRef]
37. Martinez, A.; Benavente, R. The AR Face Database. *Tech. Rep. CVC Tech. Rep.* 1998 Available online: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html (accessed on 1 March 2020)

38. Chingovska, I.; Anjos, A.; Marcel, S. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. In Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 6–7 September 2012.
39. Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D.; Hawk, S.; Knippenberg, A. Presentation and validation of the Radboud Face Database. *Cogn. Emot. Cogn. Emot.* **2010**, *24*, 1377–1388. [CrossRef]
40. Peer, P. CVL Face Database. 2010. Available online: http://lrv.fri.uni-lj.si/facedb.html (accessed on 1 March 2020).
41. Kasiński, A.; Florek, A.; Schmidt, A. The PUT face database. *Image Process. Commun.* **2008**, *13*, 59–64.
42. Thomaz, C.; Giraldi, G. A new ranking method for Principal Components Analysis and its application to face image analysis. *Image Vis. Comput.* **2010**, *28*, 902–913. [CrossRef]
43. Ma, D.; Correll, J.; Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **2015**, *47*. [CrossRef] [PubMed]
44. AMSL Face Morph Image Data Set. 2018. Available online: https://omen.cs.uni-magdeburg.de/disclaimer/index.php (accessed on 1 December 2019).
45. Phillips, P.J.; Hyeonjoon Moon.; Rizvi, S.A.; Rauss, P.J. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1090–1104. [CrossRef]
46. Dabouei, A.; Soleymani, S.; Dawson, J.M.; Nasrabadi, N.M. Fast Geometrically-Perturbed Adversarial Faces. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa Village, HI, USA, 7–11 January 2019. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.