IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 6, 2020, accepted May 28, 2020, date of publication June 5, 2020, date of current version June 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000254

Differential Detection of Facial Retouching: A Multi-Biometric Approach

C. RATHGEB[1], C.-I. SATNOIANU[2], N. E. HARYANTO[1], K. BERNARDO[1], AND C. BUSCH[1], (Senior Member, IEEE)

[1]da/sec–Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany
[2]Department of Information Security and Communication Technology, Danmarks Tekniske Universitet, Lyngby, Denmark

Corresponding author: C. Rathgeb (christian.rathgeb@h-da.de)

This work was supported in part by the German Federal Ministry of Education and Research, and in part by the Hessen State Ministry for Higher Education, Research, and the Arts, through the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT Facial retouching apps have become common tools which are frequently applied to improve one's facial appearance, *e.g.* before sharing face images via social media. Beautification induced by retouching has the ability to substantially alter the appearance of face images and hence might represent a challenge for face recognition. Towards deploying secure face recognition as well as enforcing anti-photoshop legislations, a robust and reliable detection of retouched face image is needed. Published approaches consider a single image-based (no-reference) scenario where a potentially retouched face image serves as sole input to the retouching detector. However, in many cases a trusted unaltered face image of a subject examined is available which enables an image pair-based (differential) detection scheme. In this work, ICAO-compliant subsets of the FERET and FRGCv2 face databases are used to automatically create a database containing 9,078 retouched face images together with unconstrained probe images. In evaluations employing the commercial Cognitec FaceVACS and the open-source ArcFace face recognition system, it is shown that facial retouching can negatively impact face recognition performance. Further, a differential facial retouching detection system is proposed which processes pairs of a potentially retouched reference image and corresponding unaltered probe image of single subjects. Estimated differences in feature vectors obtained from texture descriptors, facial landmarks, and deep face representations are leveraged by machine learning-based classifiers of which the detection scores are fused to distinguish between retouched and unaltered face images. The proposed scheme is evaluated in a cross-database scenario where training and testing are performed on the FERET and FRGCv2 databases and vice versa. In the scenario where the used retouching algorithm is known by the detection algorithm, a competitive average D-EER of approximately 2% is achieved. Further, the scenario in which the employed retouching algorithm is not known by the detection algorithm is evaluated. In the latter scenario, the proposed approach obtains an average D-EER below 10% and is shown to outperform several state-of-the-art single image-based detection schemes.

INDEX TERMS Biometrics, face recognition, facial retouching, beautification, differential detection.

I. INTRODUCTION

Face recognition has been an active field of research for several decades [1]–[4]. In the past years, the introduction of deep convolutional neural networks has shown impressive performance improvements in facial recognition technologies [4]–[9]. A number of covariates have been identified that can negatively affect recognition accuracy, *e.g.* fluctuations in pose, facial expression, or image quality [3], [10]. Additionally, *facial beautification* [11] induced by plastic surgery, cosmetics, or beautification in the digital domain, *i.e.* *retouching*, was determined to be able to significantly alter the perceived shape and texture of a human face and therefore to negatively affect the accuracy of face recognition systems.

Facial retouching causes alterations similar to those achieved by plastic surgery or makeup. Beyond that, further changes can be made to face images in the digital domain, *e.g.* enlarging of the eyes. Besides professional image editing software, *e.g.* Photoshop, there exist plenty of mobile applications, *i.e.* apps, which provide dozens of filters and special beautification effects that can be applied easily even by unskilled users. Such apps might as well be employed

The associate editor coordinating the review of this manuscript and approving it for publication was Marina Gavrilova.

to reduce the fish-eye effect or unwanted front-facing camera lens distortions [20]. Fig. 1 shows examples of facial retouching.

(a) FotoRus [12]

(b) InstaBeauty [13]

FIGURE 1. Example images before (left) and after (middle) facial retouching as well as (right) main differences for (a) a female and (b) a male face image using different mobile retouching apps. Shown alterations include slimming of the face, chin and nose, smoothing of the skin, enlarging of the eyes, removal of dark eye circles and skin impurities.

Retouching can have an impact on various scenarios where facial recognition technologies are used. If face recognition is applied to images from social media such as Facebook or Instagram, *e.g.* as part of a forensic investigation, the use of retouching is very likely. Nowadays, more and more facial images are taken with smartphones, *e.g.* by making ''selfies'' [21]. For best results, users often edit these images before sharing them. This use case can be of great importance for face recognition technologies in the future, considering the increasing use of social media and the amount of retouching applications available. Moreover, in many countries, the photo used for issuing electronic travel documents is provided by the applicant. That is, various types of image editing, including facial retouching, can be performed prior to the issuance and hence negatively affect the performance of a facial recognition system, *e.g.* for automated border control.

Deviating from the above mentioned scenarios, the necessity of a reliable recognition of digitally beautified facial images will be increased by the introduction of the so-called ''Photoshop law'' [22]. People's behavior is often influenced through advertising based on digitally manipulated images and, as a result, their preferences are often badly formed. In response, the State of Israel passed a law in 2014 to mitigate the dangers of growing eating disorders caused by digitally retouched images in advertising. A similar law has been in force in France since 2017, while several other countries, *e.g.* Belgium, Spain, Italy, and Germany, discuss appropriate regulations and laws, too. Consequently, digitally retouched photos must be marked as ''edited photo'' [23]. This means that even if facial recognition systems achieve robustness to facial retouching, reliable detection systems are still required as a tool for enforcing this type of legislation.

Besides retouching, further image manipulation techniques can be applied to digitally change the appearance of face images including replacement or reenactment [24], [25], which are frequently referred to as ''face swapping'' or ''deep-fakes'', and morphing [26], [27]. It was found that human observers achieve only low accuracy in detecting said types of face image manipulations [28], [29] including facial retouching [18]. Further, different benchmarks [29], [30] have been conducted to compare the performance of published detection schemes.

In this work, subsets of two public available face databases are used to automatically create a database of retouched face images. The database comprises more than nine thousand images generated by six different retouching apps. Two state-of-the-art face recognition systems (open-source and commercial) are employed to investigate the impact of facial retouching on face recognition performance. An image pair-based, *i.e. differential*, facial retouching detection system is introduced which takes as input a potentially retouched reference image and an unaltered but unconstrained probe image. This scenario, which allows the estimation of differences between a processed image pair, is motivated by the assumption that in many real-world scenarios, *e.g.* automated border control, it is plausible that at least one other unaltered image of a depicted subject is available during detection. Detection scores are obtained from machine learning-based classifiers analyzing differences in texture descriptors, facial landmarks, and deep face representations. A fusion of detection scores is performed in order to distinguish between unaltered, *i.e. bona fide*, and retouched face images. In cross-database experiments, the scenarios in which the applied retouching app is known and unknown is evaluated. In the latter case, which has been hardly considered in related works, the proposed retouching detection system significantly outperforms several published single image-based approaches.

The remainder of this paper is organized as follows: related works are revisited in Sect. II. Subsequently, the image databases used in this work are described in detail in Sect. III. The proposed differential retouching detection approach is presented in Sect. IV. Conducted experiments are summarized in Sect. V and conclusions are drawn in Sect. VI.

II. RELATED WORK

Table 1 lists the most important works examining the effects of facial retouching on facial recognition, along with proposed detection systems, used databases, applied methods, and reported results. Performance rates are mostly reported using standardized metrics for measuring biometric performance [31], *e.g.* Equal Error Rate (EER) or Rank-1 Identification Rate (R-1). For detection schemes the Correct

TABLE 1. Most relevant works on the impact and detection of facial retouching in face recognition (adapted from [11]).

Reference	Database	Method(s)	Performance rates		Retouching Detection	Remarks
			Unaltered	Retouched		
Ferrara *et al.* [14]	AR face (118 subjects)	2× COTS, SIFT	∼0% EER (COTS)	∼2%,∼5%,∼17% EER for low/medium/high intensity (COTS)	–	3 intensities of retouching with LiftMagic, small amount of comparisons
Bharati *et al.* [15]	ND-IIITD Retouched Faces (325 subjects), Celebrity (165 subjects)	Recognition: COTS, OpenBR Detection: patch-based deep supervised RBM with SVM	100% R-1 (COTS)	97.67% R-1 (average, COTS)	87.1% CCR on ND-IIITD Retouched Faces, 96.2% CCR on Celebrity	7 types of retouching with PortraitPro Studio Max
Bharati *et al.* [16]	Multi-Demographic Retouched Faces (600 subjects)	Sub-class supervised sparse Autoencoder	–	–	94.3% CCR (on average)	2 types of retouching with PortraitPro Studio Max and BeautyPlus
Jain *et al.* [17]	ND-IIITD Retouched Faces	CNN with SVM	–	–	99.65% CCR	–
Wang *et al.* [18]	Automatically generated based on OpenImage and Flickr (1.1M face images)	Dilated Residual Network	–	–	90% CCR	Detection of Photoshop image warping operation, manually created test set
Rathgeb *et al.* [19]	Manually generated based on FRGCv2 (100 subjects)	Recognition: COTS Detection: PRNU analysis	0% FNMR at 0.1% FMR	0% FNMR at 0.1% FMR	13.7% D-EER	5 types of retouching with mobile apps

Classification Rate (CCR), which corresponds to the Detection Equal Error Rate (D-EER), is frequently used.

Ferrara *et al.* [14], [32] were the first to measure the influence of digital beautification on facial recognition systems. Among other image manipulation techniques, such as geometric distortion or morphing, they reported significant performance degradation for various facial recognition systems after the application of strong facial retouching. These findings have been confirmed by Bharati *et al.* [15], [16] while Rathgeb *et al.* [19] showed that face recognition systems might be robust to the application of moderate facial retouching.

Different facial retouching detection procedures were proposed by Bharati *et al.* [15], [16]. To distinguish between unaltered and retouched facial images, different deep learning-based techniques were proposed. A sufficient number of retouched facial images was automatically generated for training purposes. The system proposed in [15] demonstrably outperformed a re-implementation of an image forensic approach [33] with respect to detection accuracy. Interestingly, it has also been reported that the approach proposed in [15] achieves competitive performance for the task of make-up detection on a database where no retouching was applied. This indicates that this scheme recognizes exaggerated facial looks that can also be caused by facial cosmetics. Further, Bharati *et al.* [16] analyzed retouching detection across demographic groups. Two different software packages were used to retouch facial images of two sexes, male and female, and three ethnic groups, Indian, Chinese and Caucasian. The authors present the limitations of different state-of-the-art methods, *i.e.* algorithms based on universal texture descriptors and the scheme of [15], in demographic cross-evaluations. Further, in [16] it was shown that the performance of these algorithms is negatively affected when trained on different demographic groups. A deep learning approach to detecting any kind of facial retouching (including GAN-based changes) was proposed by Jain *et al.* [17]. In terms of retouching detection, impressive performance rates (>99% CCR) were reported when training and test were performed on disjunctive subsets of the database introduced in [15].

More recently, Wang *et al.* [18] introduced a deep learning-based facial retouching detection scheme which is specifically designed to detect image warping operations performed using the Adobe Photoshop software. Rathgeb *et al.* [19] proposed a facial retouching detection scheme which makes use of well-established image forensics techniques. Specifically, different spatial and spectral features extracted from Photo Response Non-Uniformity (PRNU) patterns across image regions are analyzed.

In summary, the following key findings can be made:

- Promising detection performance rates have been reported in many works on facial retouching, in particular for deep learning-based retouching detection schemes. However, the majority of works considers an experimental setup in which training and test images are taken from a single database and are retouched using a single retouching algorithm. Unlike traditional image forensics-based manipulation detection schemes [33], [34], further studies are needed to investigate whether these methods are affected by overfitting. Facial retouching detection should be evaluated in a scenario in which the applied retouching algorithm is not known to the detection scheme, *i.e.* not seen in the training stage. Such a scenario is expected to better reflect real-world cases.
- In a single image scenario, in which only a potentially manipulated image is processed by the detection system, the detection performance can highly depend on the quality of the manipulated image as well as applied image post-processing. It was found that image compression can severely impact face recognition performance [35] as well as retouching detection methods [17], [19]. Similar effects are to be expected for other

(a) FERET references (b) FRGCv2 references

(c) FERET probes (d) FRGCv2 probes

FIGURE 2. Examples of reference and probe images of both used databases.

types of post-processing, *e.g.* color-space transformations or even print-scan transformations.

- Unfortunately, the majority of the revisited facial retouching detection systems is not publicly available, in particular pre-trained detection models. Since some of the aforementioned related works require an extensive training, large datasets of retouched images would be required in order to train re-implementations. In addition, important optimizations might have been omitted in proposed retouching detection schemes. Due to these facts, a direct comparison of the presented detection scheme with published approaches in terms of detection performance is often hampered.

III. DATABASES

Used face image datasets are composed of subsets of two publicly available face image databases, *i.e.* FERET [36] and FRGCv2 [37]. The following subsections describe the choice of reference and probe images (Sect. III-A) and the generation of retouched face images (Sect. III-B).

A. REFERENCE AND PROBE IMAGES

For reference images frontal faces with neutral expression have been manually chosen and ICAO compliance has been verified. In particular, the inter-eye-distance of a face has to be at least 90 pixels [38]. Further, probe images were selected which exhibit variations in pose, expression, focus and illumination. If possible, probe images were preferably chosen from different acquisition session in order to obtain a realistic scenario. Examples of probe and reference images of both face image subsets are depicted in Fig. 2. The number of subjects, corresponding reference and probe images, as well as the resulting genuine and impostor comparisons are listed in Table 2.

TABLE 2. Overview of chosen face image subsets from the FERET and FRGCv2 face databases: amount of subjects, corresponding reference and probe images as well as resulting number of genuine and impostor comparisons ("f" and "m" denote female and male, respectively).

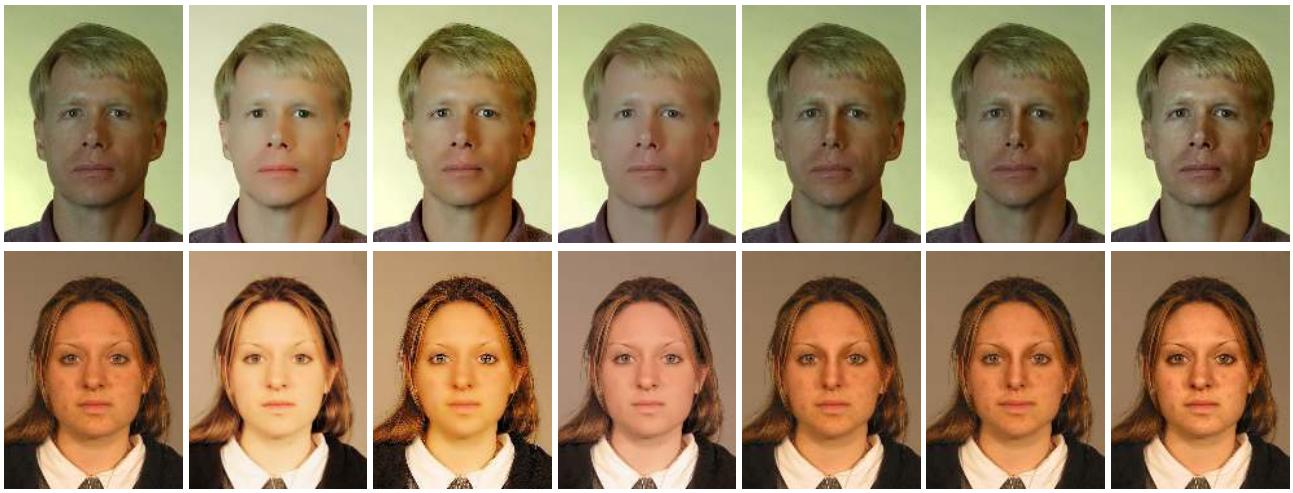
Database	Subjects (f/m)	Images		Comparisons	
		Reference	Probe	Genuine	Impostor
FERET	529 (200/329)	529	791	791	147,712
FRGCv2	533 (231/302)	984	1,726	3,298	144,032

B. AUTOMATIC RETOUCHING

In order to retouch reference face images different freely available apps from the Google PlayStore [39] were selected. It is important to emphasis that free apps are more likely to be applied by users compared to costly desktop applications which have been employed in related works [15], [16]. Moreover, the users' ratings of eligible apps and the number of downloads are considered as selection criteria. It is assured that apps provide results of sufficient quality, *i.e.* apps which produce doll-like looking faces are neglected. Finally, easy-to-use apps which allow for an (all-in-one) automatic beautification are favored since these apps facilitate an automatic creation of retouched images as will be explained hereafter.

Based on mentioned criteria the following six apps were chosen for the database creation:

1) *AirBrush* [40] slightly enlarges the eyes, makes the face slightly slimmer and more shiny, eliminates minor wrinkles and skin impurities, and reduces dark rings under the eyes;
2) *BeautyPlus* [41] enlarges the eyes (and makes them more shiny), makes the face more shiny, eliminates minor wrinkles and skin impurities, and reduces dark rings under the eyes;
3) *Bestie* [42] makes the face slightly slimmer and more shiny, eliminates minor wrinkles and skin impurities, and reduces dark rings under the eyes;



(a) original (b) AirBrush (c) BeautyPlus (d) Bestie (e) FotoRus (f) InstaBeauty (g) YouCamPerfect

FIGURE 3. Selected retouching apps applied to a male (top) and a female (bottom) face image: (a) original face images and (b)-(g) resulting retouched face images.

4) *FotoRus* [12] enlarges the eyes, makes the face slimmer, performs a nose thinning/lifting, and reduces dark rings under the eyes;
5) *InstaBeauty* [13] enlarges the eyes, makes the face slightly slimmer and smilingly, performs a slight nose thinning, and reduces small skin impurities;
6) *YouCam Perfect* [43] enlarges the eyes, makes the cheeks more rosy, eliminates minor wrinkles and skin impurities, and smooths the hair.

Fig. 3 depicts examples of applications of each selected app to a male and a female face image.

The automated generation of retouched images was performed on a Samsung Galaxy S6 device with Android version 7.0 and an Apple MacBook Pro. The Automate app [44], which is an Android automation app, was used to automatically apply FotoRus and InstaBeauty to all reference images of both databases. For the remaining apps the desktop click recording software Cliclick [45] was used together with the Android app ApowerMirror [46], which enables a mirroring of a smartphone device to a desktop device. This automated process resulted in a total number of (529+984)×6=9,078 retouched face images.

IV. DIFFERENTIAL DETECTION OF FACIAL RETOUCHING

The proposed differential retouching detection system processes image pairs of potentially retouched reference face images and trusted unaltered probe face images. Different types of features, *i.e.* texture descriptors (TD), facial landmarks (FL), and deep face representations (DFR), are extracted from both images and difference vectors are estimated. Detection scores from separately trained machine learning-based classifiers, *i.e.* Support Vector Machines (SVMs), are then fused to distinguish between bona fide and retouched reference images. An overview of the workflow of the proposed differential detection system is depicted in Fig. 4. In the following subsections, the feature extraction (Sect. IV-A) as well as the training and classification (Sect. IV-B) of the proposed system are described in detail.

A. FEATURE EXTRACTION

The following three types of features are extracted from a pair of reference and probe face image:

1) *Texture descriptors (TD)*: in the pre-processing both face images are normalized 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 [47] 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 normalized images are cropped to regions of 320 × 320 pixels centered around the tip of the nose. Cropped face parts are then converted to a grayscale image.
At feature extraction the pre-processed face image is divided into 4 × 4 cells to retain local information. Local Binary Patterns (LBP) [48] are extracted from each cell of the pre-processed face images. LBP feature vectors are extracted in their simplest form employing a radius of one where eight neighboring pixel values are processed within 3 × 3 pixel patches. For details on the extraction of LBP feature vectors the reader is referred to [48]. Obtained feature values are aggregated in a corresponding histograms. The final feature vector is formed as a concatenation of histograms extracted from each cell.
LBP has been found to be a powerful feature for texture classification. In the context of facial retouching detection, it is expected that LBP-based feature vectors extracted from the reference and probe image clearly differ, if the reference image has been

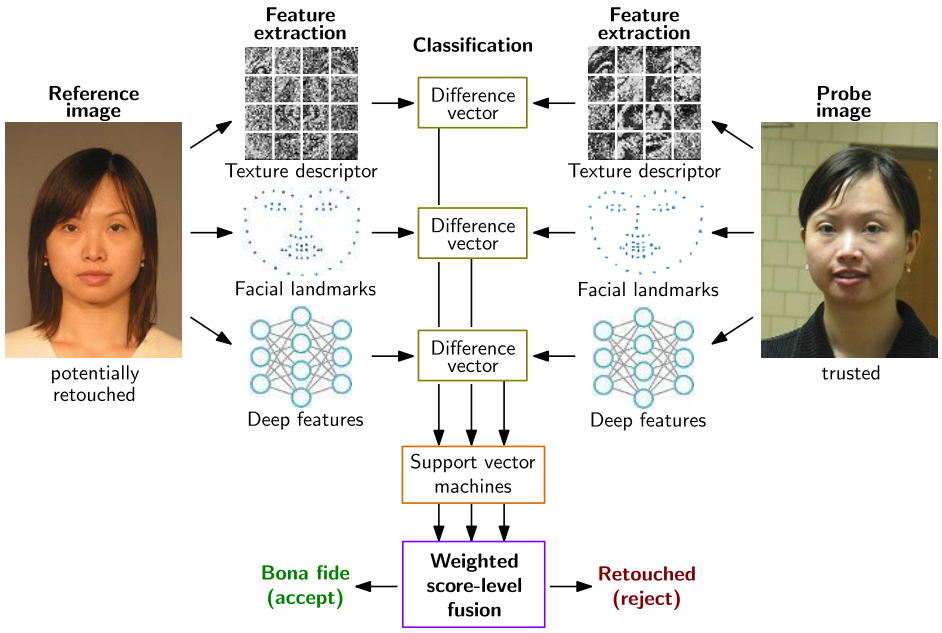


FIGURE 4. Overview of the proposed differential retouching detection system.

heavily retouched. Specifically, if skin smoothing operations are applied to eliminate wrinkles and impurities, LBP-based feature vectors are expected to significantly vary.

2) *Facial landmarks (FL)*: the previously mentioned *dlib* landmark detector [47] is used to extract a total number of 68 two-dimensional facial landmarks from each reference and probe face image. Extracted landmarks describe the jawline, eyebrows, nose, eyes and lips of a face. Again, facial landmark positions are normalized according to eye coordinates.
   Focusing on the task of retouching detection, facial landmarks are expected to change if anatomical alterations are induced by a retouching algorithm. In particular, thinning/lifting of the nose, enlarging the eyes, or a slimming of the the entire face will greatly modify facial landmark positions.
3) *Deep face representation (DFR)*: lastly, deep face representations are extracted from the reference and probe image using the *ArcFace* algorithm [49]. This algorithm is based on the ResNet-50 convolutional neural network architecture and uses Additive Angular Margin Loss to obtain highly discriminative features for face recognition. ArcFace was shown to achieve state-of-the-art recognition performance on various challenging datasets. The publicly available pre-trained deep face recognition network is used as feature extractor, *i.e.* the deep representations extracted by the neural network (on the lowest layer). Since this algorithm uses some internal pre-processing no cropping (or grayscale conversion) is applied prior to the feature extraction. Feature vectors comprising 512 floats are extracted from the reference and probe face images.
   Deep face recognition systems leverage very large databases of face images to learn rich and compact representations of faces. It is expected that alterations induced by facial retouching will also be reflected in extracted deep face features. Due to the high generalization capabilities of deep face recognition systems with respect to textural changes of skin, such changes might be more pronounced in case anatomical alterations are induced through retouching.

Many alternative algorithms with similar properties have been proposed for each considered type of feature extraction over the past years, which could also be applied, see [4], [50], [51] for recent surveys. However, a rigorous analysis of the worthiness of different feature extraction techniques for the task of differential retouching detection is out of scope in this first study on differential retouching detection.

B. TRAINING AND CLASSIFICATION

At training and classification, difference vectors are estimated from pairs of feature vectors extracted from a reference and probe face image. Specifically, an element-wise subtraction of feature vectors is performed. For the facial landmark-based feature vectors x- and y-coordinates are subtracted separately, resulting in a difference vector of length 2×68. Note that, resulting difference vectors also retain the direction of differences as opposed to a distance vector, which would only comprise absolute differences between the feature vectors.

In the training stage, difference vectors are extracted for each feature extractor and SVMs with Radial Basis Function (RBF) kernels are trained to distinguish between bona fide and retouched face images. The *scikit-learn* library [52] is used to train SVMs. Data-normalization is applied as the feature elements of extracted feature vectors are expected

have different ranges. This is particularly the case in cross-database experiments and hence represents an essential processing step. The normalization process aims to rescale the feature elements to have a mean of 0 and a standard deviation of 1. To this end, the StandardScaler of the scikit-learn library is employed. During training, a regularization parameter of $C = 1$ and a kernel coefficient Gamma of $1/n$ is used, where n denotes the number of feature elements.

While a concatenation of difference vectors would allow a feature level fusion by training a single SVM, separate SVMs are trained due to the difference in the nature of the extracted feature vectors and their dimensions. Trained SVMs generate a normalized attack detection score in the range [0, 1]. Subsequently, a weighted score-level fusion is performed by testing different combinations of weights. The sum-rule is used to obtain a fused score based on which the final decision is made. Alternatively, more sophisticated fusion techniques could be employed, which is beyond the scope of this work. Due to the varying nature of retouching algorithms, machine learning-based classifiers could also be employed for the purpose of score-level fusion. This would be particularly beneficial, if a large number of feature extractors are applied, which could also be subject to future investigations.

V. EXPERIMENTS

Firstly, used evaluation metrics are summarized (Sect. V-A). Experimental results on the impact of facial retouching on face recognition performance are presented (Sect. V-B). Subsequently, the detection performance of the proposed system is evaluated in scenarios where the applied retouching algorithms are known (Sect. V-C) and unknown (Sect. V-D) during training. Finally, the detection performance of the differential retouching detection system is compared to several published single image-based detection methods (Sect. V-E).

A. EVALUATION METRICS

Biometric performance is evaluated in terms of False Non-Match Rate (FNMR) and False Match Rate (FMR). More precisely, the FNMR at a FMR of 0.1%, referred to as $\text{FNMR}_{0.1}$, is reported which represents the operation point recommended in the guidelines of European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX) [53]. In addition, as a measure of decidability $d' = |\mu_g - \mu_i| / \sqrt{\frac{1}{2}(\sigma_g^2 + \sigma_i^2)}$ is reported, where μ_g and μ_i represent the means of the genuine (mated comparison trials) and the impostor (non-mated comparison trials) score distributions and σ_g and σ_i their standard deviations, respectively. The amount of genuine and impostor comparisons for bona fide and retouched images (per retouching algorithm) on each of the used databases are summarized in Table 2.

The performance of the detection algorithms is reported according to metrics defined in ISO/IEC 30107-3 [54]. The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same presentation attack instrument species incorrectly classified as bona fide presentations in a specific scenario. The Bona Fide Presentation Classification Error Rate (BPCER) is defined as the proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario. The D-EER, *i.e.* the operation point where detection accuracy APCER = BPCER, is reported for different detection methods. In addition, the BPCER10, *i.e.* the operation point where APCER = 10%, and BPCER20, *i.e.* the operation point where APCER = 5%, are estimated.

In experiments on facial retouching detection, training and testing is conducted on the disjoint datasets. On the one hand, all bona fide and retouched face images of FRGCv2 are used for training and evaluations are performed on the FERET database for individual retouching algorithms. On the other hand, the FERET database is used for training and the FRGCv2 for testing in the same manner. The number of bona fide and retouched comparisons (per retouching algorithm) on the FERET and FRGCv2 databases is equal to the number of genuine comparisons for each database listed in Table 2.

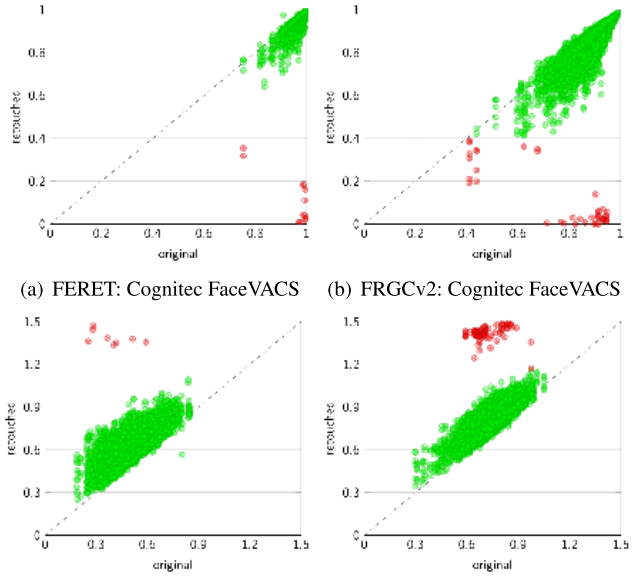
B. IMPACT ON FACE RECOGNITION

Two different face recognition systems are used in the evaluation, *i.e.* the Cognitec FaceVACS v9.3 [55] and ArcFace [49]. While the first system is a frequently deployed commercial product the latter represents an open-source algorithm which is widely used in the biometrics research community. Given a pair of face images the Cognitec FaceVACS returns a similarity score in the range [0, 1] (*i.e.* high values indicate high similarity) while the ArcFace system returns a distance score in the range [0, 1.5] (*i.e.* low values indicate high similarity).

To obtain fixed thresholds for the $\text{FNMR}_{0.1}$ values for both face recognition systems impostor comparisons are obtained using bona-fide images. With respect to the $\text{FNMR}_{0.1}$ a fixed decision threshold of 0.4 and 1.15 was estimated for the normalized comparison scores of the commercial system and the open-source system, respectively. While the ArcFace successfully processed all reference images the unconstrained probe images caused a Failure to Extract Rate (FTX) of 1.76% and 0.3% on the FERET and FRGCv2 probe images, respectively. For the Cognitec FaceVACS system a zero FTX was achieved for all images. Note that the FTX is ignored when estimating the FNMR [31]. Obtained performance rates are summarized in Table 3. Generally, the Cognitec FaceVACS achieves higher d' values compared to the ArcFace system, which indicates clearer separation of genuine and impostor score distributions. In terms of $\text{FNMR}_{0.1}$ both systems achieve similar performance. Moreover, it can be concluded that the AirBrush app has the most severe impact on recognition accuracy, followed by FotoRus, InstaBeauty, and YouCam Perfect. The least impact is observed for Bestie and BeautyPlus. Scatter plots of genuine comparison scores before and after retouching across all apps are shown in Fig. 5. From the scatter plots it can be seen that retouching causes a general deterioration of comparison scores. Further, it can be observed that non-matches (red dots) mostly result from highly deteriorated comparison scores.

TABLE 3. Performance results for both face recognition systems on both databases ($FNMR_{0.1}$ in %).

System	Retouching	FERET		FRGCv2	
		d'	$FNMR_{0.1}$	d'	$FNMR_{0.1}$
Cognitec FaceVACS v9.3	none	22.79	0.0	14.87	0.03
	AirBrush	12.62	0.77	10.06	0.82
	BeautyPlus	21.39	0.0	11.09	0.30
	Bestie	22.30	0.0	12.01	0.30
	FotoRus	18.57	0.13	10.44	0.42
	InstaBeauty	17.91	0.13	9.44	0.76
	YouCam Perfect	16.01	0.51	11.33	0.58
ArcFace	none	9.23	0.0	7.99	0.0
	AirBrush	7.89	0.77	6.66	0.91
	BeautyPlus	9.09	0.0	7.13	0.24
	Bestie	9.29	0.0	7.42	0.24
	FotoRus	8.51	0.13	7.04	0.30
	InstaBeauty	8.20	0.13	6.56	0.69
	YouCam Perfect	8.87	0.26	7.15	0.55



(a) FERET: Cognitec FaceVACS (b) FRGCv2: Cognitec FaceVACS

(c) FERET: ArcFace (d) FRGCv2: ArcFace

FIGURE 5. Scatter plots of genuine comparison scores of both face recognition systems on both databases before and after applying facial retouching (red dots mark scores resulting in a rejection after beautification using a decision threshold yielding a FMR of 0.1%).

Compared to results reported for other face recognition systems a few years ago, *e.g.* [14], [15], considered state-of-the-art face recognition systems appear relatively robust to facial retouching, maintaining $FNMR_{0.1}$s $<1\%$ across all retouching apps and both databases. In even more challenging scenarios, *e.g.* in case reference images are acquired under unconstrained conditions as well, a more severe impact of facial retouching is to be expected.

C. DETECTION PERFORMANCE FOR KNOWN RETOUCHING

In the first facial retouching detection experiments, the training and testing are performed utilizing a single retouching algorithm. This experiment corresponds to the scenario where the potentially used retouching method is known beforehand. The performance is estimated for detection schemes trained with a single feature extraction. Subsequently, results for the proposed (weighted) score-level fusion of all classifiers are presented.

Table 4 lists obtained detection performance rates for different configurations of the proposed system. In this setting, similar detection error rates are obtained for training on FERET and testing on FRGCv2 and vice versa. It can be observed that the detection performance of detection schemes based on individual feature extraction methods highly varies across retouching apps. For instance, the TD-based detector achieves the best results for detecting images which have been retouched applying Bestie or AirBrush, which perform severe textural alterations on the entire face region, *i.e.* skin smoothing. The FL-based detection scheme obtains only moderate detection performance with lowest error rates for FotoRus and InstaBeauty. These retouching apps induce anatomical changes, *e.g.* thinning of the nose, which cause larger differences in facial landmarks. Best overall detection performance is achieved for DFR, associated to average D-EERs of 3.71% and 6.59% on the FRGCv2 and the FERET database, respectively. Due to the applied deep-learning, resulting face representations are expected to comprise both, textural as well as anatomical information. Moreover, it can be seen that retouching detection becomes more challenging in case only small alterations are performed by a retouching app. For example, the YouCam Perfect app only slightly edits face images which leads to higher detection errors for all individual detection systems. However, "minor" image edits are being excluded from discussed photoshop legislations [23].

It can be observed that a fusion of detection scores using the sum-rule significantly improves accuracy. Error rates drop across nearly all retouching apps and both databases, resulting in average D-EERs of 2.49% on the FRGCv2 and 3.04% on the FERET database. This general decrease of error rates shows that the individual detection systems based on a single feature extraction methods complement each other. Therefore, the proposed fusion outperforms the best single feature extraction-based detection systems in the vast majority of cases. For the weighted fusion with weights of 0.4, 0.1 and 0.5 for TD, FL, and DFR, respectively, have been found to reveal best detection performance yielding competitive average D-EERs of 1.43% and 2.43% on FRGCv2 and FERET, respectively. Note that the weight assigned to FL is rather low due to its high error rates. However, this weight might be increased in cases where an image is suspected to have undergone alteration of facial shape.

D. DETECTION PERFORMANCE FOR UNKNOWN RETOUCHING

In the second experiment on facial retouching detection, training is performed on all but one retouching app. This retouching app is then applied in the testing stage. This scenario corresponds to the case where the potentially used retouching method is not known beforehand. Note that this

TABLE 4. Performance results for differential detection of known retouching.

System	Retouching	**Training:** FERET -- **Test:** FRGCv2			**Training:** FRGCv2 -- **Test:** FERET			*Average*		
		D-EER	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**
TD	AirBrush	0.64	0.0	0.0	2.55	0.69	1.37	1.59	0.35	0.68
	BeautyPlus	8.15	4.08	24.20	2.64	0.27	1.03	5.39	2.17	12.62
	Bestie	0.64	0.0	0.0	5.78	2.91	6.52	3.21	1.46	3.26
	FotoRus	4.08	0.89	2.29	5.88	3.88	7.07	4.98	2.39	4.68
	InstaBeauty	3.19	0.25	1.91	5.60	3.04	6.04	4.39	1.64	3.97
	YouCamPerfect	27.01	49.17	64.71	22.93	39.36	50.44	24.97	44.26	57.57
	Average	**7.28**	**9.07**	**15.52**	**7.56**	**8.36**	**12.08**	**7.42**	**8.71**	**13.80**
FL	AirBrush	27.31	52.84	71.30	34.07	71.27	84.42	30.69	62.06	77.86
	BeautyPlus	25.41	50.19	69.03	32.96	72.01	84.29	29.19	61.10	76.66
	Bestie	28.70	58.15	73.70	36.51	70.95	84.66	32.61	64.55	79.18
	FotoRus	8.09	6.45	13.15	22.59	43.51	58.94	15.34	24.98	36.05
	InstaBeauty	6.57	4.43	9.61	15.83	24.32	37.42	11.20	14.37	23.51
	YouCamPerfect	32.74	67.51	77.12	37.30	74.74	87.27	35.02	71.13	82.19
	Average	**21.47**	**39.93**	**52.32**	**29.88**	**59.47**	**72.83**	**25.67**	**49.7**	**62.58**
DFR	AirBrush	0.89	0.13	0.13	3.54	0.73	2.31	2.21	0.43	1.22
	BeautyPlus	6.18	2.17	8.02	6.49	3.73	7.86	6.34	2.95	7.94
	Bestie	6.12	3.82	7.13	13.83	18.43	31.79	9.97	11.13	19.46
	FotoRus	0.13	0.0	0.0	0.56	0.0	0.0	0.34	0.0	0.0
	InstaBeauty	0.0	0.0	0.0	0.42	0.0	0.0	0.21	0.0	0.0
	YouCamPerfect	8.92	7.52	15.80	14.71	20.35	32.40	11.81	13.93	24.10
	Average	**3.71**	**2.27**	**5.18**	**6.59**	**7.21**	**12.39**	**5.15**	**4.74**	**8.78**
Fusion	AirBrush	0.71	0.0	0.0	0.38	0.0	0.0	0.55	0.0	0.0
	BeautyPlus	9.32	8.99	12.27	4.72	0.38	3.44	7.02	4.69	7.86
	Bestie	2.54	0.0	0.67	1.66	0.13	0.25	2.10	0.06	0.46
	FotoRus	0.07	0.0	0.0	0.63	0.0	0.0	0.36	0.0	0.0
	InstaBeauty	0.03	0.0	0.0	0.76	0.0	0.0	0.40	0.0	0.0
	YouCamPerfect	2.32	0.94	1.39	10.06	10.06	33.89	6.19	5.50	17.64
	Average	**2.49**	**1.65**	**2.38**	**3.04**	**1.76**	**6.26**	**2.77**	**1.71**	**4.33**
Weighted Fusion	AirBrush	0.41	0.0	0.0	0.25	0.0	0.0	0.33	0.0	0.0
	BeautyPlus	4.11	1.55	3.34	3.25	0.89	1.15	3.68	1.22	2.24
	Bestie	3.05	0.0	0.36	3.06	0.13	2.29	3.05	0.06	1.33
	FotoRus	0.03	0.0	0.0	0.25	0.0	0.0	0.14	0.0	0.0
	InstaBeauty	0.03	0.0	0.0	0.25	0.0	0.0	0.14	0.0	0.0
	YouCamPerfect	2.73	0.67	1.15	7.52	4.33	33.38	5.12	2.50	17.27
	Average	**1.43**	**0.64**	**0.82**	**2.43**	**0.89**	**6.14**	**1.93**	**0.76**	**3.48**

scenario better reflects a real-world case in which it must not be assumed that the potentially applied retouching algorithm is known. In this setting, the same amount of retouched images are used during training. These are alternately chosen from the five remaining sets of retouched face images. For instance, if testing is performed for AirBrush, the first retouched training image is chosen from the BeautyPlus set, the second from the Bestie set and so on and so forth.

Obtained detection performance rates for different configurations of the proposed system are summarized in Table 5. Corresponding DET curves are depicted in Fig. 6. It can be observed that overall the detection performance drastically drops in this more challenging setting. Again, no trend with respect to differences in terms of detection performance is noticeable across the used databases. More specifically, TD and FL obtain lower error rates on the FERET database while this is not the case for DFR. This indicates that TD and FL are impacted by high variations with respect to pose, illumination, and focus as it is the case for the FRGCv2 probe set. This causes generally larger differences between reference and probe images of a single subject and hence hampers a reliable detection of retouching in the differential scenario.

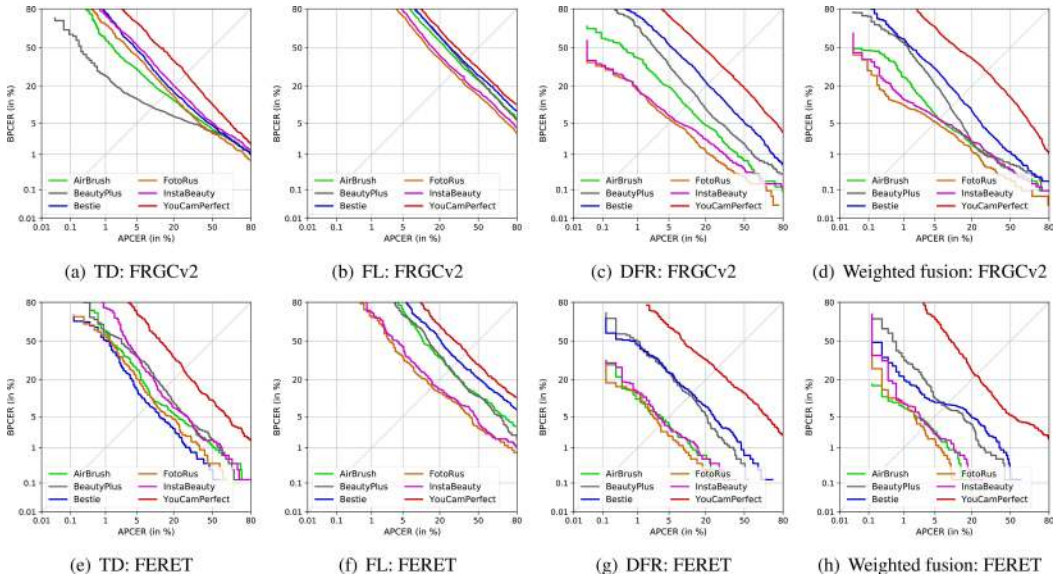
The proposed fusion-based detection system improves the detection performance compared to the ones based on each single feature extractor. Again, applying weights in the fusion yields further improvements resulting in average D-EERs of 11.71% and 8.16% on the FRGCv2 and the FERET database, respectively.

E. COMPARISON WITH SINGLE IMAGE DETECTION METHODS

In the last experiment, the proposed system is compared to other published single image-based approaches in the challenging scenario where the potentially used retouching app is unknown. The following single image facial retouching detection methods have evaluated: a generic image forgery detection tool which aims at detecting inconsistencies in noise variances [56], the approach by Wang *et al.* [18], Rathgeb *et al.* [19], DFR, and TD (not that the FL-based method is not expected to reveal competitive results in a single image detection scenario). While the first scheme does not require any training, a pre-trained model is available for the scheme of [18]. The remaining methods are trained in the same previously described manner like the proposed

TABLE 5. Performance results for differential detection of unknown retouching.

System	Retouching	**Training:** FERET -- **Test:** FRGCv2			**Training:** FRGCv2 -- **Test:** FERET			*Average*		
		D-EER	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**
TD	AirBrush	15.50	20.91	32.14	10.06	9.94	26.75	12.78	15.42	29.45
	BeautyPlus	9.71	9.44	13.26	14.54	23.14	36.54	12.12	16.29	24.90
	Bestie	18.41	31.32	47.92	8.41	7.13	14.52	13.41	19.23	31.22
	FotoRus	17.09	27.62	42.76	9.94	8.917	20.64	13.51	18.27	31.70
	InstaBeauty	20.25	36.60	52.23	13.12	20.38	37.45	16.69	28.49	44.84
	YouCamPerfect	28.35	55.12	70.65	26.55	56.13	71.93	27.45	55.63	71.29
	Average	**18.22**	**30.17**	**43.16**	**13.77**	**20.94**	**34.64**	**16.01**	**25.56**	**38.90**
FL	AirBrush	35.42	71.71	84.05	26.55	55.63	74.34	30.99	63.67	79.20
	BeautyPlus	36.78	76.99	87.20	27.43	61.44	71.05	32.11	69.22	79.13
	Bestie	37.75	76.74	87.20	33.12	68.39	81.54	35.44	72.57	84.37
	FotoRus	29.50	59.49	75.41	15.55	22.12	35.27	22.53	40.81	55.34
	InstaBeauty	31.02	63.43	78.84	16.81	26.80	45.13	23.92	45.12	61.99
	YouCamPerfect	40.69	80.65	89.14	39.95	80.66	88.62	40.32	80.66	88.88
	Average	**35.19**	**71.5**	**52.32**	**26.57**	**52.51**	**65.99**	**30.88**	**62.01**	**59.16**
DFR	AirBrush	10.23	10.42	19.56	4.20	1.65	3.44	7.22	6.04	11.50
	BeautyPlus	14.65	20.92	35.01	11.66	13.38	24.71	13.16	17.15	29.86
	Bestie	21.14	39.11	52.20	12.74	15.03	24.33	16.94	27.07	38.27
	FotoRus	5.62	3.64	6.17	3.31	0.76	1.91	4.46	2.20	4.04
	InstaBeauty	6.07	4.46	6.89	3.82	1.66	3.06	4.95	3.06	4.98
	YouCamPerfect	32.77	63.86	75.58	26.88	46.88	61.78	29.83	55.37	68.68
	Average	**15.08**	**23.74**	**32.57**	**10.44**	**13.23**	**19.87**	**12.76**	**18.49**	**26.22**
Fusion	AirBrush	7.74	5.37	12.27	3.57	0.89	1.78	5.65	3.13	7.03
	BeautyPlus	10.94	12.48	21.74	8.60	6.75	17.83	9.77	9.62	19.79
	Bestie	13.70	18.04	31.73	11.97	15.03	25.86	12.84	16.54	28.80
	FotoRus	6.63	4.13	8.68	2.04	0.64	0.89	4.34	2.38	4.79
	InstaBeauty	8.16	6.47	12.57	2.67	0.76	1.40	5.42	3.62	6.99
	YouCamPerfect	28.08	51.14	65.11	24.84	47.90	64.46	26.46	49.52	64.79
	Average	**12.54**	**16.27**	**25.35**	**8.95**	**12.0**	**18.7**	**10.75**	**14.14**	**22.03**
Weighted Fusion	AirBrush	5.95	3.95	7.19	3.57	0.89	1.78	4.76	2.42	4.49
	BeautyPlus	9.50	8.84	21.01	7.84	6.62	10.70	8.67	7.73	15.86
	Bestie	13.52	19.07	32.40	8.41	8.15	9.04	10.96	13.61	20.72
	FotoRus	5.21	3.13	5.31	2.93	0.128	1.53	4.07	1.63	3.42
	InstaBeauty	5.97	4.58	6.71	2.67	0.76	1.40	4.32	2.67	4.06
	YouCamPerfect	30.11	53.69	67.20	22.68	49.17	68.66	26.40	51.43	67.93
	Average	**11.71**	**15.54**	**23.31**	**8.16**	**11.04**	**15.75**	**9.94**	**13.29**	**19.53**



(a) TD: FRGCv2 (b) FL: FRGCv2 (c) DFR: FRGCv2 (d) Weighted fusion: FRGCv2

(e) TD: FERET (f) FL: FERET (g) DFR: FERET (h) Weighted fusion: FERET

FIGURE 6. DET curves for facial retouching detection systems based on individual feature extraction methods.

system. In order to use DFR and TD in a single image-based detection, the SVM-based classifiers are directly trained with feature vectors obtained from the retouched and bona fide reference images. Obtained detection performance rates are shown in Table 6. Lowest error rates are obtained by the approach scheme of [19] achieving rather high average

TABLE 6. Performance results for other single image-based retouching detection methods.

System	Retouching	**Training:** FERET **--Test:** FRGCv2			**Training:** FRGCv2 **--Test:** FERET			*Average*		
		D-EER	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**	**D-EER**	**BPCER10**	**BPCER20**
Noise Variance [56] (no training)	AirBrush	45.64	81.53	89.10	40.51	70.10	76.37	43.08	75.82	82.74
	BeautyPlus	45.61	87.78	94.86	40.35	67.85	74.76	42.98	77.82	84.81
	Bestie	40.72	87.78	94.79	43.73	72.83	83.60	42.23	80.31	89.20
	FotoRus	49.77	89.31	94.86	49.20	89.07	93.89	49.49	89.19	94.38
	InstaBeauty	49.88	89.79	94.86	49.04	89.23	93.89	49.46	89.51	94.38
	YouCamPerfect	47.24	88.96	94.38	49.20	89.23	93.41	48.22	89.10	93.90
	Average	**46.48**	**87.52**	**93.81**	**45.34**	**79.72**	**85.99**	**45.91**	**83.62**	**89.90**
Wang et al. [18] (no training)	AirBrush	43.04	83.61	91.53	37.86	59.0	72.67	40.45	71.31	82.10
	BeautyPlus	2.92	0.55	1.46	4.98	2.89	4.82	3.95	1.72	3.14
	Bestie	28.60	51.88	61.26	13.83	15.76	19.45	21.22	33.82	40.35
	FotoRus	6.84	4.31	10.56	15.43	19.45	31.19	11.13	11.88	20.88
	InstaBeauty	3.96	1.25	2.98	9.01	7.878	13.34	6.48	4.56	8.16
	YouCamPerfect	34.15	64.86	75.14	17.36	24.60	30.23	25.76	44.73	52.69
	Average	**19.92**	**34.41**	**40.49**	**16.41**	**21.6**	**28.62**	**18.17**	**28.01**	**34.56**
Rathgeb et al. [19]	AirBrush	22.42	38.23	42.17	24.02	72.23	92.03	23.22	55.23	67.10
	BeautyPlus	13.93	25.66	38.53	14.05	16.72	24.53	13.99	21.19	31.53
	Bestie	18.04	28.04	50.33	20.33	42.78	60.88	19.19	35.41	55.61
	FotoRus	15.34	30.68	49.19	17.07	38.87	65.82	16.21	34.78	57.51
	InstaBeauty	2.45	0.43	1.77	1.27	0.0	0.90	1.86	0.22	1.34
	YouCamPerfect	24.97	75.16	92.21	25.71	78.53	94.44	25.34	76.85	93.33
	Average	**16.19**	**33.03**	**45.70**	**17.08**	**41.52**	**56.43**	**16.63**	**37.28**	**51.07**
DFR (single image)	AirBrush	15.03	21.67	35.66	14.79	24.12	35.85	14.91	22.90	35.76
	BeautyPlus	18.37	29.17	50.08	24.94	49.52	68.65	21.66	39.35	59.37
	Bestie	26.89	49.95	66.13	26.85	56.27	68.17	26.87	53.11	67.15
	FotoRus	9.99	10.02	16.51	14.95	20.90	29.74	12.47	15.46	23.13
	InstaBeauty	9.35	8.98	17.72	17.85	26.37	36.33	13.60	17.68	27.03
	YouCamPerfect	35.07	69.53	82.03	36.66	72.35	82.96	35.87	70.94	82.50
	Average	**19.12**	**31.55**	**44.69**	**22.67**	**41.59**	**53.62**	**20.90**	**36.57**	**49.16**
TD (single image)	AirBrush	6.52	4.76	7.74	15.27	18.81	26.37	10.90	11.79	17.05
	BeautyPlus	6.95	4.88	8.53	24.94	49.52	68.65	15.94	27.20	38.59
	Bestie	9.65	9.07	16.36	29.10	44.37	60.45	19.37	26.72	38.41
	FotoRus	8.77	7.47	12.32	17.20	22.67	27.65	12.98	15.07	19.99
	InstaBeauty	12.75	15.05	23.31	18.49	26.37	34.73	15.62	20.71	29.02
	YouCamPerfect	44.93	85.25	93.47	49.20	90.51	95.82	47.07	87.88	94.65
	Average	**14.93**	**21.08**	**23.79**	**22.67**	**36.55**	**45.28**	**18.80**	**28.82**	**34.54**

D-EERs of 16.19% and 17.08% on the FRGCv2 and the FERET database, respectively. Similarly, the remaining schemes obtain average D-EERs slightly below 20%, except for the method of [56] which obtains results close to guessing.

In this challenging cross-database evaluation in which the retouching algorithm is unknown, it is observable that the proposed differential detection method clearly outperforms all considered single image-based detection systems. This confirms the worthiness of the proposed differential detection concept. Overall, it can be observed that in this scenario obtained error rates of evaluated single image detection schemes are considerably higher than what has been reported in corresponding publications. Most of these works have not investigated in a cross-database experiment in which a variety of retouching apps is used for training and the retouching app used during training is unknown. Hence, it is reasonable to assume that similar effects would be observable for other single image-based facial retouching detection schemes.

VI. CONCLUSIONS

Ever-increasing progress in image manipulation has raised many concerns, which might lead to a loss of trust in digital content. This is especially critical for facial images used in biometric recognition system. In this context, facial retouching algorithms play an important role since numerous easy-to-use mobile apps have become freely available. Said apps are frequently used in different scenarios where face recognition technologies are employed.

In this work, the concept of differential facial retouching detection was firstly introduced. The up until now largest dataset of retouched face image was created by applying six popular mobile retouching apps in an automated manner to subsets of two public face image databases. State-of-the-art face recognition systems were benchmarked on the generated dataset. Subsequently, a differential retouching detection system was proposed, which analyses differences in various types of extracted features in order to detect observed alterations induced by applied retouching algorithms. A weighted sum-rule-based score-level fusion of detection scores obtained from separately trained SVMs revealed good detection performance in challenging cross-database evaluations. In the challenging scenario in which the retouching algorithm is not known at the training stage, the proposed differential facial retouching detection system

was shown to significantly outperform different types of single image-based methods. Similar scenarios have previously been considered for the task of morphing attack detection, *e.g.* in the works of Ferrara *et al.* [57] and Scherhag *et al.* [58]. As the quality of image manipulation is steadily improving, differential detection scenarios are expected to become even more relevant for image manipulation detection in future research.

ACKNOWLEDGMENTS

The authors thank Cognitec Systems GmbH for providing a research license of Cognitec FaceVACS face recognition system SDK v9.3.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.
[2] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognit. Lett.*, vol. 28, no. 14, pp. 1885–1906, Oct. 2007.
[3] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*. London, U.K.: Springer, 2011.
[4] M. Wang and W. Deng, "Deep face recognition: A survey," 2018, *arXiv:1804.06655*. [Online]. Available: https://arxiv.org/abs/1804.06655
[5] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[6] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–12.
[7] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "Deep learning for understanding faces: Machines may be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.
[8] M. Kawulok, M. E. Celebi, and B. Smolka, *Advances in Face Detection and Facial Image Analysis*, 1st ed. Cham, Switzerland: Springer, 2016.
[9] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, p. 92, 2018.
[10] National Institute of Standards and Technology (NIST). (2018). *Face Recognition Vendor Test (FRVT) Quality Assessment*. Accessed: Nov. 2019. [Online]. Available: https://www.nist.gov/programs-projects/frvt-quality-assessment
[11] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152667–152678, 2019.
[12] Fotoable, Inc. *FotoRus—Camera & Photo Editor*. Accessed: Jun. 2019. [Online]. Available: https://play.google.com/store/apps/details?id=com.wantu.activity
[13] Fotoable Inc. *InstaBeauty -Makeup Selfie Cam*. Accessed: Jun. 2019. [Online]. Available: https://play.google.com/store/apps/details?id=com.fotoable.fotobeauty
[14] M. Ferrara, A. Franco, D. Maltoni, and Y. Sun, "On the impact of alterations on face photo recognition accuracy," in *Image Analysis and Processing—ICIAP*. Berlin, Germany: Springer, 2013, pp. 743–751.
[15] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting facial retouching using supervised deep learning," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 1903–1913, Sep. 2016.
[16] A. Bharati, M. Vatsa, R. Singh, K. W. Bowyer, and X. Tong, "Demography-based facial retouching detection using subclass supervised sparse autoencoder," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 474–482.
[17] A. Jain, R. Singh, and M. Vatsa, "On detecting GANs and retouching based synthetic alterations," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–7.
[18] S.-Y. Wang, O. Wang, A. Owens, R. Zhang, and A. A. Efros, "Detecting photoshopped faces by scripting photoshop," 2019, *arXiv:1906.05856*. [Online]. Available: http://arxiv.org/abs/1906.05856
[19] C. Rathgeb, A. Botaljov, F. Stockhardt, S. Isadskiy, L. Debiasi, A. Uhl, and C. Busch, "PRNU-based detection of facial retouching," *IET Biometrics*, vol. 9, no. 4, pp. 154–164, Jul. 2020.
[20] B. Ward, M. Ward, O. Fried, and B. Paskhover, "Nasal distortion in short-distance photographs: The selfie effect," *J. Amer. Med. Assoc. Facial Plastic Surg.*, vol. 20, no. 4, pp. 333–335, 2018.
[21] S. Diefenbach and L. Christoforakos, "The selfie paradox: Nobody seems to like them yet everyone has reasons to take them. An exploration of psychological functions of selfies in self-presentation," *Frontiers Psychol.*, vol. 8, p. 7, Jan. 2017.
[22] J. Szewczyk, "Photoshop law: Legislating beauty in the media and fashion industry," *SSRN Electron. J.*, 2014.
[23] N. Eggert. *BBC News: Is She Photoshopped? In France, They Now Have to Tell You*. Accessed: Sep. 30, 2017. [Online]. Available: https://www.bbc.com/news/world-europe-41443027
[24] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2387–2395.
[25] J. Thies, M. Zollhöfer, and M. Niessner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 66:1–66:12, Jul. 2019.
[26] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23012–23026, 2019.
[27] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–7.
[28] D. J. Robertson, R. S. S. Kramer, and A. M. Burton, "Fraudulent ID using face morphs: Experiments on human and automatic recognition," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173319.
[29] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1–11.
[30] M. Ngan, P. Grother, K. Hanaoka, and J. Kuo, "Face recognition vendor test (FRVT) part 4: MORPH performance of automated face morph detection," Nat. Inst. Technol. (NIST), Rourkela, India, Tech. Rep. NISTIR 8292, 2020.
[31] *Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework*, Standard ISO/IEC IS 19795-1:2006, 2006.
[32] M. Ferrara, A. Franco, and D. Maltoni, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*. Cham, Switzerland: Springer, 2016, pp. 195–222.
[33] E. Kee and H. Farid, "A perceptual metric for photo retouching," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 50, pp. 19907–19912, Dec. 2011.
[34] E. Kee, J. F. O'brien, and H. Farid, "Exposing photo manipulation from shading and shadows," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 165:1–165:21, Aug. 2014.
[35] W. Funk, M. Arnold, C. Busch, and A. Munde, "Evaluation of image compression algorithms for fingerprint and face recognition systems," in *Proc. 6th Annu. IEEE Syst., Man Cybern. (SMC) Inf. Assurance Workshop*, Jun. 2005, pp. 72–78.
[36] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.
[37] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 947–954.
[38] *Machine Readable Passports—Part 9—Deployment of Biometric Identification and Electronic Storage of Data in eMRTDs*, International Civil Aviation Organization (ICAO), Montreal, QC, Canada, 2015.
[39] Google. *Google Play*. Accessed: Jun. 2019. [Online]. Available: https://play.google.com/store.GoogleLLC
[40] Meitu Technology, Inc. *AirBrush: Easy Photo Editor for the Best Moments*. Accessed: Jun. 2019. [Online]. Available: http://www.appairbrush.com/en/
[41] Meitu Technology Inc. *BeautyPlus—Easy Photo Editor & Selfie Camera*. Accessed: Jun. 2019. [Online]. Available: http://global.meitu.com/
[42] PinGuo. *Bestie—Camera for Selfies*. PinGuo Inc. Accessed: Jun. 2019. [Online]. Available: https://play.google.com/store/apps/details?id=us.pinguo.selfie
[43] Perfect Corp. *YouCam Perfect—Foto Editor & Selfie Camera App*. Accessed: Jun. 2019. [Online]. Available: https://www.perfectcorp.com/app/ycp

[44] H. L. AD&D. *Automate*. LlamaLab. Accessed: Jun. 2019. [Online]. Available: https://llamalab.com/automate/
[45] C. Blüm. *Cliclick—Commandline Interface Click*. Accessed: Jun. 2019. [Online]. Available: https://www.bluem.net/de/projekte/cliclick/
[46] Apowersoft. *ApowerMirror—Mirror/Control Android Screen From PC*. Apowersoft Ltd. Accessed: Jun. 2019. [Online]. Available: https://www.apowersoft.com/phone-mirror
[47] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jul. 2009.
[48] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, T. Pajdla and J. Matas, Eds. Berlin, Germany: Springer, 2004, pp. 469–481.
[49] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019.
[50] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 74–109, Jan. 2019.
[51] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *Int. J. Comput. Vis.*, vol. 127, no. 2, pp. 115–142, Feb. 2019.
[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[53] *Best Practice Technical Guidelines for Automated Border Control ABC Systems*, Frontex, Warsaw, Poland, Sep. 2015.
[54] *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, Standard ISO/IEC IS 30107-3:2017, International Organization for Standardization and International Electrotechnical Committee, 2017.
[55] Cognitec: The Face Recognition Company. (2018). *FaceVACS Engine 9.3*. [Online]. Available: http://www.cognitec.com/technology.html
[56] A. Levandoski and J. Lobo. *Image Forgery Detection: Developing a Holistic Detection Tool*. Accessed: Feb. 2020. [Online]. Available: https://github.com/andrewlevandoski/Image-Forgery-Detection
[57] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1008–1017, Apr. 2018.
[58] U. Scherhag, C. Rathgeb, and C. Busch, "Towards detection of morphed face images in electronic travel documents," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Apr. 2018, pp. 187–192.

C. RATHGEB is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is a Principal Investigator of the National Research Center for Applied Cybersecurity (ATHENE). He coauthored over 100 technical articles in the field of biometrics. His research interests include pattern recognition, iris and face recognition, the security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems. He is a member of the European Association for Biometrics (EAB), the Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG), and an Editorial Board Member of *IET Biometrics* (IET BMT). He was a Winner of the European Biometrics Research Award 2012 from the EAB, the Austrian Award of Excellence 2012, the Best Poster Paper Awards (IJCB 2011, IJCB 2014, and ICB 2015), and the Best Paper Award Bronze (ICB 2018). He has served for various program committees and conferences, such as ICB, IJCB, BIOSIG, and IWBF, and journals as a Reviewer such as the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE, and IET BMT.

C.-I. SATNOIANU has completed the master's degree at Danmarks Tekniske Universitet, Lyngby, Denmark, on the topic of facial retouching detection. She did a research visit with the da/sec–Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany.

N. E. HARYANTO is currently pursuing the bachelor's degree with Hochschule Darmstadt, Germany. She works as a Student Research Assistant with the da/sec–Biometrics and Internet Security Research Group, Hochschule Darmstadt. Her research interests include biometrics, in particular facial recognition.

K. BERNARDO is currently pursuing the bachelor's degree with Hochschule Darmstadt, Germany. He works as a Student Research Assistant with the da/sec–Biometrics and Internet Security Research Group, Hochschule Darmstadt. His research interests include biometrics, in particular facial recognition.

C. BUSCH (Senior Member, IEEE) is a member of the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. Furthermore, he has been a Lecturer of biometric systems with the Technical University of Denmark (DTU), since 2007. He coauthored more than 400 technical papers and has been a speaker at international conferences. He is a convenor of WG3 in ISO/IEC JTC1 SC37 on biometrics and an active member of CEN TC 224 WG18. He served for various program committees, such as NIST IBPC, ICB, ICHB, BSI-Congress, GI-Congress, DACH, WEDELMUSIC, and EUROGRAPHICS, and served for several conferences, journals, and magazines as a Reviewer such as ACM-SIGGRAPH, ACM-TISSEC, the IEEE COMPUTER GRAPHICS AND APPLICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and the *Computers and Security* journal (Elsevier). Furthermore, on behalf of Fraunhofer, he chairs the biometrics working group of the TeleTrusT association as well as the German standardization body on biometrics (DIN-NIA37). He is also an Appointed Member of the Editorial Board of the *IET Biometrics* journal and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY journal.