This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3044723, IEEE Access

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Makeup Presentation Attacks: Review and Detection Performance Benchmark

C. RATHGEB[1], P. DROZDOWSKI[1], AND C. BUSCH[1] (Member, IEEE)

[1] da/sec – Biometrics and Internet-Security Research Group at Hochschule Darmstadt, Darmstadt, Germany (e-mail: christian.rathgeb@h-da.de)

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

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

ABSTRACT The application of facial cosmetics may cause substantial alterations in the facial appearance, which can degrade the performance of facial biometrics systems. Additionally, it was recently demonstrated that makeup can be abused to launch so-called makeup presentation attacks. More precisely, an attacker might apply heavy makeup to obtain the facial appearance of a target subject with the aim of impersonation or to conceal their own identity.

We provide a comprehensive survey of works related to the topic of makeup presentation attack detection, along with a critical discussion. Subsequently, we assess the vulnerability of a commercial off-the-shelf and an open-source face recognition system against makeup presentation attacks. Specifically, we focus on makeup presentation attacks with the aim of impersonation employing the publicly available Makeup Induced Face Spoofing (MIFS) and Disguised Faces in the Wild (DFW) databases. It is shown that makeup presentation attacks might seriously impact the security of face recognition systems. Further, we propose different image pair-based, i.e. differential, attack detection schemes which analyse differences in feature representations obtained from potential makeup presentation attacks and corresponding target face images. The proposed detection systems employ various types of feature extractors including texture descriptors, facial landmarks, and deep (face) representations. To distinguish makeup presentation attacks from genuine, i.e. bona fide presentations, machine learning-based classifiers are used. The classifiers are trained with a large number of synthetically generated makeup presentation attacks utilising a generative adversarial network for facial makeup transfer in conjunction with image warping. Experimental evaluations conducted using the MIFS database and a subset of the DFW database reveal that deep face representations achieve competitive detection equal error rates of 0.7% and 1.8%, respectively.

INDEX TERMS Biometrics, face recognition, presentation attack detection, makeup, makeup attack detection.

I. INTRODUCTION

BIOMETRIC recognition has quickly established itself as one of the most pertinent means of authenticating individuals in a reliable and fast manner by analysing their biological and/or behavioural characteristics [1], [2]. The constantly growing use of biometric systems requires security-related investigations of these technologies. Potential attack vectors against biometric systems were first established in [3]. Due to the fact that many biometric characteristics are not secret, in particular the face, so-called presentation attacks (PAs) or "spoofing" attacks represent one of the most critical attack vectors against biometric systems [4]. In contrast to software-based attacks, no access to the internal modules of a biometric system is necessary to launch PAs. The vulnerability of biometric systems with regard to PAs has been confirmed by experts in the past years and published in international media [5]. One of the earliest media effective examples was presented at Black Hat 2009 [6], one of the world's leading conferences on technical security. Researchers showed how facial recognition systems introduced by three different laptop manufacturers could be circumvented with photos of legitimate users. This vulnerability has since been listed in the National Vulnerability Database of the National Institute of Standards and Technology (NIST) [7].

Since then, many efforts have been made towards robust and reliable presentation attack detection (PAD) in the field

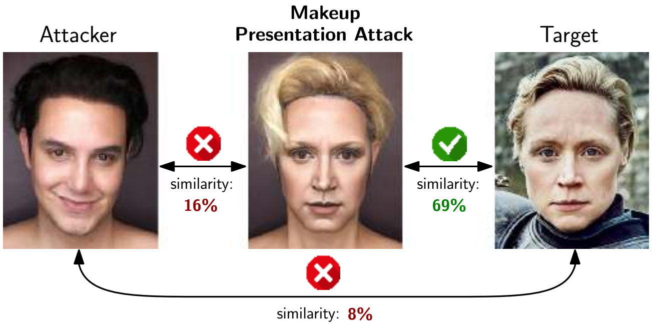


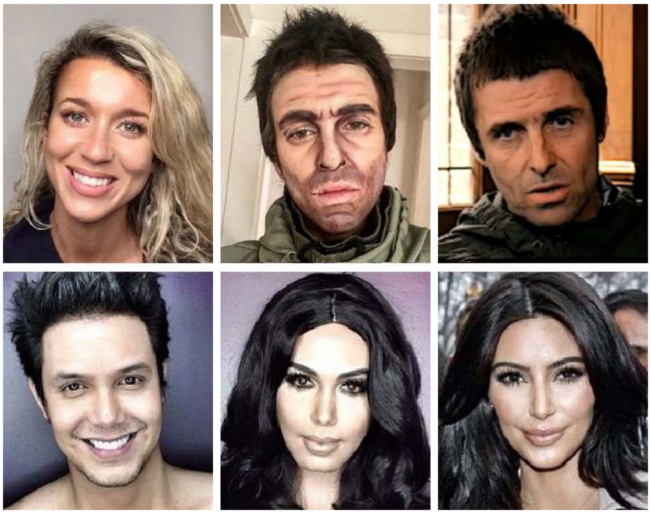
FIGURE 1: Illustration of a makeup presentation attack based on web-collected examples of face images of a makeup artist: before (left) and after the application of makeup (middle) with the intention of obtaining the facial appearance of a target subject (right). Similarity scores were obtained applying a COTS face recognition system.

of face recognition. Various research projects, *e.g.* EU-FP7 TABULA RASA [8] or IARPA Odin [9], have been conducted and numerous face PAD methods have been published in the scientific literature [10], [11]. Further, the International Organization for Standardization (ISO) achieved significant progress w.r.t. a standardised evaluation of PAs and PAD methods. In 2009, ISO/IEC JTC1 SC27 on information security, cybersecurity and privacy protection published the international standard ISO/IEC 19792 on "security evaluation of biometric data" [12], which contains a clause on the evaluation of weaknesses of biometric systems. In 2016, ISO/IEC JTC1 SC37 on biometrics published the standard ISO/IEC 30107 on "Biometric presentation attack detection" [13], which exclusively focuses on PAs and PAD. The standard ISO/IEC 19989 on "criteria and methodology for security evaluation of biometric systems" [14] is currently being prepared.

In the case of PAs, the attacker presents a so-called Presentation Attack Instrument (PAI) against a biometric capture device, *e.g.* a face printout or a 3D face mask, or attempts to disrupt the biometric system through their behaviour, *e.g.* movement of the head. The aim of the attacker is to be recognised as a (certain) target subject registered in the biometric system, *i.e. impersonation*, or to prevent being recognised, *i.e. concealment* [13]. According to [13], PAIs can be roughly divided into two classes: "artificial" and "human". Artificial PAIs, *i.e.* artefacts, are further divided into "complete" and "partial". The former refers to the creation of a fully artificial PAI, *e.g.* a display showing a face, a 3D face mask, or a face printout. The latter includes partial artificial PAIs, *e.g.* a fake nose or non-permanent makeup as stated in [13].

Makeup may substantially alter the perceived facial texture and shape which can pose a challenge to automated face recognition [15], [16]. Most prominently covered in the media, the Computer Vision Dazzle Camouflage campaign [17] showed how hairstyling and makeup designs can be used for identity concealment, *i.e.* to camouflage from face recognition technologies. When applied by skilled users or professional makeup artists, makeup can also be abused with the aim of impersonation [18]. In the latter case, makeup is applied in a way that the face of an attacker looks similar to that of a target subject, see figure 1, which is the core topic of this work. In 2013, the feasibility of such Makeup PAs (M-PAs) was first demonstrated in the TABULA RASA Spoofing Challenge [19], where a researcher put on makeup in such way that she was able to successfully impersonate a male target subject. Additionally, different makeup artists have showcased the possibility of transforming a face to that of a target subject through the mere application of makeup, see figure 2.

M-PAs pose a serious risk since they cannot be prevented by simply detecting makeup. Facial cosmetics are socially



(a) Before (b) After (c) Target

FIGURE 2: Web-collected examples of face images of makeup artists (a) before and (b) after the application of makeup with the intention of obtaining the facial appearance of (c) target celebrity. Note that in both examples the target subjects have a different sex than the one of the makeup artists.

TABLE 1: Popular PAIs and appropriate software- and hardware-based PAD methods proposed in the scientific literature. "Image, Video" refers to the visible spectrum. (Since makeup is also applied by bona fide subjects, hardware-based PAD methods are unsuitable for detecting M-PAs)

	PAD Analysis						
PAI	**Image, Video**	**Motion**	**Depth**	**Multi-spectral**	**Thermal**	**Light-field**	**Challenge response**
Printout	✓	✓	✓	✓	✓	✓	✓
Image on display	✓	✓	✓	✓	✓	✓	✓
Video on display	✓		✓	✓	✓	✓	✓
3D printout	✓	✓		✓	✓		✓
3D paper mask	✓			✓	✓		
3D silicone mask	✓			✓	✓		
Makeup	✓		(✓)	(✓)			

TABLE 2: Most relevant works on the topic of M-PAs and M-PAD.

Reference	Year	Database	Analysis	Contribution	Results	Remarks
Wang and Kumar [20]	2017	DMFaces	Concealment	Vulnerability analysis of various face recognition systems in verification mode	Significant drop in recognition performance for M-PAs	Database also contains different types of disguise
Chen *et al.* [18]	2017	MIFS	Impersonation	Vulnerability analysis of COTS and open-source face recognition systems in verification and identification mode	Higher comparison scores and ranks for M-PAs (compared to impostors)	—
Singh *et al.* [21], [22]	2019	DFW	Concealment, Impersonation	Vulnerability analysis of various face recognition systems for three degrees of difficulty in verification mode	Significant drop in recognition performance for M-PAs	Database also contains different types of disguise
Kotwal *et al.* [23]	2019	AIM	Concealment	Vulnerability analysis of open-source face recognition system in verification mode, CNN-based M-PAD system	Significant drop ($\sim$ 15%) of genuine comparison score, high M-PAD performance (D-EER $<$ 10%)	High detection performance on other makeup (not M-PA) database
Liu *et al.* [24]	2019	SiW-M	Concealment, Impersonation	M-PAD system based on Deep Tree Learning	Moderate M-PAD performance (D-EER $>$ 10%)	Method tested on numerous types of PAs
Rathgeb *et al.* [25]	2020	MIFS	Impersonation	Vulnerability analysis of open-source face recognition system in verification mode, M-PAD system based on depth data analysis and 3D reconstruction	Low IAPMRs ($<$ 15%) for relevant FMRs, moderate M-PAD performance ($\sim$ 20%)	Simulation of depth data through 3D reconstruction
Rathgeb *et al.* [26]	2020	MIFS	Impersonation	Vulnerability analysis of COTS face recognition system in verification mode, M-PAD system based on deep face representations	Low IAPMRs ($<$ 15%) for relevant FMRs, very high M-PAD performance (D-EER $<$ 1%)	Synthetic M-PAs for classifier training
Arab *et al.* [27]	2020	In-house	Concealment	M-PAD system based on GAN-based makeup removal and short-wave infrared images	High M-PAD performance (D-EER $\sim$ 2%)	Method tested on in-house M-PA database

acceptable in many parts of the world and cultural communities. They have become a daily necessity for many women to improve facial aesthetics in a simple and cost-efficient manner [15]. This is underlined by the huge and steadily growing market value of the facial cosmetics industry, *e.g.* €77.6 billion in Europe in 2017 [28] and $63 billion in the US in 2016 [29]. That is, the mere application of makeup *must not* be interpreted as a PA, in contrast to other face PAIs species which have mainly been considered in the scientific literature, *e.g.* face printouts or masks. Makeup might be used in an innocent manner (bona fide subjects, who are interacting with the capture device in the fashion intended by the policy of the biometric system). Nevertheless, it might as well be used in a malicious manner (by subjects with the intent to impersonate an enrolled target).

Due to the aforementioned reasons, a reliable detection of M-PAs turns out to be challenging and, so far, only relatively few research efforts have focused on the topic of M-PAD, *e.g.* in the Odin research program [30]. For such a nascent field of research, it is therefore desirable to establish a foundation w.r.t. existing works and available resources, evaluation metrics and protocols, as well as an overview of the suitable detection approaches and a benchmark thereof. Accordingly, this article serves the need of a single and comprehensive reference-point for future research in this field.

A. CONTRIBUTION AND ORGANISATION

The contributions of this work are:

- A comprehensive review and detailed discussion of related works dedicated to vulnerability assessment of M-PAs and recent approaches to M-PAD.
- An evaluation of the vulnerability of Commercial Off-The-Shelf (COTS) as well as open-source face recognition systems against M-PAs for impersonation. To this end, a standardised ISO/IEC methodology and metrics [31] is employed. Publicly available MIFS [32] and DFW [33] datasets are used.
- An investigation of different image pair-based (*i.e. differential*) M-PAD systems which make use of various feature extractors, including deep (face) representations, texture descriptors, and facial landmarks. In this differential detection scenario, the M-PAD systems take as input a potential M-PA and a target reference image to analyse differences between their features. Detection scores are then obtained by employing machine learning-based classifiers.
- A training database of M-PAs (and bona fide presentations) where the synthetic M-PA samples are generated using image warping and a Generative Adversarial Network (GAN) for facial makeup transfer. This synthetic database is shown to be suitable to train aforementioned machine learning-based classifiers of the proposed M-PAD methods.

This paper is organised as follows: related works are discussed in section II. The proposed M-PAD system is described in detail in section III. The experimental setup is summarised in section IV and experimental results are presented and discussed in section V. Finally, section VI contains a summary, concluding remarks, and future work items.

II. RELATED WORKS

In the recent past, diverse countermeasures against PAs, *i.e.* Presentation Attack Detection (PAD) methods, have been proposed for face recognition systems to prevent well-known attacks. For surveys on this topic, the interested reader is referred to [10], [11]. Published approaches can be categorised into software- and hardware-based PAD schemes where the latter make use of additional sensors, *e.g.* depth or near-infrared (NIR) capture devices [11]. It has been shown that different types of PAIs can be detected reliably, in particular by taking advantage of recent advances in deep learning-based image analysis and skin detection based on specialised hardware. Table 1 provides an overview of popular PAIs that have been investigated in the scientific literature along with indications which type of analysis has been reported for a reliable detection thereof. Due to the previously mentioned reasons, it can be observed that most conventional hardware-based PAD methods turn out to be less suitable for the detection of M-PAs. orblueMost recent works in general face PAD (*e.g.* in the context of the ChaLearn PAD challenge [34]) utilised depth and/or near-infrared data coupled with CNNs. While they achieved promising results, the requirement of additional sensors to acquire the multi-modal information may be prohibitive for operational systems, which overwhelmingly rely on 2D RGB images only.

All of the aforementioned calls for alternative Makeup-PAD (M-PAD) methods using only 2D RGB images, which reliably detect M-PAs, in particular the ones aimed at impersonation which is the focus of this work. The following subsections revisit related works on the impact of M-PAs on face recognition systems (subsection II-A) as well as M-PAD (subsection II-B) in more detail. Table 2 provides an overview of the most relevant works on both topics.

A. MAKEUP PRESENTATION ATTACKS

Makeup induces non-permanent alterations with the ability to substantially change the facial appearance. According to Dantcheva *et al.* [15], makeup can be applied mainly in three regions of the face, *i.e.* eyes, lips, and skin. Prominent examples of makeup alterations include changing of the perceived contrast of the eyes, size of the mouth, as well as skin colour [15], [16]. Further, the application of makeup can be categorised w.r.t. intensity [15], namely as light makeup (makeup cannot be easily perceived, since the applied colours correspond to natural skin, lip, and eye colours) and heavy makeup (makeup is clearly perceptible).

Dantcheva *et al.* [15] firstly investigated the impact of facial makeup on face recognition systems. Degradations in biometric performance were observed in the case where makeup has been applied either to the reference or probe image. Similar studies confirming these findings were presented in [35] and [20]. Further, Ueda and Koyama [36] have shown that the application of heavy makeup decreases humans' ability to recognise faces. The aforementioned early works make use of web-collected databases containing face image pairs before and after the application of mostly light makeup, *e.g.* [15], [35], which can be categorised as bona fide presentations.

In order to overcame potential performance drops resulting from makeup, various researchers have introduced specific face feature extraction and comparison techniques. Of the presented approaches some fused information obtained from multiple types of features, *e.g.* [37], [38], [39]. By maximising the amount of extracted biometric information, accuracy may improve. Within such approaches, potential improvements come at the cost of additional computational cost and appropriate biometric fusion techniques are required. Further, it has been suggested to use deep learning, *e.g.* [40], [41], in order to learn to extract makeup-resilient features from faces. More precisely, deep learning-based face recognition might be (re-)trained to gain robustness against facial makeup. Obviously, such schemes require a huge amount of training data. In addition, various methods to detect makeup have been proposed, *e.g.* [42], [43], [44], [45]. Such makeup detection schemes generally analyse facial colour, shape, and texture. In particular, skin features such as colour and smoothness were effectively extracted by applying suitable texture descriptors, *e.g.* Local Binary Patterns and Histogram of Oriented Gradients, together with machine learning-based classifiers. In case the use of facial makeup has been detected in a captured facial image, the face recognition system can react accordingly, *e.g.* by applying the feature extraction with different parameters [16].

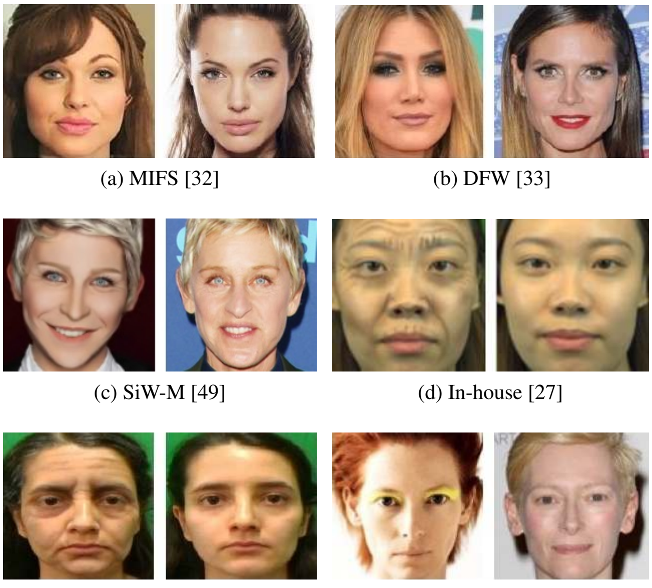
It is worth noting that the aforementioned research efforts devoted to makeup-resilient face recognition may additionally prevent M-PAs aiming at identity concealment. In other words, by achieving more tolerance and hence more robustness towards makeup-induced face alterations, M-PAs aiming at concealment are hampered. On the contrary, makeup-resilient face recognition algorithms may inadvertently facilitate M-PAs with the aim of impersonation. In case a face recognition system tolerates facial alterations resulting from the use of makeup, the success chance of impersonation M-PAs is expected to increase. This marks a potential disadvantage of the mentioned approaches w.r.t. the overall security of the face recognition system.

More recent works introduced databases containing images of faces with heavy makeup (un)intentionally applied for concealment or impersonation. Properties of face databases containing M-PAs are listed in table 3 and example images are depicted in figure 3. Kumar and Wang [20] reported performance degradations of face recognition, *i.e.* a decrease in genuine comparison scores, in the presence of heavy makeup. The application of heavy makeup to the probe sample of the image pair can be seen as concealment M-PAs. Kotwal *et al.* [23] investigated age-induced concealment M-PAs in which makeup is applied by professional artists to make the attacker look significantly older. Again, a large drop in genuine comparison scores was reported, which confirms the feasibility of this type of M-PA. Lastly, Singh *et al.* presented the first competition on disguised faces in the wild (DFW) [22]. On a web-collected database partially containing images with heavy makeup for the purpose of concealment, several submitted face recognition algorithms have been benchmarked [21]. Obtained results confirm the findings of previous works.

Chen *et al.* [18] were the first to investigate the potential of M-PAs for the purpose of impersonation. To this end, the authors introduced the MIFS database, which was collected from YouTube makeup video tutorials containing face images of subjects before and after the application of makeup, as well as images of target subjects. It was reported that different automated face recognition systems are also vulnerable to M-PAs. Recently, Rathgeb *et al.* [25], [26] confirmed these results for different state-of-the-art face recognition systems. Furthermore, the database introduced by Singh *et al.* [21]

TABLE 3: Overview of M-PA databases which are available for research purposes.

Database	M-PA Type(s)	Subjects	Samples
DMFaces [47]	Concealment	410	2,460
MIFS [32]	Impersonation	107	642
AIM [48]	Concealment	72	456
DFW [33]	Concealment, Impersonation	1,000	11,157
SiW-M [49]	Concealment, Impersonation	84	84
In-house [27]	Concealment	73	193



(a) MIFS [32]

(b) DFW [33]

(c) SiW-M [49]

(d) In-house [27]

(e) AIM [48]

(f) DMFaces [47]

FIGURE 3: Example image pairs of databases containing M-PAs. Left images in each pair show impersonation (a)-(c) and concealment M-PAs (d)-(f).

contains impersonation M-PAs leading to similar results. In the case of impersonation M-PAs, the associated impostor comparison scores might significantly improve, thus resulting in false matches. The chances of success for this type of M-PA increase if there is a certain degree of similarity in terms of soft biometric characteristics between the attacker and the target subject, *e.g.* sex or age, as well as facial geometry, *e.g.* eye distance or forehead height. Furthermore, appropriate know-how in handling makeup is required. If this is not available to an attacker, professional makeup artists could be hired or publicly available instructions and tutorials from the web could be consulted. This means that an impersonation M-PA requires a high degree of preparation compared to other PAIs. It should be noted that the durability of the PAI, *i.e.* the makeup, is only temporary. For this reason, makeup should ideally be applied shortly before the attack. In addition, it is difficult to guarantee consistency in the quality of the PAI, especially if the makeup is applied by different makeup artists. This hampers the evaluation of the effectiveness of such a PA.

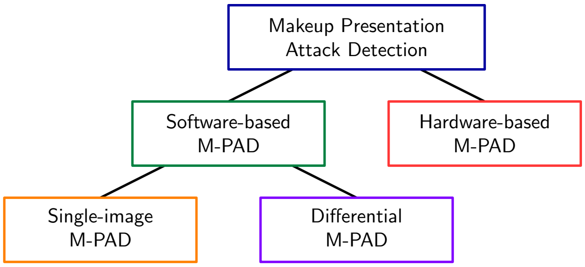


FIGURE 4: Conceptual categorisation of M-PAD methods.

Deviating from the traditional concepts of concealment and impersonation M-PAs, Zhu *et al.* [46] showed that the simulation of makeup in the digital domain can also be used to launch adversarial attacks.

B. MAKEUP PRESENTATION ATTACK DETECTION

At the time of this writing, only a few works have been published on the topic of M-PAD. Similar to traditional face PAD methods, published works on M-PAD can be categorised in software- and hardware-based methods. The former algorithms can be further divided into single-image and differential approaches, see figure 4.

Kotwal *et al.* [23] presented a single-image deep learning-based M-PAD system which was designed to detect age-induced M-PAs aimed at identity concealment. They proposed the use of a convolutional neural network (CNN) to extract features that can distinguish between presentations with age-induced facial makeups (attacks), and those without makeup (bona-fide). These feature descriptors, based on shape and texture cues, were constructed from multiple intermediate layers of a CNN. The authors report high average accuracy of 93.88%. Further, the authors presented cross-database experiments on different makeup databases including MIFS. Interestingly, the M-PAD scheme was also reported to achieve competitive detection performance on other databases in which makeup was applied for the purpose of facial beautification, *i.e.* bona fide samples. This might suggest that the M-PAD system of [23] detects different kinds of makeup. However, as mentioned before, the majority of subjects are not expected to wear makeup with the aim of identity concealment or impersonation, but merely with the intent to beautify the overall facial impression. In contrast, in this work it is assumed that bona-fide face images do not show makeup which may lead to high false-positive rates in real-world scenarios. Similarly, Arab *et al.* [27] recently proposed a method to detect age-induced concealment M-PAs (as well as other concealment M-PAs). To this end, the authors presented a GAN-based makeup removal technique. This technique is firstly applied to a given face image resulting in a "reconstructed" face image. Subsequently, the reconstructed image is compared against the original one using a conventional face recognition system. A competitive detection performance rate of approximately 2% D-EER was reported. Compared to the work in [23], this method is only expected to detect heavy use of makeup which in turn is interpreted as concealment M-PA. Moreover, in [27] it was suggested to employ a multispectral camera working in the short-wave infrared (SWIR) range of the electromagnetic spectrum. Selective experiments showed the potential of using SWIR bands in detecting malicious makeup. It is important to note that the use of additional sensors may prevent an economically cost-effective M-PAD. While the in-house database used in [27] only contains concealment M-PAs, the presented M-PAD scheme might as well be applied for detecting impersonation attacks.

Liu *et al.* [24] presented a generic face PAD scheme based on Deep Tree Learning (DTL) in order to partition potential PAs into semantic sub-groups in an unsupervised fashion. When a data sample is analysed, being a known or an unknown attack, DTN routes it to the most similar spoof cluster, and performs a binary decision. The method is evaluated on a newly introduced face database containing M-PAs aiming at concealment and impersonation, as well as twelve other types of face PAs [49]. In this challenging scenario, D-EER of approximately 50% and 10% were reported for concealment and impersonation attacks, respectively. In contrast to other related works, *e.g.* [23], [27], the authors considered different types of M-PAs in a more general scenario which led to the finding that M-PAs are hard to detect using a state-of-the-art software-based face PAD method. Traditional state-of-the-art PAD methods are usually based on deep learning and trained to detect certain types of artefacts resulting from the use of specific PAIs, *e.g.* moiré patterns on displays. Obviously, such artefacts do not appear in M-PAs.

In a preliminary study, Rathgeb *et al.* [25] presented one of the first M-PAD systems in the scientific literature with the aim of detecting impersonation M-PAs. The introduced M-PAD scheme was motivated by two observations: 1) Makeup might successfully change the appearance of a face image captured by an RGB capture device; 2) Makeup is not expected to significantly change the actual facial depth data of a subject's face. The proposed M-PAD scheme compared face depth data with face depth reconstructions obtained from RGB images of potential M-PAs. Significant variations between the two sources of information are expected to indicate facial shape alterations induced by strong use of makeup. Conceptual experiments on the MIFS database using simulated depth data confirmed the feasibility of the presented concept resulting in an average detection equal error rate (D-EER) of approximately 20%. Similar to the SWIR-based detection scheme suggested in [27], an apparent disadvantage of this method is that it requires an additional sensor. Moreover, it is expected that low-priced consumer depth cameras are not capable to capture depth data with sufficient precision to detect M-PAs, *i.e.* more expensive high precision depth sensors would be required. In another preliminary study, Rathgeb *et al.* [26] introduced the first differential M-PAD system to detect impersonation M-PAs. This M-PAD scheme analyses differences in deep face representations obtained from potential M-PAs and corresponding target face images and employs a machine learning-based classifier to distinguish M-PAs from bona fide presentations. Experimental evaluations on the MIFS database revealed a competitive D-EER of 0.7%. In this work, we build upon the idea of differential M-PAD for detecting impersonation M-PAs. Compared to the previously described M-PAD methods, an advantage of a differential M-PAD system is that it does not only rely on features extracted from single suspected face image. On the contrary, a differential M-PAD system allows for a comparison of features extracted from a given image pair (reference and probe face images). Such an image pair-based analysis might be vital to reliably detect M-PAs due to the previously mentioned fact that makeup might as well be used by bona fide subjects.

In summary, proposed single-image M-PAD methods make use of deep learning similar to state-of-the-art face PAD schemes and have been applied for detecting M-PAs with the aim of concealment and impersonation. In constrained scenarios, such algorithms reliably distinguish between subjects wearing makeup and those without makeup, see *e.g.* [23]. However, in more challenging but realistic scenarios, in which the mere detection of makeup should not be interpreted as M-PA, only moderate M-PAD detection performance is reported for such approaches, see *e.g.* [24]. A software-based removal of makeup appears to be a promising concept which would allow to unmask concealment as well as impersonation attacks. However, such approaches are strongly probabilistic and usually require a great amount of training data. In contrast, differential M-PAD schemes have shown competitive performance rates for detecting impersonation M-PAs in scenarios where bona fide subjects may also wear makeup, *e.g.* [26]. Besides a textural analysis of a single face image, a differential analysis allows for a detection of further, *e.g.* anatomical, differences between an M-PA and a corresponding target subject. Focusing on hardware-based M-PAD, the use of depth and SWIR analysis have been proposed [25], [27]. Besides the already mentioned disadvantage of additional sensor costs, so far, those schemes have either only been conceptualised and evaluated in rather constrained scenarios, *i.e.* evaluations on more realistic data are needed to confirm their worthiness.

III. MAKEUP PRESENTATION ATTACK DETECTION

The following subsections describe the considered detection scenarios (subsection III-A), the employed feature extraction methods (subsection III-B), and the synthetic generation of M-PAs (subsection III-C) which are subsequently used for machine learning-based classifier training (subsection III-D).

A. DETECTION SCENARIOS

In this work, we focus on software-based M-PAD methods which process face images in the visible spectrum. Software-based PAD mechanisms are cost-effective and generally easy to integrate into operational biometric recognition systems. Two different M-PAD scenarios, which are depicted in figure 5, can be distinguished:

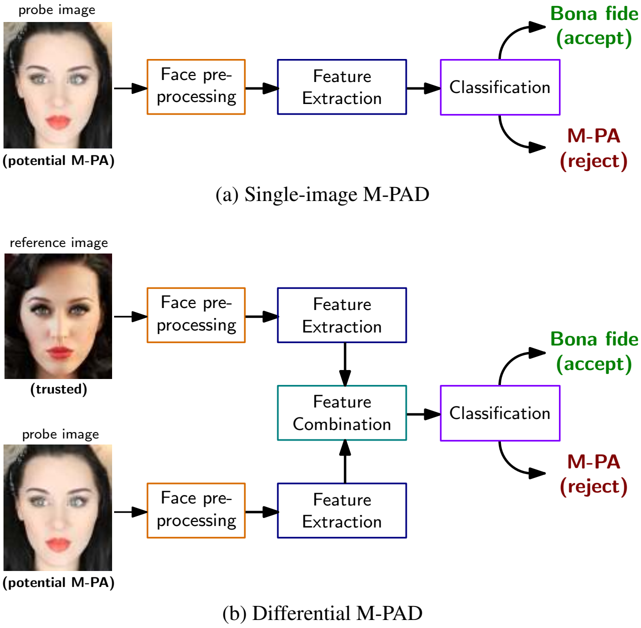


FIGURE 5: Overview of the M-PAD detection scenarios.

1) *Single-image M-PAD:* a suspected probe image serves as the only input to the M-PAD system.
2) *Differential M-PAD:* a pair consisting of a trusted reference image and a suspected probe image is processed by the M-PAD mechanism.

Compared to the differential M-PAD approach, a single-image-based M-PAD system is not expected to reliably detect M-PAs [25]. On the contrary, a single image-based M-PAD system which only analyses probe face images would most likely detect the mere presence of makeup which does not indicate M-PAs *per se*, since makeup might as well be used by bona fide subjects. A differential M-PAD system processes the stored reference image, in addition to the presented probe image. This allows for the analysis of differences between facial features extracted from a reference and a suspected probe image. Differences which indicate M-PAs can be learned in a training stage employing a machine learning-based classifier. Differential attack detection systems have already been successfully proposed for detection of face morphing [50] and facial retouching [51]. Therefore, main focus is put on differential M-PAD.

It is important to note that there is a key difference between concepts of differential image manipulation detection, *e.g.* morphing or retouching detection, and differential M-PAD: in differential M-PAD for the detection of impersonation M-PAs, the reference image, *i.e.* target subject, is considered as trusted (unaltered) face image. In an impersonation M-PA, an attacker tries to impersonate a distinct target subject. Such a target subject might have been previously enrolled to the face recognition system or the face image of the target subject is simultaneously presented during authentication, *e.g.* through an electronic travel document in automated border control. In such scenarios this is a reasonable assumption. In contrast, differential detection methods aiming to unveil digital image manipulation, *e.g.* [50], [51], consider the probe image to be a trusted live capture. In rare scenarios, a differential M-PAD might not be possible, *i.e.* if no (trusted) reference face image is available. For instance, if an M-PA is launched during the enrolment process of a face recognition system or in unconstrained, *e.g.* forensic, scenarios where the reference image is not considered as trusted. However, it is less likely that M-PAs would be performed in these scenarios, since PAs are usually targeted at actively gaining access to a biometric system.

B. FEATURE EXTRACTION

We consider different types of algorithms which can be used for feature extraction in a differential M-PAD scenario:

- *Deep Face Representations (DFR)*: deep face recognition systems employ hugh databases of face images to learn rich and compact representations of faces. Alterations induced by M-PAs would also be reflected in extracted deep face representations, *i.e.* information encoded in the parameters of the latent representations of the neural network. Said alterations are expected to be more pronounced if the application of makeup changes the perceived facial shape, since deep face recognition systems usually provide high generalisation capabilities w.r.t. variations in skin appearance.
  In principle, it would be possible to train a neural network from scratch or to apply transfer learning and re-train a pre-trained deep face recognition network to detect M-PAs. However, the high complexity of the model, represented by the large number of weights in the neural network, requires a large amount of training data. Even if only the lower layers are re-trained, the limited number of training images (and much lower number of subjects) in the available databases can easily result in overfitting to the characteristics of the training set.
- *Deep Representations (DR)*: analogous to the usage of deep face recognition networks, generic deep neural networks can be adapted for the task of M-PAD in the same way. While not explicitly trained for facial feature extraction, such networks have been reported to be suitable for similar tasks, in particular face PAD.
- *Facial Landmarks (FL)*: landmark detectors are used to extract two-dimensional facial landmarks from each reference and suspected probe face image. Extracted landmarks describe different facial features, *e.g.* the jawline, eyebrows, nose, eyes, and lips of a face. Facial landmark positions are normalised according to eye coordinates. Focusing on the task of M-PAD, positions of facial landmarks of the probe image might differ from that of the reference image if anatomical alterations induced by the M-PA do not precisely resemble that of the target subject. Similar schemes have been proposed for face image manipulation detection [51].

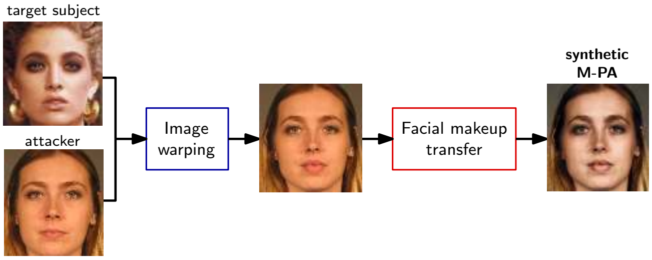


FIGURE 6: Processing steps in the generation of synthetic M-PAs.

- *Texture Descriptors (TD)*: at feature extraction, the aligned and cropped reference and probe images are converted to greyscale and divided into 4×4 cells to retain local information. Resulting feature values are aggregated in histograms. The final feature vector is constructed as a concatenation of histograms extracted from each cell. Texture descriptors have been found to be powerful features for texture classification. The texture descriptor-based feature vectors extracted from the reference and probe image are expected differ if the texture of the reference image differs from that of the probe image. Similarly, this concept has been proposed in [52] with the aim of face morphing detection.

C. GENERATION OF SYNTHETIC M-PAS

Since there exists no publicly available face database containing a sufficient number of M-PAs that could be employed to train an M-PAD system, we propose to automatically generate synthetic M-PAs. For this purpose, we apply two different face image transformations, see figure 6:

1) *Change of facial shape*: the shape of a face can be altered by heavy use of makeup, *e.g.* slimming of contour and nose or enlargement of eyes. In order to simulate said alterations with the aim of impersonating a target subject, image warping [53] is employed. More specifically, facial landmarks of a target reference image and the probe image are extracted, and image warping is applied to the probe face image w.r.t. the landmarks detected in the target reference image. Subsequently, the resulting probe face image exhibits the facial shape of the target reference image. This transformation is motivated by the fact that a skilled attacker, *e.g.* makeup artist, would be able to change the appearance of their own facial shape through the application of makeup.
2) *Change of facial texture*: the application of makeup can significantly change the perceived texture of a face.To simulate textural alterations caused by makeup for the purpose of impersonating a target subject, a GAN-based facial makeup transfer is applied (see subsection IV-A). GANs have enabled an automated transfer of full makeup styles, *e.g.* [54], [55]. Such transfer is motivated by the demand of users attempting to copy makeup styles of other individuals such as celebrities.

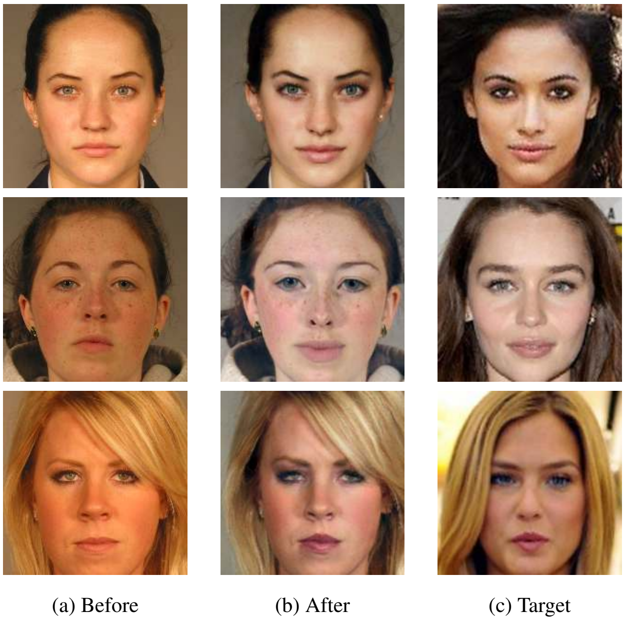


FIGURE 7: Examples of cropped face images (a) before and (b) after the generation of synthetic M-PAs intended to obtain the facial appearance of a (c) target subject.

The above described processing steps are applied to pairs of randomly chosen target reference images containing makeup and probe images of different subjects without makeup. For both types of images frontal pose, relatively neutral facial expression (*e.g.* closed mouth), and sample image quality are assured algorithmically. Figure 7 shows examples of resulting transformed probe images, *i.e.* synthetic M-PAs. It can be observed that the proposed synthetic M-PA generation algorithm produces M-PAs of reasonable quality. Therefore, alternative techniques of creating synthetic M-PAs, *e.g.* (re-)training an existing GAN architecture, were not considered but could be subject to future work. Synthetically generated M-PAs are used together with unaltered pairs of face images of the same subject which represent bona fide authentication attempts. It is important to stress that synthetically generated M-PAs are only used during the training of M-PAD methods. While GANs have been found to be extremely useful for generating photorealistic images, these may introduce specific noise to the generated images commonly referred to as GAN model fingerprint. Therefore, the use of synthetically generated M-PAs in both training and testing is deliberately avoided in order to prevent from any sort of overfitting towards potential GAN-induced model fingerprints.

The presented synthetic generation of M-PAs could be adapted in several ways. On the one hand, the image warping process could be applied with a randomised intensity in order to simulate different skill levels of attackers. On the other hand, multiple facial makeup transfer algorithms could be employed to improve robustness and avoid overfitting to potentially induced algorithm-specific artefacts. However, in the experimental setting used in this work, these adaptations did not reveal any improvements in terms of detection performance. Furthermore, the training and testing image sets were selected to be fully disjoint, stemming from different databases.

D. TRAINING AND CLASSIFICATION

At training and classification, pairs of feature vectors extracted from a reference and probe face image are combined by estimating difference vectors. Specifically, an element-wise subtraction of feature vectors is performed. For the facial landmark-based feature vectors, x- and y-coordinates are subtracted separately. Note that the 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 machine learning-based classifiers are trained to distinguish between bona fide and M-PA samples. More precisely, Support Vector Machines (SVMs) with Radial Basis Function (RBF) kernels are employed.

Due to the use of machine learning-based classifiers, more novel combinations of feature vectors is not necessary. More precisely, machine learning-based classifiers are trained to learn patterns in feature vector differences. For this purpose, a simple subtraction of feature vectors suffices as suitable weights will be assigned to distinct (combinations of) feature elements during the training of the classifier. In theory, other combinations of feature vectors could be considered. For instance, a concatenation of both feature vectors could be performed. However, this would double the size of the resulting feature vector, *i.e.* more time and data would be required during the training of machine learning-based classifiers. However, in the experimental setting used in this work, a concatenation of feature vectors did not improve the detection performance of M-PAD methods.

IV. EXPERIMENTAL SETUP

The following subsections describe the software and databases (subsection IV-A), as well as evaluation methodology and metrics (subsection IV-B) used in the proposed M-PAD system and in the experimental evaluations.

A. SOFTWARE AND DATABASES

The *dlib* algorithm (version 19.21.1) [56] is applied for face detection. Further, the algorithms of [57] and [58] implemented in *dlib* and *OpenCV* (version 4.4.0), respectively, are employed for facial landmark extraction each resulting in a feature vector of length 2×68. The detected eye coordinates are used for face alignment. Deep face representations are extracted using the well-known open-source VGG-Face [59] and ArcFace [60] systems. The extracted feature vectors consist of 128 (VGG-Face) and 512 (ArcFace) floating-point values. In addition, a COTS face recognition system is used in the vulnerability analysis. The use of the COTS face recognition system raises the practical relevance of the vulnerability analysis. While the COTS system is closed-source, it is assumed that it is based on deep learning as the vast majority of state-of-the-art face recognition systems. Therefore, it is only used in the vulnerability analysis, whereas open-source algorithms are used for the proposed M-PAD method. Focusing on deep representations, the DenseNet [61] and ResNet [62] networks are used; those networks have also been applied to general face PAD and it is therefore worth investigating if they are suitable for the M-PAD task. Extracted feature vectors consist of 1024 (DenseNet-121) and 2048 (ResNet-50) floating-point values. W.r.t. texture descriptors, Local Binary Patterns (LBP) [63] and Binarized Statistical Image Features (BSIF) [64] are extracted from each cell of the preprocessed face images. LBP and BSIF feature vectors are extracted employing a radius of 1, where eight neighboring pixel values are processed within 3×3 pixel patches. For details on the extraction of LBP and BSIF feature vectors, the reader is referred to [63], [64].

During the generation of synthetic M-PAs, image warping is applied using *OpenCV* with *dlib* landmarks and a re-implementation [65] of the *BeautyGAN* algorithm of Li *et al.* [55] is used for facial makeup transfer.

The *scikit-learn* library (version 0.23.2) [66] is used to train SVMs. Data-normalisation is applied as the feature elements of extracted feature vectors are expected to have different ranges. This is particularly the case in cross-database experiments and hence represents an essential processing step. The normalisation 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 regularisation parameter of $C = 1$ and a kernel coefficient Gamma of $1/n$ is used, where n denotes the number of feature elements. Trained SVMs generate a normalised attack detection score in the range $[0, 1]$.

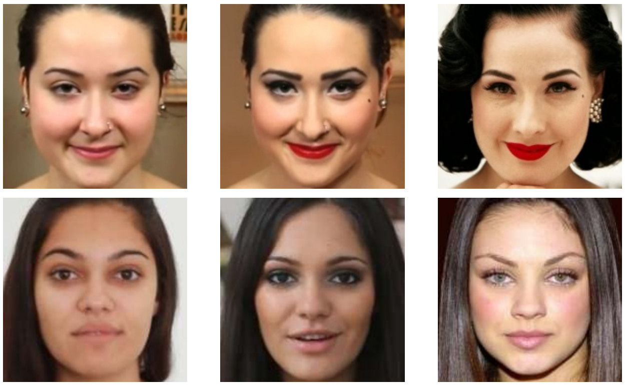
Table 4 gives an overview of the used face image databases and their purposes. The MIFS database consists of 642 images of 117 subjects. For each subject three categories of images are available, *i.e.* original face images, M-PAs, and face images of target subjects. Further, a subset (3,415 images of 880 subjects) of the DFW database was used. Specifically, comparisons labelled as “impersonation” in which makeup is (un)intentionally applied to obtain the facial appearance of a target subject are utilised. The total number of template comparisons in the vulnerability assessment was: 1,196 bona fide, 1,933 M-PA, and 483,636 zero-effort impostor attempts.

Example images of both databases are shown in figures 8 and 9, respectively. Said subset of the DFW database contains less constrained face images, thus counterbalancing the low intra-class variation of the MIFS database.

In the training stage of the proposed M-PAD system, a subset of the CelebA face database [67] was used as target references. To obtain this subset, the CelebA face database has been filtered to only contain face images with heavy

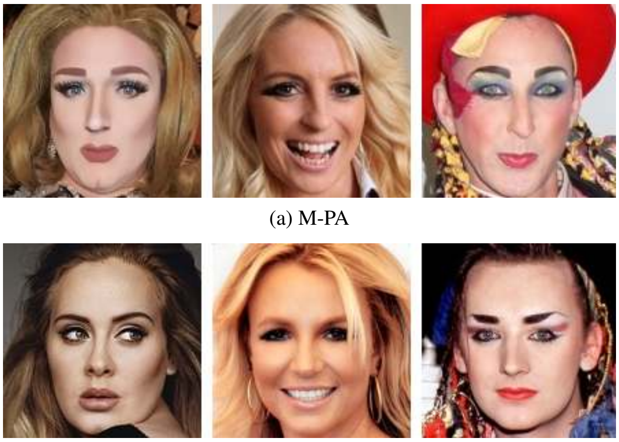
TABLE 4: Overview of used face databases.

System	Purpose	Bona Fide	M-PAs	Impostor
Face recognition	Vulnerability Assessment	MIFS, DFW	MIFS, DFW	DFW
M-PAD	Training	FRGCv2	Synthetic (FRGCv2, CelebA)	–
M-PAD	Testing	MIFS, DFW	MIFS, DFW	–



(a) Before (b) M-PA (c) Target

FIGURE 8: Examples of face images from the MIFS database (a) before and (b) after the application of makeup intended to obtain the facial appearance of a (c) target subject.



(a) M-PA

(b) Target

FIGURE 9: Examples of face images of the DFW database of (a) impersonations with makeup yielding the facial appearance of a (b) target subject.

use of makeup, frontal pose, and closed mouth. Face sample quality assurance has been conducted using the FaceQNet algorithm [68], resulting in a total number of 641 face images of different subjects. Example images of the resulting subset of the CelebA face database are depicted in figure 10. Images from the CelebA face database were randomly paired with face images from the FRGCv2 database [69] to generate 3,290 synthetic M-PAs which are used together with the genuine authentication attempts of the FRGCv2 database to train the proposed M-PAD scheme. Note that the synthetically generated M-PAs were only used for the purpose of M-PAD training and not during testing. In order to evaluate the detection performance of the M-PAD system, the aforementioned M-PAs and bona fide comparisons of the MIFS and the DFW databases were used.

FIGURE 10: Example face images of the subset of the CelebA database used for the training data generation for the proposed M-PAD system.

B. EVALUATION METRICS

The used evaluation metrics conform to the currently applicable international standards for biometric performance and presentation attack detection, *i.e.* ISO/IEC 19795-1:2006 and ISO/IEC 30107-3:2017 [70], [31]. Specifically, following metrics are reported:

- *Biometric performance:* the False Non-Match Rate (FNMR) and False Match Rate (FMR) denote the proportion of falsely classified genuine and impostor attempts in a biometric verification scenario.
- *Vulnerability assessment:* the Impostor Attack Presentation Match Rate (IAPMR) [31] defines the proportion of attack presentations using the same PAI species in which the target reference is matched. The Relative Impostor Attack Presentation Accept Rate (RIAPAR) establishes a relationship between IAPMR and $1-\text{FNMR}$, *i.e.* the biometric recognition performance of the attacked system, $\text{RIAPAR}=1+(\text{IAPMR}-(1-\text{FNMR}))$, as originally proposed by Scherhag *et al.* [71].
- *Attack detection performance:* the Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same PAI

TABLE 5: Descriptive statistics of score distributions.

System	Distribution	Mean	St. Dev.	Minimum	Maximum
COTS	Genuine	0.943	0.076	0.001	0.996
	Impostor	0.037	0.048	0.000	0.968
	M-PA MIFS	0.166	0.152	0.000	0.759
	M-PA DFW	0.240	0.184	0.001	0.988
ArcFace	Genuine	0.823	0.082	0.257	0.992
	Impostor	0.282	0.061	0.000	0.854
	M-PA MIFS	0.376	0.093	0.123	0.673
	M-PA DFW	0.431	0.096	0.145	0.900

TABLE 6: Vulnerability in relation to biometric performance (in %).

System	FMR	FNMR	IAPMR		RIAPAR	
			MIFS	DFW	MIFS	DFW
COTS	0.001	2.085	0.000	1.121	2.085	3.206
	0.010	0.751	1.878	4.746	2.629	5.497
	0.100	0.500	6.573	14.898	7.073	15.398
	1.000	0.334	29.577	44.034	29.911	44.368
ArcFace	0.001	1.090	1.168	2.742	2.258	3.832
	0.010	0.922	2.804	7.759	3.726	8.681
	0.100	0.838	7.243	21.940	8.081	22.778
	1.000	0.587	32.477	51.104	33.063	51.691

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 PAs in a specific scenario. Further, as suggested in the aforementioned standard, the BPCER10 and BPCER20 represent the operation points ensuring a security level of the system at an APCER of 10% and 5%, respectively. Additionally, the Detection Equal-Error Rate (D-EER) is reported.

V. EXPERIMENTS

The vulnerability of face recognition systems against M-PAs is analysed in subsection V-A. Subsequently, the detection performance of the proposed M-PAD systems is reported and discussed in subsection V-B. All conducted experiments were performed on the same platform using the same experimental protocol.

A. VULNERABILITY ANALYSIS

Table 5 and table 6 contain the descriptive statistics of score distributions and the results of the vulnerability assessment, respectively. Figure 11 show the histograms of said score distributions, as well as the resulting error rates. Note, that the usual dissimilarity scores of ArcFace have been normalised using min-max method and transformed to similarity scores for consistency of the table values and the plots.

The score distributions resulting from M-PAs lie between the genuine and impostor distributions. From table 5, it can be observed that the means of the M-PA distributions on MIFS and DFW are only moderately higher than those of the impostor distributions for both face recognition systems. However, the corresponding standard deviations are significantly greater since some M-PAs produce rather high similarity scores. As mentioned earlier, the chances of success for this type of M-PA highly depend on the degree of similarity between the attacker and the target subject as well as know-how in handling makeup. M-PAs yielding high similarity scores usually correspond to those of skilled makeup users. In other words, the attack potential of M-PAs can be arbitrarily high depending on the skill of the attacker.

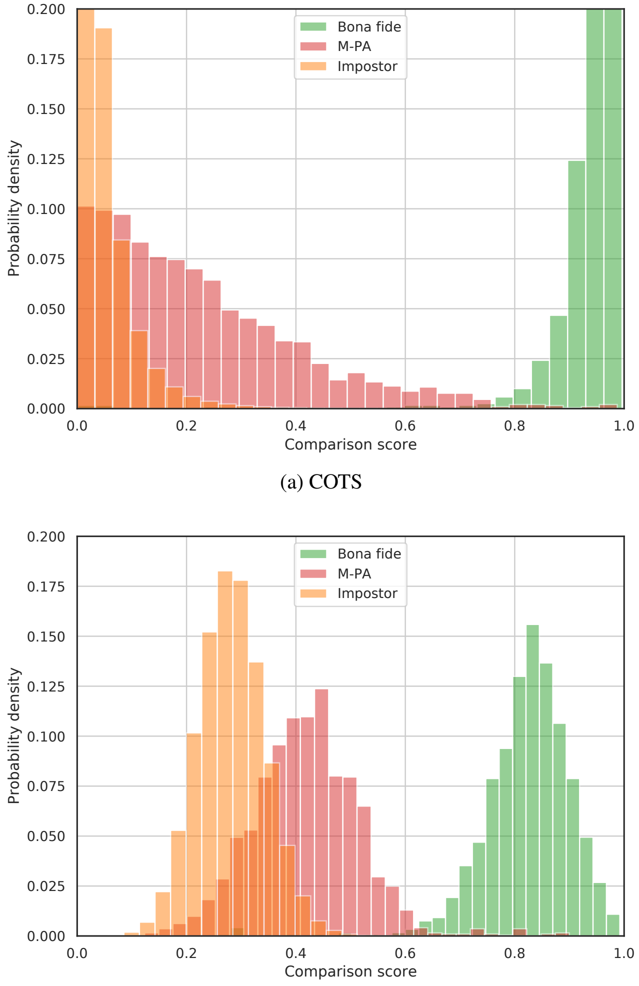
It can further be observed that for a practically relevant (see [72]) FMR of 0.1%, the success chances of the M-PAs (*i.e.* the IAPMR and RIAPAR values) are moderately high for the MIFS database (between 5% and 10%) and high for the considered subset of the DFW database (between 14% and 23%). At a more relaxed FMR of 1%, the IAPMR and RIAPAR values become extremely high, while still persisting at moderate to low levels for more stringent FMRs of 0.01% and 0.001%. Those results underline the vulnerability of face recognition systems, whether commercial or open-source, towards M-PAs. Compared to other types of face PAIs which might be detected more easily, *e.g.* printouts or masks, obtained IAPMR and RIAPAR values might be lower. Nonetheless, any deviation from the expected FMR a biometric system poses a vulnerability, since the FMR directly corresponds to the system's security. That is, even IAPMRs in the range of 10% pose a serious security risk and are thus unacceptable in operational deployments of face recognition.

Based on figure 12, it would in theory be possible to set a decision threshold with a high biometric performance and where nearly all of the M-PAs are rejected (*e.g.* around 0.6 in figure 12b). However, such an approach is linked to the specific database and recognition algorithm; furthermore, it assumes knowledge of the comparison score distributions. Consequently, it would be strongly affected by any changes to those factors (*e.g.* due to unknown attacks or degradation of sample quality). Hence, such a threshold-based system would not be flexible or generalise in a cross-database scenario. Contrary to the above, the system proposed in subsection V-B does not suffer from this limitation, as it is trained on synthetically generated M-PAs and bona fide images from a disjoint dataset.

B. DETECTION PERFORMANCE

In addition to the differential M-PAD methods, two M-PAD baseline systems are evaluated:

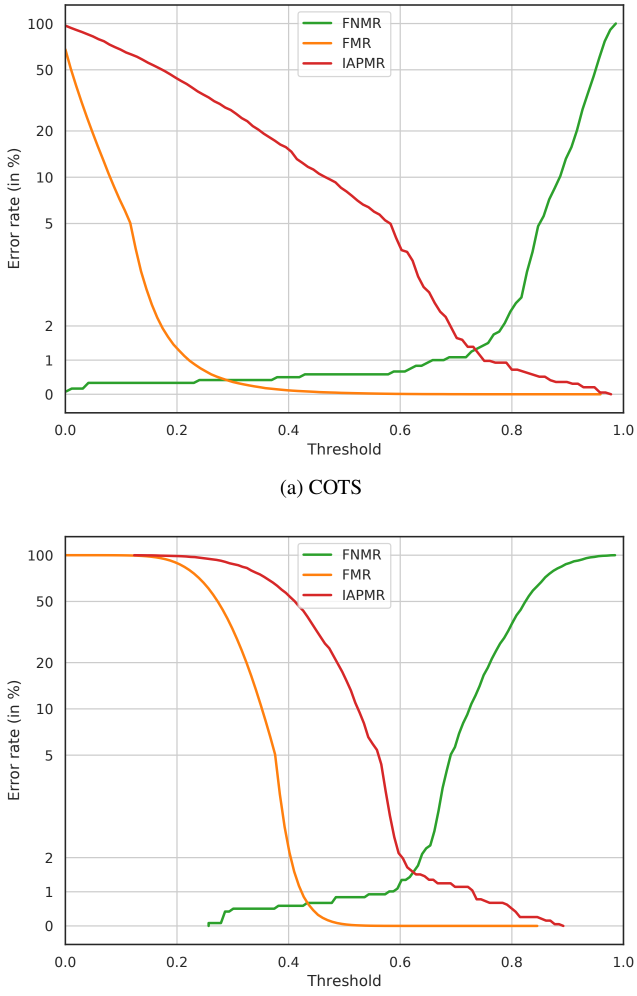
- *Depth Analysis (DA)*: the M-PAD approach of Rathgeb *et al.* [25] is based on the rationale that the depth data of an M-PA might significantly differ from the received depth data which is approximated using a depth image reconstruction method. This concept is simulated by extracting approximations of facial depth images from the reference and probe images employing the PRNet



(a) COTS

(b) ArcFace

FIGURE 11: Vulnerability assessment – comparison scores.



(a) COTS

(b) ArcFace

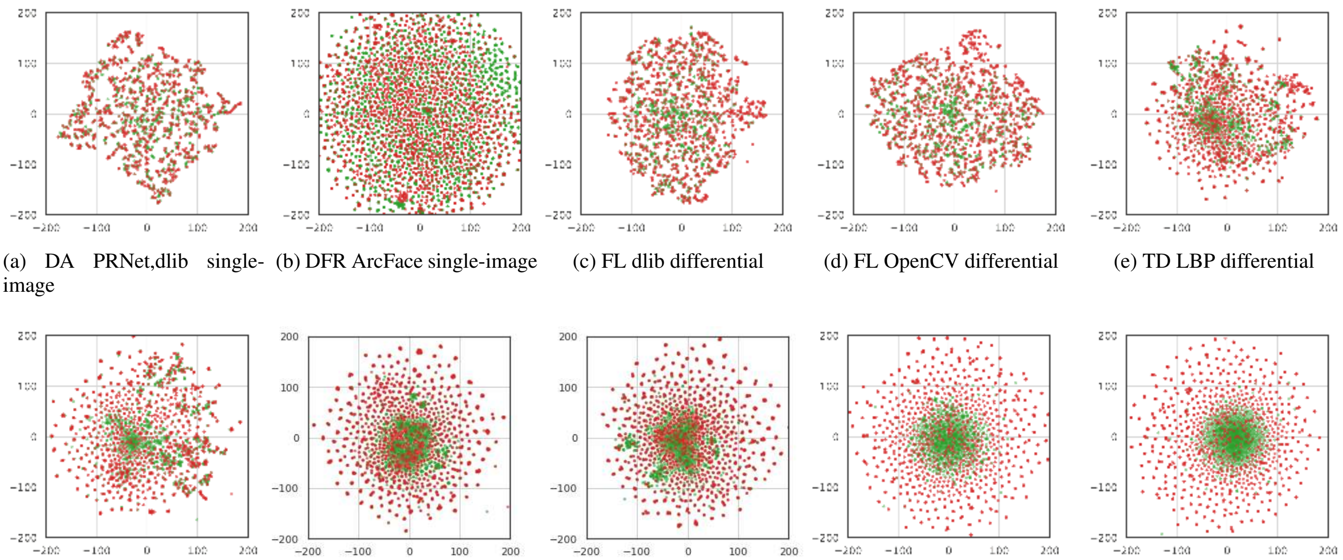
FIGURE 12: Vulnerability assessment – error rates.

algorithm [73]. The distance between the depth values at all *dlib* landmarks is computed using the mean squared error (MSE). The average pairwise MSE is returned as the final M-PAD score.

- *Single-Image Deep Face Representation*: deep face representations are extracted only from the suspected probe image using the *ArcFace* algorithm [60]. The extracted feature vector is then directly used to distinguish between M-PAs and bona fide authentications. The use of a single-image M-PAD scheme should reveal whether it is possible to detect M-PAs from single probe images. For this purpose, deep face representations are used since these represent rich textural as well as anatomical properties of face images. This scheme utilises the same training set and SVM-based classifier as the proposed differential M-PAD systems.
- *Image Quality (IQ)*: the use of (face) image quality has been suggested for the task of face PAD, *e.g.* in [74]. Hence, generic image quality estimators as well as face image sample quality assessment algorithms could also be applied for the purpose of M-PAD. Firstly, the general-purpose Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [75] is used. BRISQUE calculates a no-reference image quality score which is sensitive to various distortions, *e.g.* blur. In addition, the deep learning-based face image quality assessment algorithm FaceQNet [68] is employed.

While the suitability of additional recent CNN-based general face PAD may theoretically be considered for the M-PAD task (recall section II), those systems utilise additional information sources (depth and/or near-infrared images) which are not available for any of the publicly available M-PA datasets; conversely, the aforementioned systems do not offer pre-trained models for RGB image data alone.

In a pre-test, the suitability of the considered feature extractors is analysed employing the *scikit-learn* [66] implementation of the T-distributed Stochastic Neighbor Embedding (t-SNE) method [76]. t-SNE is a non-linear dimensionality reduction technique well-suited for embedding high-dimensional feature vectors for visualisation in a low-



(a) DA PRNet,dlib single-image (b) DFR ArcFace single-image (c) FL dlib differential (d) FL OpenCV differential (e) TD LBP differential

(f) TD BSIF differential (g) DR ResNet differential (h) DR DenseNet differential (i) DFR VGG-Face differential (j) DFR ArcFace differential

FIGURE 13: t-SNE plots for the benchmarked systems. The bona fide presentations and M-PAs are represented by green and red points, respectively.

TABLE 7: Error rates of the M-PAD systems (in %).

Feature Type	Mode	Feature Extractor	D-EER		BPCER10		BPCER20	
			MIFS	DFW	MIFS	DFW	MIFS	DFW
DA [25]	single-image	PRNet,dlib	21.864	45.759	41.414	92.386	51.515	96.818
DFR	single-image	ArcFace	38.785	44.860	83.652	89.953	90.068	92.991
IQ	single-image	BRISQUE	44.860	50.997	97.196	90.513	99.065	94.883
IQ	single-image	FaceQNet	39.720	50.113	69.159	91.067	79.439	95.556
FL	differential	dlib	19.484	40.768	28.125	78.132	45.313	87.813
FL	differential	OpenCV	25.837	41.116	43.770	76.257	52.716	86.433
TD	differential	LBP	33.803	45.168	93.569	86.023	97.917	91.250
TD	differential	BSIF	30.986	42.341	49.063	85.795	59.375	91.705
DR	differential	ResNet	33.645	48.981	66.140	87.617	74.051	94.318
DR	differential	DenseNet	35.514	50.756	62.025	85.909	74.367	91.932
DFR	differential	VGG-Face	4.984	11.824	4.361	15.040	4.984	26.177
DFR	differential	ArcFace	0.704	1.791	0.313	0.799	0.313	0.799

dimensional space. In case of single-image M-PAD, the feature vectors extracted from probe images are visualised. For differential M-PAD, difference vectors of corresponding feature vector pairs are employed. Resulting two-dimensional t-SNE plots for the combined MIFS and DFW databases are shown in figure 13. It can be observed that differential M-PAD based on DFR is expected to provide the best separation of bona fide and M-PA presentations. In contrast, single-image M-PAD based on DFR does not appear promising.

The performance rates of all M-PAD systems are listed in table 7, while figure 14 shows the DET curves for the best performing systems.

The single-image system based on ArcFace features alone is shown to achieve a poor detection performance of around 40% D-EER, which is close to guessing. This is to be expected, as explained in subsection III-A. The proof-of-concept single-image system based on reconstructed depth data achieves results around 20% D-EER on the MIFS dataset and over 45% on the DFW dataset. The poorer rates on the DFW dataset are presumably due to it being a more unconstrained dataset than MIFS is. Those results indicate the potential of utilising depth data for the purpose of M-PAD, although only in scenarios where high-quality sample and depth data might be available and where it is not possible to take advantage of the differential approach. Finally, the use of single-image quality metrics as a basis for M-PAD does not appear to be a suitable approach. The systems based on both general-purpose and face recognition specific metrics achieve very poor D-EERs of 40% and 50% for MIFS and DFW datasets, respectively.

The differential M-PAD systems based on texture descriptors and landmarks achieve a moderate to poor detection per-

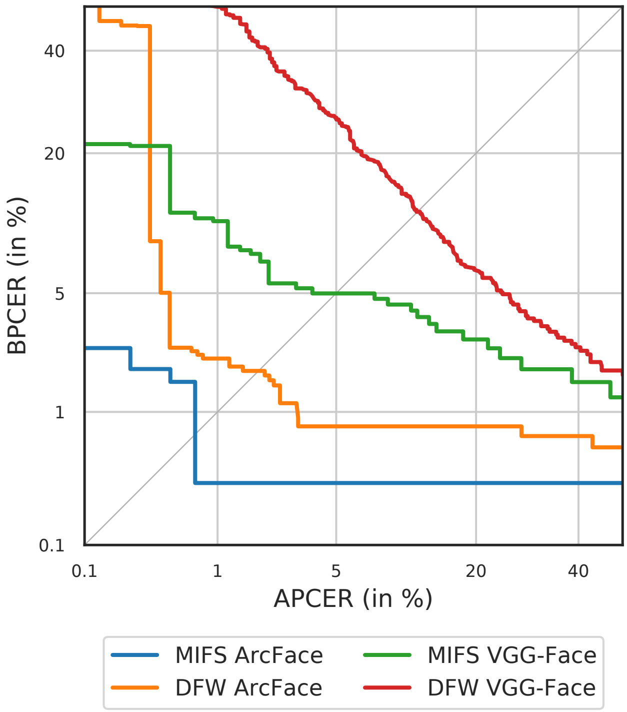


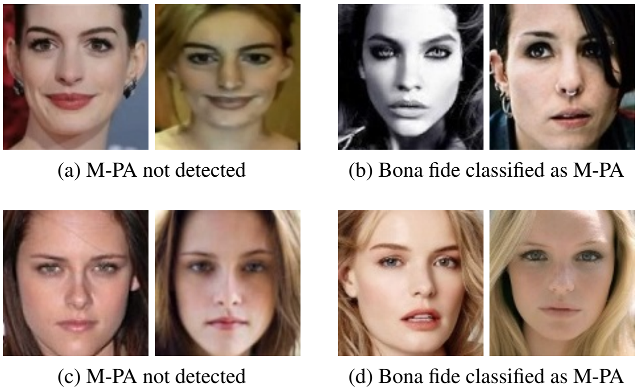
FIGURE 14: DET curves of DFR-based differential M-PAD.

formance of 19%-34% D-EER on the MIFS dataset, whereas the detection performance deteriorates on the selected subset of the DFW dataset, on which D-EERs of over 40% are achieved. A such, texture-based analysis is likely not suitable (at least on its own) to distinguish between bona fide presentations and M-PAs.

With general-purpose deep neural networks used as a basis for the differential M-PAD, poor D-EERs in the range of 35% are achieved for the MIFS dataset, whereas the detection performance is completely deteriorated when testing on the DFW dataset, where D-EERs around 50% are achieved. Those results suggest that general-purpose networks may not be sufficient for the M-PAD task and that fine-tuning to the specific domain of facial images may be necessary for achieving good detection results.

It can be observed that the differential M-PAD systems employing deep face representations significantly outperform all other schemes. Here, the system based on the more recent ArcFace achieves excellent results of below 1% and 2% D-EER for MIFS and DFW (subset) datasets, respectively. The system based on the older VGG-Face achieves higher, but still decent D-EERs of around 5% and 12% for the aforementioned datasets.

Figure 15 shows images of example failures for the ArcFace-based differential M-PAD system. Figure 15a represents an M-PA of very high quality, whereas figure 15c additionally couples the makeup application with the attacker being a look-alike of the target subject. The pairs of images for the false alarms in figures 15b and 15d exhibit high intra-class variation which might be the reason for the failed classifications.


(a) M-PA not detected

(b) Bona fide classified as M-PA

(c) M-PA not detected

(d) Bona fide classified as M-PA

FIGURE 15: Example images of detection failures by differential DFR-based M-PAD. In each pair, the reference image is to the left and the probe image to the right. Samples from MIFS databse in (a) and (b), from DFW database in (c) and (d).

VI. CONCLUSION

We assessed the vulnerability of a COTS and an open-source face recognition system against M-PAs. It was found that M-PAs of good quality, *i.e*. ones which mimic the facial texture as well as the shape of an impersonated target subject, can pose a serious risk to the security of face recognition systems. In contrast, M-PAs based on a simple makeup style transfer have a rather low success rate.

In addition, we benchmarked various *differential* M-PAD systems which analyse differences in features extracted from a pair of reference and probe face images. Detection scores were obtained from SVM-based classifiers which have been trained to distinguish difference vectors from a training set of bona fide presentations and synthetically generated M-PAs. In performance tests using the MIFS face database, the proposed M-PAD system employing deep face representations extracted by the ArcFace algorithm was shown to achieve encouraging D-EERs of approximately 0.7% and 1.8% on the MIFS and DFW (subset) databases, respectively. That is, the presented M-PAD schemes can effectively prevent M-PAs and hence improve the security of face recognition systems.

While the proposed M-PAD systems make use of a machine learning-based classifiers which are trained with a few thousand synthetically generated images, an end-to-end deep learning-based M-PAD system could technically be developed. However, such a system requires a huge amount of training data. Certainly, the presented generation of synthetic M-PAs would allow for the creation of a larger training database. This capability notwithstanding, the number of bona fide face images (in particular good quality reference images) is restricted by the employed databases. In order to avoid overfitting, large-scale face databases containing good quality images would be required to train an end-to-end deep learning-based M-PAD system.

Finally, it should be noted that the transition between PAs based on partial face masks and makeup is fluid. When using partial masks or mask elements, *e.g.* false nose, natural transitions can be created by applying makeup. For a masking that changes the appearance of a face, a professional makeup artist can use parts made of silicone (or other materials) that are glued to those areas of the face that are to be manipulated. Makeup is then used to hide transitions and create a natural and realistic look. With this approach, which is also used in film productions, the creation of an artefact takes a relatively long time. For example, the creation of such a mask in the film "The Associate" took several hours [77]. In addition, the creation of such an artefact can be very expensive compared to other types of artefacts, especially since such a partial mask might only be used once. There are currently no published studies on such an attack type. However, it can be assumed that it would be an effective PA against facial recognition systems. Since skin remains detectable on the areas of the face not covered with silicone, this type of attack might also overcome a PAD process based on material recognition.

Future work in the area of makeup presentation attacks might address the creation of larger evaluation databases suitable for training of end-to-end deep learning-based solutions (recall subsection III-B) and establishment of independent benchmarks (similar to the numerous evaluations carried out by NIST [78]). Furthermore, the impact of M-PAs on facial identification systems could be evaluated; lastly, it would be of interest to assess the demographic differentials [79] in success rates of M-PAs and M-PAD.

REFERENCES

[1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, August 2016.

[2] A. K. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*. Springer, July 2008.

[3] N. Ratha, J. Connell, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, March 2001.

[4] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, "Handbook of biometric anti-spoofing: Presentation attack detection," 2019.

[5] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, August 2015.

[6] Black Hat USA, https://www.blackhat.com/html/bh-usa-09/bh-usa-09-archives.html, 2009, last accessed: Oct. 2020.

[7] National Institute of Standards and Technology, "National Vulnerability Database," https://nvd.nist.gov/, 2020, last accessed: Oct. 2020.

[8] EU-FP7 Trusted Biometrics under Spoofing Attacks (TABULA RASA) , http://www.tabularasa-euproject.org/, 2016, last accessed: Oct. 2020.

[9] IARPA Odin, https://www.iarpa.gov/index.php/research-programs/odin, 2016, last accessed: Oct. 2020.

[10] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, December 2014.

[11] R. Raghavendra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–37, March 2017.

[12] ISO/IEC JTC1 SC27 Information security, cybersecurity and privacy protection, *ISO/IEC 19792. Information technology – Security techniques – Security evaluation of biometrics*, International Organization for Standardization and International Electrotechnical Committee, August 2009.

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

[14] ISO/IEC JTC1 SC27 Information security, cybersecurity and privacy protection, *ISO/IEC FDIS 19989-1. Information technology – Criteria and methodology for security evaluation of biometric systems – Part 1: Framework*, International Organization for Standardization and International Electrotechnical Committee, 2020.

[15] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, September 2012, pp. 391–398.

[16] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152 667–152 678, October 2019.

[17] A. Harvey, "Computer Vision Dazzle Camouflage," https://cvdazzle.com/, 2010, last accessed: Oct. 2020.

[18] C. Chen, A. Dantcheva, T. Swearingen, and A. Ross, "Spoofing faces using makeup: An investigative study," in *International Conference on Identity, Security and Behavior Analysis (ISBA)*. IEEE, February 2017, pp. 1–8.

[19] TABULA RASA Spoofing Challenge in conjunction with the 6th International Conference of Biometrics (ICB 2013), http://www.tabularasa-euproject.org/events/tabula-rasa-spoofing-challenge, 2013, last accessed: Oct. 2020.

[20] T. Y. Wang and A. Kumar, "Recognizing human faces under disguise and makeup," in *International Conference on Identity, Security and Behavior Analysis (ISBA)*. IEEE, February 2016, pp. 1–7.

[21] M. Singh, R. Singh, M. Vatsa, N. K. Ratha, and R. Chellappa, "Recognizing disguised faces in the wild," *Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, vol. 1, no. 2, pp. 97–108, March 2019.

[22] M. Singh, M. Chawla, R. Singh, M. Vatsa, and R. Chellappa, "Disguised faces in the wild 2019," in *International Conference on Computer Vision Workshop (ICCVW)*. IEEE, October 2019, pp. 542–550.

[23] K. Kotwal, Z. Mostaani, and S. Marcel, "Detection of age-induced makeup attacks on face recognition systems using multi-layer deep features," *Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, pp. 1–11, October 2019.

[24] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4675–4684.

[25] C. Rathgeb, P. Drozdowski, D. Fischer, and C. Busch, "Vulnerability assessment and detection of makeup presentation attacks," in *International Workshop on Biometrics and Forensics (IWBF)*. IEEE, April 2020, pp. 1–6.

[26] C. Rathgeb, P. Drozdowski, and C. Busch, "Detection of makeup presentation attacks based on deep face representations," 2020, arXiv 2006.05074.

[27] M. A. Arab, P. Azadi Moghadam, M. Hussein, W. Abd-Almageed, and M. Hefeeda, "Revealing true identity: Detecting makeup attacks in face-based biometric systems," in *Proceedings of the 28th ACM International Conference on Multimedia*. Association for Computing Machinery, 2020, p. 3568–3576.

[28] Cosmetics Europe, "Socio-economic contribution of the European cosmetics industry," https://www.ft.com/content/4721ed6a-f797-11e5-96db-fc683b5e52db, May 2018, last accessed: Oct. 2020.

[29] L. Whipp, "Changing face of cosmetics alters $63bn US beauty market," https://www.ft.com/content/4721ed6a-f797-11e5-96db-fc683b5e52db, April 2016, last accessed: Oct. 2020.

[30] L. Ericson, "Overview of the odin program on presentation attack detection, International Face Performance Conference (IFPC)," 2018.

[31] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 30107-3. Information Technology – Biometric presentation attack detection – Part 3: Testing and Reporting*, International Organization for Standardization and International Electrotechnical Committee, September 2017.

[32] Makeup Induced Face Spoofing (MIFS), http://antitza.com/makeup-datasets.html, 2017, last accessed: Oct. 2020.

[33] Disguised Faces in the Wild (DFW), http://iab-rubric.org/DFW/2019Competition.html, 2019, last accessed: Oct. 2020.

[34] A. Liu, J. Wan, S. Escalera, H. Jair Escalante, Z. Tan, Q. Yuan, K. Wang, C. Lin, G. Guo, I. Guyon, and S. Li, "Multi-modal face anti-spoofing attack detection challenge at CVPR2019," in *Conference on Computer Vision and*

Pattern Recognition Workshops (CVPRW). IEEE, June 2019, pp. 1601–1610.

[35] M. Eckert, N. Kose, and J.-L. Dugelay, "Facial cosmetics database and impact analysis on automatic face recognition," in *International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, September 2013, pp. 434–439.

[36] S. Ueda and T. Koyama, "Influence of make-up on facial recognition," *Perception*, vol. 39, no. 2, pp. 260–264, February 2010.

[37] A. Moeini, H. Moeini, F. Ayatollahi, and K. Faez, "Makeup-invariant face recognition by 3D face: Modeling and dual-tree complex wavelet transform from women's 2D real-world images," in *International Conference on Pattern Recognition (ICPR)*. IEEE, August 2014, pp. 1710–1715.

[38] N. Kose, L. Apvrille, and J.-L. Dugelay, "Facial makeup detection technique based on texture and shape analysis," in *International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1. IEEE, May 2015, pp. 1–7.

[39] C. Chen, A. Dantcheva, and A. Ross, "An ensemble of patch-based subspaces for makeup-robust face recognition," *Information Fusion*, vol. 32, no. B, pp. 80–92, November 2016.

[40] K. Zhang, Y. Chang, and W. Hsu, "Deep disguised faces recognition," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2018, pp. 32–324.

[41] N. Kohli, D. Yadav, and A. Noore, "Face verification with disguise variations via deep disguise recognizer," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2018, pp. 17–24.

[42] R. Feng and B. Prabhakaran, "Quantifying the makeup effect in female faces and its applications for age estimation," in *International Symposium on Multimedia (ISM)*. IEEE, December 2012, pp. 108–115.

[43] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *International Conference on Biometrics (ICB)*. IEEE, June 2013, pp. 1–8.

[44] G. Guo, L. Wen, and S. Yan, "Face authentication with makeup changes," *Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 24, no. 5, pp. 814–825, August 2014.

[45] S. Wang and Y. Fu, "Face behind makeup," in *Conference on Artificial Intelligence*. AAAI, February 2016, pp. 58–64.

[46] Z. Zhu, Y. Lu, and C. Chiang, "Generating adversarial examples by makeup attacks on face recognition," in *International Conference on Image Processing (ICIP)*. IEEE, September 2019, pp. 2516–2520.

[47] The Hong Kong Polytechnic University Disguise and Makeup Faces Database (DMFaces), https://www4.comp.polyu.edu.hk/~csajaykr/DMFaces.htm, 2016, last accessed: Oct. 2020.

[48] Age Induced Makeup (AIM), https://www.idiap.ch/dataset/aim, 2016, last accessed: Oct. 2020.

[49] Spoofing in the Wild with Multiple Attacks Database (SiW-M), http://cvlab.cse.msu.edu/siw-m-spoof-in-the-wild-with-multiple-attacks-database.html, 2019, last accessed: Nov. 2020.

[50] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Deep face representations for differential morphing attack detection," *Transactions on Information Forensics and Security (TIFS)*, 2020.

[51] C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch, "Differential detection of facial retouching: A multi-biometric approach," *IEEE Access*, vol. 8, pp. 106 373–106 385, June 2020.

[52] U. Scherhag, C. Rathgeb, and C. Busch, "Towards detection of morphed face images in electronic travel documents," in *International Workshop on Document Analysis Systems (DAS)*. IEEE, April 2018, pp. 187–192.

[53] C. A. Glasbey and K. V. Mardia, "A review of image-warping methods," *Journal of Applied Statistics*, vol. 25, no. 2, pp. 155–171, April 1998.

[54] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "PairedCycleGAN: Asymmetric style transfer for applying and removing makeup," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2018, pp. 40–48.

[55] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin, "BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network," in *International Conference on Multimedia (MM)*. ACM, October 2018, pp. 645–653.

[56] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research (JMLR)*, vol. 10, pp. 1755–1758, December 2009.

[57] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Conference on Computer Vision and Pattern Recognition (CVPR*. IEEE, June 2014, pp. 1867–1874.

[58] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 FPS via regressing local binary features," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2014, pp. 1685–1692.

[59] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference (BMVC)*. BMVA Press, September 2015, pp. 1–6.

[60] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2019, pp. 4690–4699.

[61] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[63] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European Conference on Computer Vision (ECCV)*. Springer, May 2004, pp. 469–481.

[64] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *International Conference on Pattern Recognition (ICPR)*. IEEE, Nov 2012, pp. 1363–1366.

[65] H. Zhang, "BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network," https://github.com/Honlan/BeautyGAN, last accessed: Oct. 2020.

[66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research (JMRL*, vol. 12, no. 85, pp. 2825–2830, October 2011.

[67] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *International Conference on Computer Vision (ICCV)*. IEEE, December 2015, pp. 3730–3738.

[68] J. Hernandez-Ortega, J. Galbally, J. Fiérrez, R. Haraksim, and L. Beslay, "FaceQnet: Quality assessment for face recognition based on deep learning," in *International Conference on Biometrics (ICB)*. IEEE, June 2019, pp. 1–8.

[69] 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 *Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, June 2005, pp. 947–954.

[70] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, International Organization for Standardization and International Electrotechnical Committee, April 2006.

[71] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis *et al.*, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, September 2017, pp. 1–7.

[72] Research and Development Unit, "Best practice technical guidelines for automated border control ABC systems," European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union, Tech. Rep. TT-02-16-152-EN-N, September 2015.

[73] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3D face reconstruction and dense alignment with position map regression network," in *European Conference on Computer Vision (ECCV)*. Springer, September 2018, pp. 534–551.

[74] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, 2014.

[75] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[76] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research (JMLR)*, vol. 9, pp. 2579–2605, November 2008.

[77] The Christian Science Monitor, "In 'Associate,' Whoopi Shatters Glass Ceiling," https://www.csmonitor.com/1996/1107/110796.feat.film.1.html, 1996, last accessed: Oct. 2020.

[78] National Institute of Standards and Technology, "Biometric evaluations homepage," https://www.nist.gov/itl/iad/image-group/resources/biometrics-evaluations, last accessed: Oct. 2020.

[79] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, "Demographic bias in biometrics: A survey on an emerging challenge," *Transactions on Technology and Society (TTS)*, vol. 1, no. 2, pp. 89–103, June 2020.

DR. CHRISTIAN RATHGEB is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is a Principal Investigator in the National Research Center for Applied Cybersecurity ATHENE. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design and privacy enhancing technologies for biometric systems. He co-authored over 100 technical papers in the field of biometrics. He is a winner of the EAB - European Biometrics Research Award 2012, the Austrian Award of Excellence 2012, Best Poster Paper Awards (IJCB'11, IJCB'14, ICB'15) and the Best Paper Award Bronze (ICB'18). He is a member of the European Association for Biometrics (EAB), a Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG) and an editorial board member of IET Biometrics (IET BMT). He has served for various program committees and conferences (*e.g*. ICB, IJCB, BIOSIG, IWBF) and journals as a reviewer (*e.g*. IEEE TIFS, IEEE TBIOM, IET BMT).

DR. PAWEL DROZDOWSKI is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. His research interests include biometrics, information security and privacy, pattern recognition, and algorithmic fairness. He co-authored over 20 technical publications in the field of biometrics. He won the Best Student Paper Runner-Up Award (WIFS'18) and Best Poster Award (BIOSIG'19). He is a member of the European Association for Biometrics (EAB) and represents the German Institute for Standardization (DIN) in ISO/IEC SC37 JTC1 SC37 on biometrics.

PROF. DR. CHRISTOPH BUSCH is member of the Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with Hochschule Darmstadt (HDA), Germany. Further he lectures Biometric Systems at Denmark's DTU since 2007. On behalf of the German BSI he has been the coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg and NFIQ2.0. He was/is partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD, RESPECT, TReSPsS, iMARS and others. He is also principal investigator in the German National Research Center for Applied Cybersecurity (ATHENE) and is co-founder of the European Association for Biometrics (EAB). Christoph co-authored more than 500 technical papers and has been a speaker at international conferences. He is member of the editorial board of the IET journal on Biometrics and formerly of the IEEE TIFS journal. Furthermore, he chairs the TeleTrusT biometrics working group as well as the German standardization body on Biometrics and is convenor of WG3 in ISO/IEC JTC1 SC37.