Makeup Presentation Attack Potential Revisited: Skills Pay the Bills

P. Drozdowski, S. Grobarek, J. Schurse, C. Rathgeb, F. Stockhardt, C. Busch
dasec – Biometrics and Internet Security Research Group
Hochschule Darmstadt, Germany
{pawel.drozdowski,christian.rathgeb,fabian.stockhardt,christoph.busch}@h-da.de

Abstract—Facial appearance can be substantially altered through the application of facial cosmetics. In addition to the widespread, socially acceptable, and in some cases even expected use for the purpose of beautification, facial cosmetics can be abused to launch so-called makeup presentation attacks. Thus far, the potential of such attack instruments has generally been claimed to be relatively low based on experimental evaluations on available datasets.

This paper presents a new dataset of such attacks with the purpose of impersonation and identity concealment. The images have been collected from online sources, concentrating on seemingly highly skilled makeup artists. A vulnerability assessment of face recognition with respect to probe images contained in the collected dataset is conducted on state-of-the-art open source and commercial off-the-shelf facial recognition systems with a standardised methodology and metrics. The obtained results are especially striking for the impersonation attacks: the obtained attack success chance of almost 70% at a fixed decision threshold corresponding to 0.1% false match rate is significantly higher than results previously reported in the scientific literature.

Index Terms—Biometrics, face recognition, makeup presentation attacks, concealment, impersonation

I. INTRODUCTION

Presentation attacks (PAs), frequently referred to as “spoofing”, pose a serious threat to face recognition systems [1], [2]. To launch a PA, the attacker presents a so-called presentation attack instrument (PAI), e.g. a face printout or a 3D face mask, to a biometric capture device or attempts to disrupt the biometric system through their behaviour, e.g. movement of the head. That is, the goal of the attacker is either to be recognised as a (certain) target subject registered in the biometric system, i.e. impersonation, or to prevent being recognised, i.e. concealment [3]. In the past years, various researchers have confirmed the vulnerability of face recognition systems to PAs, in particular state-of-the-art systems utilising deep convolutional neural networks [4].

Besides the aforementioned common PAIs, facial cosmetics may also be used to perform so-called makeup PAs (M-PAs) [5]. Makeup can be applied in a way that it substantially alters the perceived facial texture and shape which poses a challenge to automated face recognition [6], [7]. In 2010, the Computer Vision Dazzle Camouflage campaign [8] showed how makeup designs can be applied for identity concealment, i.e. to camouflage from face detection. When applied by skilled users or professional makeup artists, M-PAs may also be performed with the aim of impersonation [9]. In this case, makeup is applied such that an attacker’s face looks similar to that of a target subject, while concealing the facial appearance of the attacker, see figure 1. In 2013, a female researcher successfully impersonated a male target subject by putting on makeup in the TABULA RASA Spoofing Challenge [10]. Moreover, on social media various makeup artists have showcased the feasibility of transforming their facial appearance to that of a target subject through the mere application of makeup.

Recently, different researchers have studied the vulnerability of face recognition systems to M-PAs for impersonation and concealment. For a comprehensive survey on M-PAs and detection approaches the interested reader is referred to [5]. Additionally, some face datasets containing M-PAs have been made available by different research laboratories. Evaluations on these datasets have revealed a moderate attack potential, i.e. success chance, of M-PAs which diminishes with rather restrictive decision thresholds [11]. However, it was also found that the chances of success for M-PAs can be arbitrarily high depending on the skill of the attacker and the degree of similarity between the attacker and the target subject in case of impersonation [5].

In this work, a new facial image dataset containing M-PAs for concealment and impersonation is introduced. In contrast to existing datasets, the collected dataset comprises M-PAs involving highly skilled makeup artists. The dataset is...

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.
made available to the research community\(^1\). A vulnerability analysis using state-of-the-art open source and commercial-of-the-shelf (COTS) face recognition systems reveals that the attack potential of M-PAs has been clearly underestimated in previous works.

The remainder of this paper is organised as follows: section II briefly revisits related works on M-PAs. The data collection is described in section III and the vulnerability assessment is presented in section IV. Finally, conclusions are drawn in section V.

II. RELATED WORK

Dantcheva \textit{et al.} [6] first investigated the impact of daily makeup applied by bona fide subjects, \textit{i.e.} bona fide makeup, on face recognition performance. It was found that the changes in facial texture and shape induced by makeup can negatively affect face recognition. Motivated by these findings, makeup-resilient face recognition schemes have been proposed by different research groups, see [7] for a survey on the effects of facial beautification on face recognition.

More recent works introduced datasets containing images of faces with heavy makeup (un)intentionally applied for concealment or impersonation. Properties of existing face datasets containing M-PAs are listed in table I. Example images of said datasets are shown in figure 2. Most available datasets contain web-collected facial images while some have been created with the help of professional makeup artists.

Different datasets contain various types of concealment M-PAs. Kumar and Wang [12] introduced the Disguise and Makeup Faces (DMFaces) dataset which comprises face images with heavy makeup. The application of heavy makeup contained in the probe sample of the image pair can be seen as concealment M-PAs for which the authors observed a decrease in genuine comparison scores various face recognition systems. Kotwal \textit{et al.} [13] investigated age-induced concealment M-PAs and published the Age Induced Makeup (AIM) dataset. For this dataset, makeup was applied by professional artists to make the attacker look significantly older. Using an open source face recognition system, the authors observed a large drop in mated comparison scores (\textasciitilde15\%), which confirms the feasibility of this type of M-PA. A similar in-house dataset was used by Arab \textit{et al.} [14]. Singh \textit{et al.} [15] presented the first competition on Disguised Faces in the Wild (DFW) dataset. This web-collected dataset partially contains images with heavy makeup for the purpose of concealment. Several submitted face recognition algorithms have been benchmarked and obtained results confirm the findings of previous works [16]. Finally, a small number of M-PAs are contained in the Spoof in the Wild dataset with Multiple Attack Types (SiW-M) introduced by Liu \textit{et al.} [17].

Focusing on impersonation M-PAs, Chen \textit{et al.} [9] introduced the Makeup Induced Face Spoofing (MIFS) dataset, 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 COTS and open source face recognition systems are vulnerable to impersonation M-PAs. Recently, Rathgeb \textit{et al.} [11]. [23] confirmed these results for different state-of-the-art face recognition systems reporting success rates of approximately 8\% and 30\% for decision thresholds corresponding to false match rates (FMRs) of 0.1\% and 1\%, respectively. Additionally, on the DFW dataset introduced by Singh \textit{et al.} [16], which contains impersonation M-PAs, Rathgeb \textit{et al.} [5] obtained significantly higher success chances for impersonation M-PAs, \textit{i.e.} up to 20\% at a FMR of 0.1\% and 50\% for a FMR of 1\%. It was found that the chances of success for impersonation M-PAs increase if there is a certain degree of similarity in terms of soft biometric characteristics between the attacker and the target subject, \textit{e.g.} sex or age, as well as facial geometry, \textit{e.g.} eye distance or forehead height. In summary, the evaluation of the effectiveness of impersonation M-PAs is generally conceded as difficult, since the attack potential depends on various factors such that consistency in the quality of M-PAs can not be guaranteed, especially if the makeup is applied by different makeup artists [5].

\begin{table}
\centering
\caption{Overview of M-PA datasets which are available for research purposes}
\begin{tabular}{|c|c|c|c|}
\hline
Dataset & M-PA Type(s) & Subjects & Samples \\
\hline
DMFaces [18] & Concealment & 410 & 2,460 \\
AIM [19] & Concealment & 72 & 456 \\
In-house [14] & Concealment & 73 & 193 \\
DFW [20] & Concealment, Impersonation & 1,000 & 11,157 \\
MIFS [21] & Impersonation & 107 & 642 \\
SiW-M [22] & Concealment, Impersonation & 84 & 84 \\
\hline
\end{tabular}
\end{table}

\begin{figure}[h]
\centering
\subfigure[DMFaces [18]]{\includegraphics[width=0.24\textwidth]{fig2a.png}} \quad \subfigure[AIM [19]]{\includegraphics[width=0.24\textwidth]{fig2b.png}} \\
\subfigure[In-house [14]]{\includegraphics[width=0.24\textwidth]{fig2c.png}} \quad \subfigure[DFW [20]]{\includegraphics[width=0.24\textwidth]{fig2d.png}} \\
\subfigure[MIFS [21]]{\includegraphics[width=0.24\textwidth]{fig2e.png}} \quad \subfigure[SiW-M [22]]{\includegraphics[width=0.24\textwidth]{fig2f.png}}
\caption{Example image pairs of datasets containing M-PAs. Left images in each pair show concealment (a)-(c) and impersonation M-PAs (d)-(f).}
\end{figure}

\(^1\)https://dasec.h-da.de/research/biometrics/hda-facial-makeup-presentation-attack-database/
Images in the presented dataset were found through manual Internet search. The collected dataset consists of two parts:

**Bona fide makeup** In recent years, many videos of makeup application were posted on popular public video sharing platforms. Among them are videos where a model is shown applying and wearing makeup\(^2\). From such videos, a reference image without makeup along with multiple (up to 20) images with different makeup styles were extracted.

**Concealment and impersonation makeup** Recently, instructions on how to imitate the facial appearance of another individual, as shown in figure 1, have gained popularity on video platforms. Several makeup artists have posted images and/or videos\(^3\), where they demonstrate their skills. From such sources triplets of images (artist without makeup, artist with makeup, and target reference) were extracted. Images of artists without and with makeup represent concealment M-PAs and artists with makeup and the corresponding target reference are used as impersonation M-PAs.

The key aim of the data collection was that the collected images be of high quality (which is in contrast to many previously published datasets, recall section II). This was ensured in several ways: only high definition (1080p) videos were considered; from each of the found videos, frames corresponding to different makeup styles were extracted and subsequently manually selected based on general image quality (e.g., considering blur and sharpness). Furthermore, extreme facial expressions, partially covered faces, and strongly non-frontal poses were avoided. Table II provides a numerical overview of the collected dataset, while example images are shown in figure 3.

### TABLE II: Overview of the collected dataset

<table>
<thead>
<tr>
<th>Makeup type</th>
<th>Average resolution</th>
<th>Neutral</th>
<th>Makeup</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bona fide</td>
<td>552 × 757</td>
<td>33</td>
<td>366</td>
<td>—</td>
</tr>
<tr>
<td>Concealment</td>
<td>493 × 542</td>
<td>41</td>
<td>88</td>
<td>—</td>
</tr>
<tr>
<td>Impersonation</td>
<td>500 × 547</td>
<td>7</td>
<td>54</td>
<td>54</td>
</tr>
</tbody>
</table>

To facilitate reproducible research and further experiments in this field, the collected dataset along with the links to all the used source videos and websites are made available upon request.

### IV. VULNERABILITY ASSESSMENT

Subsection IV-A describes the details of the experimental setup for the conducted vulnerability assessment, while subsection IV-B presents and discusses the obtained results.

#### A. Experimental Setup

The vulnerability analysis on the collected datasets was conducted using a strong open source system (ArcFace [24]) with a pre-trained model provided by its authors. ArcFace produces feature vectors of 512 elements, whose dissimilarity can be computed using Euclidean distance. For the purposes of visualisation of the results, those dissimilarity scores were mapped into the range \([0, 1]\) using min-max normalisation and converted into similarity scores. While the use of this publicly available and well-known tool ensures the reproducibility of the experiments, an evaluation with a state-of-the-art commercial off-the-shelf (COTS) system was additionally conducted to increase the practical relevance of the obtained results.

To establish a large reference set of mated and non-mated comparison scores, as well as decision thresholds for several fixed, operationally relevant security levels, a subset of FRGCv2 dataset [25] yielding around 9,000 mated and 5,000,000 non-mated comparisons was used.

The results of the vulnerability analysis are reported using metrics standardised by ISO/IEC [26], [27]. Specifically, biometric recognition performance is reported using false match rate (FMR) and false non-match rate (FNMR); the efficacy of concealment attacks is reported using concealer attack presentation non-match rate (CAPNMR), while that of impersonation attacks using impostor attack presentation match rate (IAPMR). Additionally, to establish a relationship between the attack efficacy and biometric recognition performance, the relative impostor attack presentation accept rate (RIAPAR), is reported using the formula \[1 + (\text{IAPMR} - (1 - \text{FNMR}))\] as originally proposed by Scherhag et al. [28].

#### B. Results

Figure 4 shows boxplots of the obtained comparison scores, while table III presents the corresponding descriptive statistics of the comparison scores. The figures also visualise decision thresholds for several security levels based on a fixed FMR value between 0.001\% and 1.0\%.

### TABLE III: Descriptive statistics of the comparison scores

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcFace</td>
<td>Non-mated</td>
<td>0.171</td>
<td>0.042</td>
<td>0.359</td>
<td>0.463</td>
<td>0.000</td>
<td>0.638</td>
</tr>
<tr>
<td></td>
<td>Mated</td>
<td>0.667</td>
<td>0.028</td>
<td>-2.987</td>
<td>11.702</td>
<td>0.407</td>
<td>1.000</td>
</tr>
<tr>
<td>Bona fide</td>
<td>Non-mated</td>
<td>0.585</td>
<td>0.085</td>
<td>-0.307</td>
<td>1.873</td>
<td>0.249</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Mated</td>
<td>0.303</td>
<td>0.078</td>
<td>0.909</td>
<td>2.065</td>
<td>0.127</td>
<td>0.573</td>
</tr>
<tr>
<td>Concealment</td>
<td>Non-mated</td>
<td>0.364</td>
<td>0.080</td>
<td>-0.224</td>
<td>0.116</td>
<td>0.150</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>Mated</td>
<td>0.214</td>
<td>0.026</td>
<td>1.057</td>
<td>9.731</td>
<td>0.000</td>
<td>0.933</td>
</tr>
<tr>
<td>Impersonation</td>
<td>Non-mated</td>
<td>0.059</td>
<td>0.062</td>
<td>2.518</td>
<td>9.731</td>
<td>0.000</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>Mated</td>
<td>0.090</td>
<td>0.086</td>
<td>-1.266</td>
<td>9.731</td>
<td>0.000</td>
<td>0.933</td>
</tr>
</tbody>
</table>

Based on the figures and the table above, several interesting observations can be made:

**Bona fide makeup** This type of makeup certainly leads to a degradation and wider spread of mated comparison scores. Nevertheless, given a fixed FMR value of 0.1\%, which is recommended as a security level by FRONTEX in several operationally relevant scenarios [29], the vast majority of verification transactions would be successful.
This type of makeup may, however, lead to problems when much more stringent security levels are applied, cf. CAPNR at FMR value of 0.001%.

**Concealment makeup** In this scenario, the applied makeup vastly degrades the mated comparison scores, although the resulting distributions do not fully overlap with the non-mated score distributions, *i.e.* the concealment is not perfect. This notwithstanding, the attacks would prove successful in a large proportion of verification cases for all but the most strict security level depicted in the figures.

**Impersonation makeup** This type of makeup produces score distributions which lie between the mated and non-mated ones. In other words, for the most part, the artists manage to alter their appearance such that it mimics that of the target subject at least in some ways. A large proportion of the attacks would prove successful, even at the most stringent of the depicted security levels. It should be noted, that the initial appearance of the artists and possible similarity to the target does not seem to play a major role – the score distributions of comparisons between the artist without makeup and the target did not exhibit deviations from the non-mated score distributions.

Furthermore, some artists actually manage successful impersonation despite a large age difference between the artist and the target or even when the artist and target are not of the same sex. This suggests that the skill of the artist is a much more important factor than their initial appearance.

Table IV shows the biometric recognition and presentation attack error rates for the three types of makeup at the aforementioned security (fixed FMR) levels. It can be seen that the tested systems exhibit a near-optimal biometric performance for normal samples – a very low FNMR even at the most strict of the reported security levels. However, when quantifying the observations made on the score distributions above, significant error rates expressing the vulnerability of the face recognition systems are observed for samples with makeup. The bona fide makeup causes only small to moderately high CAPNMR values, which suggests that current face recognition systems are strongly capable of dealing with data subjects wearing daily makeup. The results are much more dramatic for the concealment and impersonation attacks. Concealment attacks reach a CAPNMR of around 40% and 70% for the open source and COTS systems, respectively.
already at the lowest reported security level (fixed FMR of 1%); this quickly increases to beyond 70 and even 90% for higher security levels (i.e., lower FMR values). Alarmingly, a decent proportion (10-20%) of the impersonation attacks are successful even at the most stringent of the tested security levels. For less strict security levels, this proportion increases to well above 50% IAPMR: around 70% and 90% at fixed FMRs of 0.1% and 1%, respectively.

Figure 5 shows example image pairs of successful and failed attacks. The subfigure (b) shows the remarkable strength of facial recognition systems – despite extremely strong makeup covering the whole face, a seemingly different shape of the upper lip, as well as a changed hair colour and hair style, both tested systems correctly recognise the depicted data subject. The successful M-PA in subfigure (c) is particularly impressive, as the makeup artist and the attack target are actually of a different sex. This emphasises that sufficiently skilled attackers can successfully impersonate targets to which they bear no immediate resemblance. This anecdotal observation is confirmed by the quantitative evaluation shown in figure 6, which shows that the initial resemblance of the artist and the target does not seem to be a prerequisite for a successful impersonation attack.

The results obtained in the vulnerability analysis of the collected dataset indicate that, provided a sufficiently skilled artist, makeup presentation attacks pose a much more severe threat to biometric face recognition systems than previously reported in the literature and summarised in section II.
attacks with the purpose of impersonation. While the IAPRs protocols and metrics.

source and commercial biometric facial recognition systems concealment or impersonation. Using the collected dataset, presentation attacks utilising makeup for the purposes of identity verification systems.

conducted benchmark shows that the collected high-quality biometric anti-spoofing: Presentation attack detection,” 2019.

The obtained results are especially striking for the makeup attacks with the purpose of impersonation. While the IAPRs achieved on previously available datasets (e.g. MIFS and DFW) indicated impersonation M-PAs to only be a minor or moderate risk for current facial recognition systems [5], the conducted benchmark shows that the collected high-quality impersonation M-PAs exhibit an very high efficacy (see table IV) for practically relevant security levels in operational biometric verification systems.

V. CONCLUSION

This paper presents a new dataset of high-quality presentation attacks utilising makeup for the purposes of identity concealment or impersonation. Using the collected dataset, a biometric performance benchmark of state-of-the-art open-source and commercial biometric facial recognition systems has been conducted with ISO/IEC standardised evaluation protocols and metrics.

The obtained results are especially striking for the makeup attacks with the purpose of impersonation. While the IAPRs achieved on previously available datasets (e.g. MIFS and DFW) indicated impersonation M-PAs to only be a minor or moderate risk for current facial recognition systems [5], the conducted benchmark shows that the collected high-quality impersonation M-PAs exhibit an very high efficacy (see table IV) for practically relevant security levels in operational biometric verification systems.

To facilitate reproducible research and future experiments in this field, the dataset is made available upon request.

REFERENCES


