Spoofing Faces Using Makeup: An Investigative Study
Cunjian Chen, Antitza Dantcheva, Thomas Swearingen, Arun Ross

HAL Id: hal-01430020
https://hal.archives-ouvertes.fr/hal-01430020
Submitted on 9 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

Makeup can be used to alter the facial appearance of a person. Previous studies have established the potential of using makeup to obfuscate the identity of an individual with respect to an automated face matcher. In this work, we analyze the potential of using makeup for spoofing an identity, where an individual attempts to impersonate another person’s facial appearance. In this regard, we first assemble a set of face images downloaded from the internet where individuals use facial cosmetics to impersonate celebrities. We next determine the impact of this alteration on two different face matchers. Experiments suggest that automated face matchers are vulnerable to makeup-induced spoofing and that the success of spoofing is impacted by the appearance of the impersonator’s face and the target face being spoofed. Further, an identification experiment is conducted to show that the spoofed faces are successfully matched at better ranks after the application of makeup. To the best of our knowledge, this is the first work that systematically studies the impact of makeup-induced face spoofing on automated face recognition.

1. Introduction

Biometrics refers to the automated recognition of individuals based on their biological traits such as face, fingerprints and iris. A typical biometric system acquires the biometric data of an individual using a sensor; extracts a set of salient features from the data; and uses these features to determine or verify the identity of an individual [11]. In spite of its advantages, a biometric system is vulnerable to spoofing, where an adversary can spoof the biometric trait of another individual in order to circumvent the system [16, 19, 2, 25]. Unlike obfuscation, which entails deliberately obscuring one’s own identity, spoofing entails taking on another person’s identity, with the purpose of accessing privileges and resources associated with that person [17, 21]. Spoofing, in the context of face recognition,
to systematically study this effect.

1.1. Background and related work

Recent work has demonstrated the impact of commonly used facial makeup on automated face recognition systems [6, 8, 9, 12, 5], and automated face-based gender and age estimation systems [4]. Makeup can be used to alter the perceived (a) facial shape; (b) nose, mouth, eye and eye brow shape; (c) nose, mouth, eye and eye brow size; (d) facial contrast and (e) facial skin quality and color. It can also be used to conceal wrinkles, dark shadows and circles underneath the eyes, and camouflage birth moles, scars and tattoos [18].

Considering the widespread use of makeup and its implications in altering facial appearance (e.g., facial aesthetics [7]), in this work we focus on the use of makeup for spoofing. Unlike previous work [6, 8, 9], where commonly used makeup was observed to affect face recognition systems by obscuring a person’s identity, here we consider the scenario where makeup is used by an individual to mimic the facial appearance of another individual (see Figure 2).

Current face spoof detection schemes either rely on physiological cues such as eye blinking, mouth movements, and macro- and micro-expression changes [19, 16], or textural attributes of the face image [16, 24, 23, 14]. But none of these methods represent a viable mechanism for detecting makeup induced spoofing (especially since makeup is widely used). Also, in contrast to other face alteration techniques such as plastic surgery, makeup is non-permanent and cost efficient. This makes makeup-based spoofing a realistic threat to the integrity of a face recognition system.

The rest of the paper is organized as follows. Section 2 discusses the assembled spoofing dataset. Section 3 introduces the two spoofing indices for quantifying the effect of makeup on automated face recognition. Section 4 discusses face recognition methods used in this study to evaluate the impact of makeup induced spoofing. Section 5 presents the related experiments. Results of the experiments are discussed in Section 6, followed by a summary of the paper in Section 7.

2. Makeup Induced Face Spoofing (MIFS) Dataset

In order to investigate the problem of makeup induced face spoofing, we first assemble a dataset consisting of 107 makeup-transformations taken from random YouTube makeup video tutorials. We refer to this dataset as the Makeup Induced Face Spoofing (MIFS) dataset. There are two before-make-up and two after-make-up images per subject. Since each subject is attempting to spoof a target identity, we also have two face images of the target identity from the Web. Thus, this dataset has three sets of face images: images of a subject before makeup; images of the same subject after makeup with the intention of spoofing; and images of the target subject who is being spoofed. However, it is important to note that the target images are not necessarily those used by the spoofer as a reference during the makeup transformation process. This is important to point out because the spoofed celebrities can often change their facial appearance and this will have an effect on the match score between the after-make-up image of the impersonator and the target image of the celebrity. When we searched the Web for face images of the target identity, we tried to select images that most resembled the after-make-up image.

All the acquired images are subjected to face cropping. This routine eliminates hair and accessories [6]. Examples of cropped images, based on a Commercial Off-The-Shelf (COTS) face detector, are shown in Figure 3.

We make the following observations about this dataset: (a) the subjects in the before and after makeup sets do not overlap with subjects in the target set; (b) there are duplicate identities of subjects attempting to spoof — this is because there are subjects attempting to spoof different target identities (see Figure 2); (c) there are duplicate identities in the target set — this is because there are multiple subjects attempting to spoof the same target identity (see Figure 4); (d) the images in the dataset include variations in expression, illumination, pose, resolution and quality.

**Makeup-transformation:** In the MIFS dataset, subjects use different types of makeup to alter their appearance and resemble a target identity. While the makeup application process varies across the dataset (depending upon the sub-

---

1 Note that the makeup video tutorials do not include images of the target identity, if any, used by the subject during the spoofing process. In fact, it is likely that some subjects are attempting to resemble the target identity from memory.

---

**Figure 2.** The subject on the top attempts to resemble identities in the bottom row (labeled “Target Subjects”) through the use of makeup. The result of these attempts can be seen in the second row. Images were obtained from the WWW.
Figure 3. Examples of images in the MIFS dataset after cropping. Here, (a), (b) and (c) represent three spoofing attempts. In each case, the image on the left shows the subject before makeup, the one in the middle is the subject after makeup, and the image on the right is the target identity that the subject is attempting to impersonate (see text for explanation).

For each subject $p$, we have the following image samples: $\{B_p^1, B_p^2\} \in B_P$, $\{A_p^1, A_p^2\} \in A_P$, and $\{T_p^1, T_p^2\} \in T_P$. Let $\phi(x, y)$ denote the similarity match score between two images $x$ and $y$ as computed by a matcher (the greater the value, the higher the similarity between two faces). The similarity scores are normalized in the $[0, 1]$ interval.

A spoofing attack can be deemed to be successful for subject $p$, when the similarity score between the after-make-up images, $A_p^i$, and the target images, $T_p^j$, (i.e., $\phi(A_p^i, T_p^j)$) increases. However, it is not easy to make this assessment, since any change in match score has to be viewed with respect to the entire score distribution of the matcher (and not just the absolute change in value). Hence, we consider two spoofing indices.

The two spoofing indices that we introduce below describe the similarity score $\phi(A_p^i, T_p^j)$ with respect to two types of genuine scores: (a) reference genuine scores $\phi(T_p^1, T_p^2)$, where the similarity between two samples of
the same target identity is computed (spoofing index 1, \(SI_1\)), and (b) reference genuine scores \(\phi(A_i^P, A_j^P)\), where the similarity between two samples of the after-makeup images of a subject is computed (spoofing index 2, \(SI_2\)). The two spoofing indices are described below.

**Spoofing Index 1**: \(SI_1\) is defined as follows:
\[
SI_1 = 1 - \min_{i,j} |\phi(A_i^P, T_j^P) - \phi(T_1^P, T_2^P)|,
\]
where \(i, j \in \{1, 2\}\). Here, we examine if the similarity score between the after-makeup image and the target image is within the range of the score between two samples of the target identity. Specifically, \(\phi(A_i^P, T_j^P) \approx \phi(T_1^P, T_2^P)\) suggests that spoofing is successful and the output of \(SI_1 \approx 1\). For the case \(\phi(A_i^P, T_j^P) < < \phi(T_1^P, T_2^P)\), spoofing is not successful and the output of \(SI_1 \approx 0\).

**Spoofing Index 2**: \(SI_2\) is based on the difference in similarity score between the after-makeup and target images with respect to the after-makeup images:
\[
SI_2 = 1 - \min_{i,j} |\phi(A_i^P, T_j^P) - \phi(A_i^P, A_j^P)|.
\]
This index is very similar to \(SI_1\). However, here, we examine if the similarity score between the after-makeup image and the target image is within the range of the score between two after-makeup samples of the same person. Specifically, \(\phi(A_i^P, T_j^P) \approx \phi(A_i^P, A_j^P)\) suggests that spoofing is successful and the output of \(SI_2 \approx 1\). For the case \(\phi(A_i^P, T_j^P) < < \phi(A_i^P, A_j^P)\), spoofing is not successful and the output of \(SI_2 \approx 0\).

4. **Automated Face Recognition**

The following two face matchers were used in this study: a Commercial Off-The-Shelf (COTS) face software and the VGG face matcher.

The VGG face descriptor [20] is computed using a Convolutional Neural Network (CNN) implementation based on the VGG-Verdy-Deep-16 architecture as characterized in [20]. It is developed based on a triplet-loss training scheme to learn a face embedding that has a similar principle as metric learning.

The effectiveness of these two face matchers was verified using the BLUFR protocol [15] on the LFW dataset\(^3\) [10] (as seen in Figure 5). Compared to the original LFW protocol, the BLUFR protocol contains both verification and open-set identification scenarios to fully exploit the potential of the LFW dataset, with focus on low FARs. The results are reported using the “Verification \(\mu - \sigma\) Rate” as described in the BLUFR protocol. Table 1 shows the Verification Rate at 0.1% FAR. Both matchers exceed the baseline performance,\(^4\) with the COTS matcher surpassing the baseline by a factor of nearly two. Thus, both matchers are good candidates for use in this work.

5. **Experiments**

In this section, we conduct experiments on the MIFS dataset. By performing these experiments, we seek to understand a) whether makeup can be used to spoof faces from the perspective of an automated face matcher, and b) the extent to which makeup has the ability to spoof a face recognition system. We reiterate that the subjects represented in the MIFS dataset were not deliberately trying to spoof an automated face recognition system.

5.1. **Match Score Analysis**

To study the possibility of spoofing, we first analyze the score changes after the application of makeup. For spoofing to be successful, the similarity score between the after-makeup and target images should increase compared to that of the before-makeup and target images. Match scores were generated according to the following protocol:

1. Match \(B_i^P\) against \(T_j^P\) \((B-T)\): the image before makeup is matched against the target image.

![Figure 5. ROC curves averaged over 10 folds of the BLUFR protocol. The “Verification \(\mu - \sigma\) Rate” reports the average performance across the 10 folds and subtracts the standard deviation of that performance to yield a lower bound on performance.](image)

Table 1. Performance of the COTS and VGG face matchers when using the BLUFR protocol on the LFW dataset. The mean of the Verification Rate over the 10 folds is given by \(\mu\) and the standard deviation is given by \(\sigma\).

<table>
<thead>
<tr>
<th>Matcher</th>
<th>FAR (%)</th>
<th>Verification Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COTS</td>
<td>0.1</td>
<td>81.0</td>
</tr>
<tr>
<td>VGG Face</td>
<td>0.1</td>
<td>56.6</td>
</tr>
</tbody>
</table>

\(^{3}\)COTS template extraction failed for 113 images of the 12,896 images in the BLUFR protocol. For a fair comparison with VGG, we withheld these images from the VGG matcher as well. Note that the BLUFR protocol only uses 12,896 images of the 13,233 images in LFW.

\(^{4}\)http://www.cbsr.ia.ac.cn/users/scliao/projects/blufr
Figure 6. Boxplots displaying similarity score distributions of $B-T$, $A-T$ and $B-A$, as computed by the two face matchers. For each boxplot, the central red line is the median value and the edges of the box correspond to the 25th and 75th percentiles.

(a) COTS Matcher

(b) VGG Face Matcher

Figure 7. Observing the score change between $B-T$ and $A-T$ ($\phi(A,T) - \phi(B,T)$) via histograms for the two face matchers.

(a) COTS Matcher

(b) VGG Face Matcher

2. Match $A^p_i$ against $T^p_j$ ($A-T$): the image after makeup is matched against the target image.

3. Match $B^p_i$ against $A^p_j$ ($B-A$): the image before makeup is matched against the image after makeup.

Figure 6 illustrates the boxplots of score distributions corresponding to $B-T$, $A-T$ and $B-A$. We note an evident upward shift in $A-T$ scores when makeup is applied (compared to $B-T$). This means the application of makeup has increased the chances of a subject being matched against the target image, as reflected by the change in similarity score. Figure 7 visualizes the differences between $A-T$ and $B-T$ as a histogram. As can be seen, the majority of the differences are positive.

5.2. Spoofing Indices

Next, we observe the values of the proposed spoofing indices. As shown in Figure 8, both spoofing indices demonstrate the extent to which makeup has the ability to spoof a particular face recognition system. The higher the value, the more likely that spoofing has occurred. Both face matchers are observed to be vulnerable to spoof attacks via makeup.

5.3. Histogram Shift

In addition, we visualize the histograms of match scores that would allow us to see the improvement in makeup induced genuine scores in the context of general genuine and impostor score distributions of the matcher. More specifically, we plot the score distributions of $B-T$, $A-T$, and the LFW dataset in Figure 9. The LFW scores are derived from 13,120 images in the LFW dataset [10].

In Figure 9(a), we see that the $B-T$ scores fall in the $[0.03, 0.61]$ range and the $A-T$ scores fall in the $[0.13, 0.77]$ range. This indicates that the match scores generally increase after the application of makeup and the histogram shifts to the right. The same phenomenon can be seen in Figure 9(b), but to a lesser extent.

5.4. Identification Experiment

In this setting, the after-makeup and before-makeup images are used as probes, and the target images to be spoofed are placed in the gallery. The goal is to determine whether

5

113 of the 13,233 LFW images failed the template extraction stage of the COTS matcher. In order for a fair comparison to the Oxford VGG matcher, we withheld these 113 images from the Oxford VGG matcher as well.
target images match at a better rank with after-makeup images than with before-makeup images. We also populate the gallery with 13,120 “background” images from the LFW dataset [10]. We match the probe images against the gallery images and compute the ranks at which the target spoofs are successfully matched. In this context, rank-$k$ denotes the probability that the target identity to be spoofed occurs among the top $k$ matches for a given probe image. The results are summarized in Figure 10, from which we observe that the identification accuracy associated with after-makeup images is significantly higher than that of the before-makeup images. Figure 11 shows examples of subjects being matched at better and worse ranks after the application of makeup. This clearly illustrates that makeup is a viable spoofing method in the context of automated face recognition.

6. Observations and Future Work

In this section we summarize the main findings of this research.

- Based on our experiments with two different face matchers, we observe that the similarity scores between after-makeup and target face images are higher than those between before-makeup and target-images. This indicates that makeup can be strategically used to spoof a face.

- The identification experiment shows that the target identity being spoofed is likely to match at a better rank after the application of makeup by the impostor.

- Spoofing was successful in some cases and not in others. This is expected as the success of spoofing depends on the source face and target face as well as the makeup procedure used.

Even though subjects in the MIFS dataset were not deliberately launching a spoof attack against an automated face matcher, the results presented here demonstrate the potential of using makeup to spoof a face recognition system (see Figure 11). These observations point out the necessity for developing face recognition methods that are less impacted by the application of makeup. There are several ways to potentially address this issue:
Figure 10. Comparison of target identification results before and after the application of makeup. The gallery set consists of the target images to be spoofed as well as 13,120 images from the LFW dataset. The probe set consists of the before-makeup and after-makeup images. The application of makeup for spoofing purposes significantly increases the chance of target identities being matched at higher (i.e., better) ranks.

Figure 11. Examples of spoofing attempts with respect to the COTS face matcher. The target images are successfully retrieved at better ranks when using after-makeup images as probes in (a) – (d) (green border) and at worse ranks in (e) and (f) (red border). Note that (c) and (d) contain the same subject but different target identities.

- As indicated in the work of Short et al. [22], polarimetric thermal images are minimally impacted by the application of cosmetic paints. Specifically, their work demonstrated the efficacy of recognizing faces in the presence of makeup by using polarimetric thermal imaging.

- Extracting features that are invariant to the application of makeup can help mitigate the presented type of attack [5]. In addition, makeup detection schemes [3] can be employed for detecting and preprocessing face images prior to face matching.
7. Summary

In this work, we presented preliminary results on makeup-induced face spoofing. We observed in Section 5.1 that similarity scores between a face image and the target face to be spoofed did indeed increase after the application of makeup. In Sections 5.2 and 5.3, we noted that the increase in match score is significant. Finally, in Section 5.4, we showed that a face image is matched with the target identity at a better rank after applying makeup to the for-
tion of makeup. In Sections 5.2 and 5.3, we noted that the
increase in match score is significant. Finally, in Section
5.4, we showed that a face image is matched with the target
issue of makeup has to be accounted for in the context of
authentication applications, this research suggests that the
mer. With the increasing use of face recognition systems in
identity at a better rank after applying makeup to the for-
tion of makeup. In Sections 5.2 and 5.3, we noted that the
increase in match score is significant. Finally, in Section
5.4, we showed that a face image is matched with the target

References

face materials on polarimetric-thermal measurements: appli-