

Can Facial Cosmetics Affect the Matching Accuracy of Face Recognition Systems?*

Antitza Dantcheva, Cunjian Chen, Arun Ross

Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown WV 26506, USA

andantcheva@mail.wvu.edu, cchen10@mix.wvu.edu, arun.ross@mail.wvu.edu

Abstract

The matching performance of automated face recognition has significantly improved over the past decade. At the same time several challenges remain that significantly affect the deployment of such systems in security applications. In this work, we study the impact of a commonly used face altering technique that has received limited attention in the biometric literature, viz., non-permanent facial makeup. Towards understanding its impact, we first assemble two databases containing face images of subjects, before and after applying makeup. We present experimental results on both databases that reveal the effect of makeup on automated face recognition and suggest that this simple alteration can indeed compromise the accuracy of a biometric system. While these are early results, our findings clearly indicate the need for a better understanding of this face altering scheme and the importance of designing algorithms that can successfully overcome the obstacle imposed by the application of facial makeup.

1. Introduction

Motivated by the need to deploy highly reliable face recognition systems in security applications, we seek to initiate research that studies the impact of facial makeup on face recognition. Facial makeup is an example of a cosmetic alteration that can change the perceived appearance of the face. Other alterations include aging (a natural biological change) and plastic surgery (a medically induced change). Recent work has focused on the impact of plastic surgery on face recognition [1][14][13][3]. However, such surgical alterations are generally costly and permanent. On the other hand, non-permanent cosmetic alterations, such as makeup, tend to be simple, cost efficient and socially acceptable; at the same time, they have the potential to sub-

stantially change appearance. Specifically, such alterations can (a) alter the perceived facial shape by accentuating contouring techniques; (b) alter the perceived nose shape and size by contouring techniques; (c) enhance or reduce the perceived size of the mouth; (d) alter the appearance and contrast of the mouth by adding color; (e) alter the perceived form, color and location of eyebrows; (f) alter the perceived shape, size and contrast of the eyes; (g) conceal dark circles underneath the eyes; and (h) alter the perceived skin quality and color. In addition to the aforementioned effects, cosmetics can also be used to successfully camouflage as well as affect wrinkles, birth moles, scars and tattoos. A vast cosmetics market¹ - typically targeted towards women - attempts to improve facial aesthetics while projecting good health. The beautification effects induced by cosmetics have been studied in recent research literature [6][4].

The impact of makeup on *human* ability to recognize faces has been studied by Ueda and Koyama [16]. The authors concluded that light makeup slightly increases human recognizability, whereas heavy makeup significantly decreases it. While the accentuation of distinctive characteristics using light makeup is helpful towards recognition, heavy makeup increases bilateral size and symmetry of the eyes and lips leading to a decreased characteristic distinctiveness of faces [11].

In spite of the aforementioned observations, there is no research that establishes the impact of cosmetic makeup on *automated* face recognition systems. While extensive research has been done to quantify the effect of pose, illumination and expression (commonly referred to as PIE) on face matching [7] [10], the biometric literature is largely silent about the effects of cosmetic makeup on face recognition. This is particularly of interest given that cosmetic makeup is commonly used by females in many parts of the world¹. To address this aspect, we consider the following question: Can cosmetic facial makeup affect the matching accuracy of automated face recognition schemes? In or-

*This work was funded by the Center for Identification Technology Research (CITeR).

¹<http://www.japaninc.com/article.php?articleID=1390>

der to answer this question, we first assemble two different face databases consisting of female subjects with and without makeup. Subsequently we test the matching accuracy of multiple face recognition algorithms (both academic and commercial) on these two databases. Experimental results suggest that face recognition can be substantially impacted by the application of facial makeup. To the best of our knowledge, this is the first work in the biometric literature to explicitly demonstrate this effect.

The rest of the paper is organized as follows. Section 2 discusses the databases that were assembled for this study. Section 3 describes the face recognition techniques that were used to assess the matching accuracy. Section 4 presents the experimental results. Section 5 discusses the key outcomes of this study. Section 6 addresses the challenge and provides a preliminary solution towards mitigating the effect of makeup on automated face recognition. Section 7 concludes the paper.

2. Facial Makeup Databases

A review of the face recognition literature suggests that most databases available to biometric researchers are suitable for studying the effect of pose, illumination, expression, aging, plastic surgery, occlusion, motion blur, etc. on face recognition. However, there are no publicly available face databases that can be directly used to perform the study proposed in this work. The paper by Ueda and Koyama [16] discusses a proprietary database consisting of Japanese females. The work by Scherbaum et al. [12] specifies a private database containing 56 subjects, each with and without makeup. However, these databases could not be obtained from the respective researchers.

Hence, in order to conduct this study, we assembled two databases. The first database, which we refer to as the YouTube MakeUp (YMU) database, consists of face images of subjects gleaned from YouTube video makeup tutorials. The second database is a synthetic database, which we refer to as the Virtual MakeUp (VMU) database, where face images of Caucasian female subjects in the FRGC repository² were synthetically modified to simulate the application of makeup. A publicly available software³ was used to perform this alteration. A brief description of these two databases is presented below.

2.1. YouTube Makeup (YMU) database

We assembled a dataset consisting of 99 subjects, specifically Caucasian females, from YouTube makeup tutorials. We collected images of the subjects before and after the application of makeup. There are four shots per subject: two shots before the application of makeup and two shots after

the application of makeup. For a few subjects, we were able to obtain three shots each before and after the application of makeup. We use the notation \mathcal{N} for the before makeup shots and \mathcal{M} for the after makeup shots.

The makeup in these face images varies from subtle to heavy as can be seen in Figure 1. The cosmetic alteration is mainly in the ocular area, where the eyes have been accentuated by diverse eye makeup products. Additional changes are on the quality of the skin due to the application of foundation and change in lip color. This database includes some variations in expression and pose. The illumination condition is reasonably constant over multiple shots of the same subject. In few cases, the hair style before and after makeup changes drastically. URLs of videos from which face images were taken have been listed at <http://www.antitza.com>.

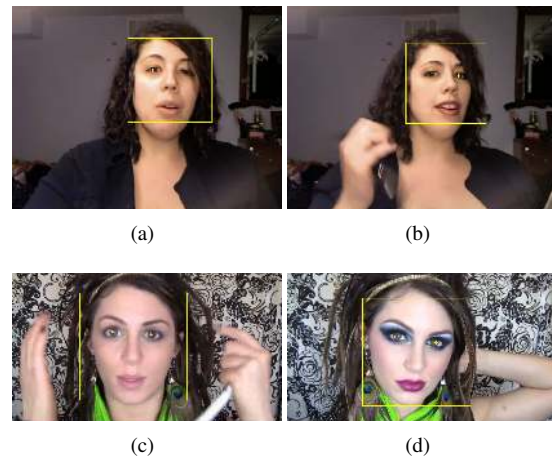


Figure 1. Examples of two subjects in the YMU database assembled by the authors, each without [(a) and (c)] and with makeup [(b) and (d)]. The subject on the top has subtle makeup, while the one on the bottom has heavy makeup. The output of the automated face and eye detection scheme, see [5], is indicated in yellow.

2.2. Virtual Makeup (VMU) database

The VMU database was assembled by synthetically adding makeup to 51 female Caucasian subjects in the FRGC database. We added makeup by using a publicly available tool from Taaz. We created three virtual makeovers: (a) application of lipstick only; (b) application of eye makeup only; and (c) application of a full makeup consisting of lipstick, foundation, blush and eye makeup. Hence, the assembled dataset contains four images per subject: one before-makeup shot⁴ and three after-makeup shots. Figure 2 provides an example. We use the notation \mathcal{N} for no-makeup, \mathcal{L} for lipstick, \mathcal{E} for eye makeup and \mathcal{F} for full makeup.

²<http://www.nist.gov/itl/iad/ig/frgc.cfm>

³<http://www.taaz.com>

⁴This image is referred to as the no-makeup shot

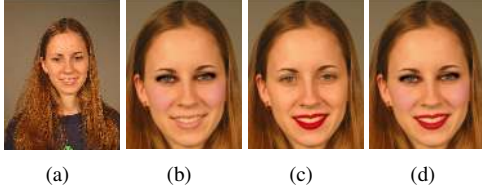


Figure 2. Example of a subject in the VMU database assembled by the authors. (a) depicts the subject without makeup, whereas, (b), (c) and (d) constitute the makeup shots. The makeup shots include the synthetic addition of (b) eye makeup (c) lipstick, and (d) full makeup, respectively, using the popular Taaz software.

The first database provides real life images of subjects with and without makeup, and helps assess the degradation in performance due to makeup. The second database contains images from the FRGC dataset that are synthetically processed by applying makeup on the portrayed subjects. *The second database ensures that intra-class variations due to pose, illumination and expression are minimized.* This allows us to focus solely on the analysis of the impact of makeup and also study of the impact of different kinds of makeup.

In the following section, we describe the face recognition methods that were used to evaluate the impact of facial makeup on face recognition systems.

3. Face Recognition Techniques

This section briefly introduces the three face recognition algorithms (i.e., matchers) that we employ in our study, viz., Gabor wavelets [8], Local Binary Pattern (LBP) [15] and the commercial Verilook Face Toolkit⁵. The choice of these matchers were based on the following observations:

- Gabor features encode both shape and texture information across different scales and orientations.
- LBP captures micro-patterns and thus represents small-scale appearance.
- Verilook is a commercial face recognition system that has demonstrated competitive performance in several public face databases.

Descriptor-based face recognition methods, such as LBP, are computational and time efficient, and do not require a learning step for feature extraction and analysis. Thus, given a pair of face images, these methods can generate a matching score without requiring an explicit training phase. Hence, they were selected for the experiments conducted in this work.

Subspace based learning methods such as principal component analysis (PCA) and linear discriminant analysis

(LDA) require the generation of a projection matrix. This projection matrix has to be computed from a set of training images. The result of the matching operation is, therefore, closely related to the choice of the training images used to generate the subspace. A different choice of training images can generate a different subspace, therefore biasing the matching results. To avoid this concern, we only consider descriptor-based schemes in our preliminary analysis.

Prior to invoking the three face matchers, individual images were converted from RGB to grayscale. Then, an automated face and eye detection routine was utilized to localize the spatial extent of the face in individual images. The face images were geometrically normalized based on the location of the eyes. Furthermore, we used Difference of Gaussian (DoG) to preprocess the images and minimize variations due to illumination. Therefore, all three face matchers operate on grayscale images of the face (see Figure 3). The spatial resolution of each face image was 130×150 .

3.1. Gabor Wavelets

Gabor wavelets are defined as follows [8]:

$$\varphi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|}{\sigma^2} e^{-\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,v}z} - e^{-\frac{\sigma^2}{2}} \right], \quad (1)$$

where μ and v denote the orientation and scale of the Gabor kernels, z denotes the pixel position, i.e., $z = (x, y)$, and $\|\cdot\|$ denotes the norm operator [17]. The wave vector $k_{\mu,v}$ is given by:

$$k_{\mu,v} = k_v e^{i\phi_\mu}, \quad (2)$$

where $k_v = k_{max}/f^v$ and $\phi_\mu = \pi\mu/8$. Here, k_{max} is the maximum frequency and f is the spacing factor between kernels in the frequency domain. The subtraction of the component $e^{-\sigma^2/2}$ imparts robustness to variations in illumination. The parameters μ and v are chosen as 8 orientations and 5 scales, respectively, resulting in a total of 40 Gabor kernels. The size of Gabor kernels is determined by the parameter z . The Gabor wavelet representation of an image is obtained by convolving it with these Gabor kernels:

$$G_{u,v}(z) = I(z) * \varphi_{u,v}(z). \quad (3)$$

The complex Gabor response has two parts, the real part $\Re_{u,v}(z)$ and the imaginary part $\Im_{u,v}(z)$. Such a convolution operation can be performed in the frequency domain via Fast Fourier Transform (FFT). The magnitude of the Gabor response (see an example in Figure 3 (c)) is computed as $A_{u,v}(z) = \sqrt{\Re_{u,v}(z)^2 + \Im_{u,v}(z)^2}$.

The Gabor features can be extracted by downsampling the output of each Gabor magnitude image, normalizing them to zero mean and unit variance, and concatenating them [15]. A simple L_2 norm distance measure (ϕ) is used

⁵<http://www.neurotechnology.com/verilook.html>

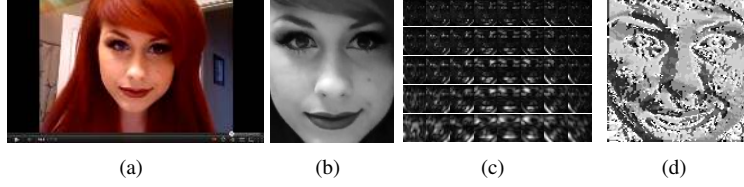


Figure 3. (a) Example makeup (\mathcal{M}) image from the YMU database. (b) Preprocessed image. (c) Gabor magnitude response (40 images). (d) LBP coded image. We note that the Gabor response around the mouth and ocular regions is more prominent than in other parts of the face.

to compare the distance between Gabor features.

$$\phi(G^1, G^2) = \sqrt{\sum_{i=1}^L (g_i^1 - g_i^2)^2}, \quad (4)$$

where G^1 and G^2 are the Gabor features extracted from two images, and L is the length of the feature vector. In our experiments, $L = 10,240$.

3.2. Local Binary Pattern (LBP)

The LBP descriptor [2] was first proposed as a type of texture descriptor that characterizes micro-patterns or micro-structures in an image by binarizing a 3×3 neighborhood based on the differences in pixel intensity between the center pixel and neighborhood pixels, and converting the resulting binary string into a decimal value. In order to better represent large-scale structures in images, the small-scale patterns were extended to accommodate neighborhoods of different sizes. LBP has many desired properties such as tolerance to monotonic illumination changes and computational efficiency.

The binary pattern for pixels evenly lying in a neighborhood $(x_i, i = 0, 1, \dots, P-1)$ with respect to the center pixel x_c , is computed as follows:

$$f(x_i - x_c) = \begin{cases} 1 & \text{if } x_i - x_c \geq \tau; \\ 0 & \text{if } x_i - x_c < \tau. \end{cases} \quad (5)$$

Next, a binomial weight 2^i is assigned to each value of $f(x_i - x_c)$ in order to generate the LBP code:

$$LBP(x_c) = \sum_{i=0}^{P-1} f(x_i - x_c) 2^i, \quad (6)$$

where P is the total number of pixels in this neighborhood or region. As an example, for a 3×3 neighborhood, the value of P will be 8. The threshold value τ is set to be zero for LBP. This results in an 8-bit code for each encoded pixel. The LBP operator is often denoted as $LBP_{P,R}$, where P refers to the number of sampling points and R is the radius of the neighboring region. To generate LBP features, each LBP coded image (see an example in Figure 3 (d)) is divided

into sub regions, where histogram information is extracted from each sub region and then concatenated to form the descriptor. The histogram intersection similarity measure (φ) that is used to compare two LBP descriptors is defined as follows [17],

$$\varphi(H^1, H^2) = \sum_{i=1}^L \min(h_i^1, h_i^2), \quad (7)$$

where H^1 and H^2 are the LBP features from images I^1 and I^2 , respectively. L is the number of histogram bins. In our experiments, $L = 15,104$. The similarity measure is then converted to a distance measure.

3.3. Neurotechnology Verilook

Verilook⁶ is a commercial face detection and recognition software, which computes matching scores between images, without a learning phase. While the underlying algorithm is not publicly disclosed, it is known that several face recognition algorithms have been combined in this software.

Having introduced the databases and the face matchers, we now proceed to describe the experimental component of the work.

4. Experiments

The purpose of the experiments was to determine the impact of facial makeup on the performance of automated face recognition algorithms. To do so, we compared the ability of such algorithms to recognize people before and after applying makeup. In this regard, we used the YMU database to understand the degradation in performance on real makeup images and the VMU database to determine which type of facial cosmetic alteration has the most impact on matching accuracy.

4.1. Experiments on the YMU Database

The YMU database contains 99 subjects. There are four images associated with each subject: two images without

⁶<http://www.neurotechnology.com/verilook.html>

makeup and two images with makeup. Let \mathcal{N}_1 and \mathcal{N}_2 denote the images without makeup, and \mathcal{M}_1 and \mathcal{M}_2 denote the images with makeup. Genuine and impostor scores for each of the three face matchers were generated according to the following protocol:

1. Matching \mathcal{N}_1 against \mathcal{N}_2 : Both the images to be compared do not have makeup (the before-makeup images).
2. Matching \mathcal{M}_1 against \mathcal{M}_2 : Both the images to be compared have makeup (the after-makeup images).
3. Matching \mathcal{N}_1 against \mathcal{M}_1 , \mathcal{N}_1 against \mathcal{M}_2 , \mathcal{N}_2 against \mathcal{M}_1 , \mathcal{N}_2 against \mathcal{M}_2 : One of the images to be compared has no makeup while the other has makeup.

The EERs (Equal Error Rates) of the matching scenarios considered in the YMU database are summarized in Table 1. The general observation here is that EERs for the \mathcal{N} vs \mathcal{M} cases are substantially higher than for the \mathcal{M} vs \mathcal{M} and \mathcal{N} vs \mathcal{N} cases. The EER for the \mathcal{N} vs \mathcal{N} case is in the range of 6.50% (LBP) to 10.85% (Verilook), whereas for the \mathcal{N} vs \mathcal{M} case the EER moves up to 23.68% (Verilook). The results show that the application of makeup results in a substantial challenge for automated face recognition algorithms, which needs to be addressed in the interest of security. The performances of the academic face matchers outperform that of the commercial matcher; this might be due to the fact that Verilook does not offer a tuning option. Figure 4 shows a subject in the YMU database whose face images have each been normalized and cropped. The figure also reports the related distance scores generated by the three face matchers. These scores illustrate the drop in similarity (and an increase in distance score) when makeup is applied to a subject's face.

Table 1. EER (%) of the face matchers on the YMU database. In all cases, comparing a makeup image (\mathcal{M}) against a no-makeup image (\mathcal{N}) decreases accuracy.

	\mathcal{M} vs \mathcal{M}	\mathcal{N} vs \mathcal{M}	\mathcal{N} vs \mathcal{N}
Gabor	11.59%	21.47%	7.01%
LBP	9.41%	18.71%	6.50%
Verilook	13.55%	23.68%	10.85%

Figure 5 illustrates the boxplot of genuine score distributions for one of the face matchers LBP. Here, we note the shift in genuine scores when makeup is applied (the center boxplot). More specifically, the distance score increases when an image with makeup is compared against an image without makeup⁷. It is reasonable to state at this time that

⁷We note that there are intra-class variations for both the \mathcal{N} versus \mathcal{N} and \mathcal{M} versus \mathcal{M} cases, which cause a decrease in the performance of the face matchers for those cases as well

makeup has the potential to reduce the matching accuracy of face recognition algorithms thereby presenting an increased security risk.

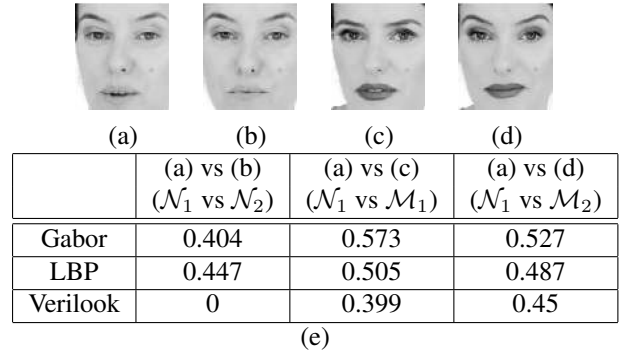


Figure 4. Sample subject from the YMU database. Intra-class variations can be observed between the no-makeup shots (a) and (b) as well as the makeup shots (c) and (d). In (e), the distance scores as reported by the three face matchers are presented for the various matching scenarios.

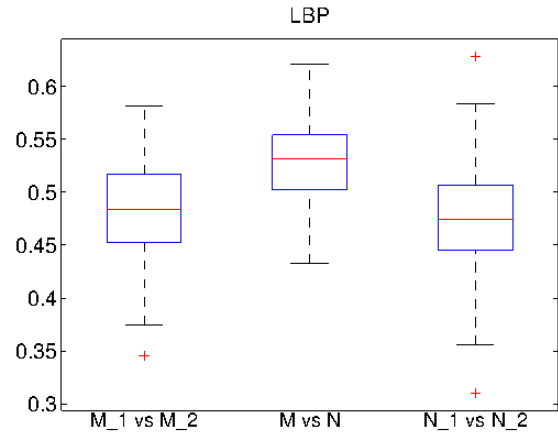


Figure 5. Boxplot of the genuine score distributions for the YMU database as computed using the LBP face matcher for the 3 matching scenarios: \mathcal{M}_1 vs \mathcal{M}_2 , \mathcal{M} vs \mathcal{N} , \mathcal{N}_1 vs \mathcal{N}_2 .

4.2. Experiments on the VMU Database

The VMU database contains 51 subjects. Each subject has one no makeup image and 3 after-makeup images. Let \mathcal{N} denote the image without makeup. Let \mathcal{L} , \mathcal{E} and \mathcal{F} denote, respectively, images with lipstick only, with eye makeup only, and with full makeup. The following configurations were considered for generating genuine and impostor matching scores for each of the three face matchers.

1. Matching \mathcal{N} against \mathcal{L} : An image without makeup is compared against the same image with lipstick added.

2. Matching \mathcal{N} against \mathcal{E} : An image without makeup is compared against the same image with eye makeup added.
3. Matching \mathcal{N} against \mathcal{F} : An image without makeup is compared against the same image with full makeup added.

The EER values corresponding to the aforementioned scenarios are reported in Table 2. Note that the impact of eye makeup is more pronounced than that of lipstick on the performance of the three face matchers. Figure 6 shows a sample subject from the VMU database whose faces have been normalized and cropped. The figure also reports the distance scores generated by the three face matchers.

Table 2. EER (%) of the face matchers on the VMU database. Matching a no-makeup image (\mathcal{N}) against a full makeup image (\mathcal{F}) results in decreased performance.

	\mathcal{N} vs \mathcal{E}	\mathcal{N} vs \mathcal{L}	\mathcal{N} vs \mathcal{F}
Gabor	8.56%	8.15%	11.38%
LBP	4.32%	3.43%	4.79%
Verilook	25%	4.79%	30%

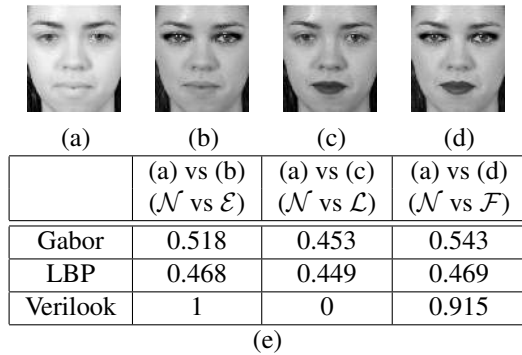


Figure 6. Example of normalized and cropped images for a subject in the VMU database. (a) No makeup. (b) Eye makeup. (c) Lipstick only. (d) Full makeup. In (e), the distance scores as reported by the three face matchers are presented for the various matching scenarios.

Figure 7 illustrates the boxplot of genuine score distributions for the LBP face matcher. Here, we note the shift in genuine scores between the different makeup styles (see the center boxplot). More specifically, we observe that the distance score due to lipstick is the lowest, while that for full makeup is the highest. We also observe that applying makeup to the ocular region (either separately or in the context of a full makeup) has a higher potential to reduce the matching accuracy of face recognition algorithms.

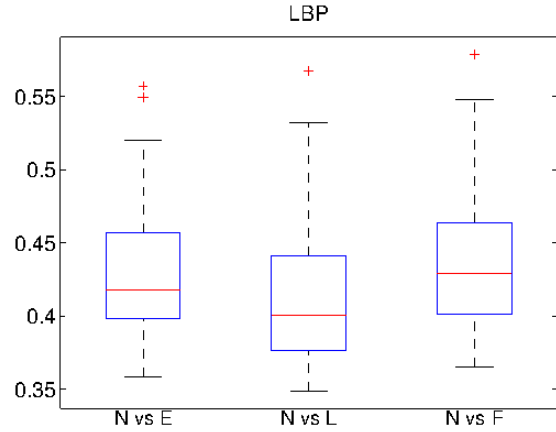


Figure 7. Boxplot of the genuine score distributions as computed by the LBP face matcher on the VMU database for the 3 matching scenarios: \mathcal{N} vs \mathcal{E} , \mathcal{N} vs \mathcal{L} and \mathcal{N} vs \mathcal{F} .

5. Observations

In this section we summarize the observations made from the experimental results.

- Although alterations due to cosmetic facial makeup are predominantly color-based, they clearly affect the performance of face matchers based on grayscale images.
- While the YMU database contains examples of natural makeovers, the VMU database depicts subjects with synthetic makeovers. However, the accuracy of face matchers decreases in both these databases when a face image without makeup is compared against a face with makeup. This is a clear indication that facial makeup introduces changes in the face that can affect the accuracy of a face recognition system.
- The impact due to the application of eye makeup is indicated to be the most pronounced. Individuals attempting to obfuscate (i.e., camouflage) their identity from a face recognition system may be able to do so by incorporating non-permanent cosmetic alterations to their ocular region. Since makeup is a socially acceptable cosmetic modification in most societies, these results indicate the possibility of compromising the security of a face or periocular biometric system [9].
- In this work, the application of makeup was not intended to deliberately undermine the security of a biometric system. However, it is not difficult to envision situations where an individual might utilize facial cosmetics to deliberately either obscure their identity or to impersonate another individual.

6. Addressing the problem

Here we investigate an existing algorithm known as Local Gabor Binary Pattern (LGBP) [17], which encodes a series of Gabor filtered images (as opposed to the original image) using LBP. We anticipate this method to be resilient, to a certain degree, to the effects of makeup. The LGBP matcher has the following steps:

1. The normalized face image is first convolved with a set of Gabor filters to generate the Gabor filter responses (Gabor magnitudes only) G_1, G_2, \dots, G_{40} .
2. The LBP coding method is applied to encode the filter responses resulting in 40 LBP maps L_1, L_2, \dots, L_{40} .
3. Each L_i is tessellated into 64 sub-regions, $S_{i1}, S_{i2}, \dots, S_{i64}$, where the size of each sub-region is 16×16 .
4. For each sub-region S_{ij} , $i = \{1, 2, \dots, 40\}$, $j = \{1, 2, \dots, 64\}$, a histogram, H_{ij} is extracted, which has 16 levels (bins). Each histogram value can be denoted as H_{ijl} , $l = \{1, 2, \dots, 16\}$.
5. For two face images, I^1 and I^2 , their respective histograms H^1 and H^2 each of length 40,960 are compared as $\phi(H^1, H^2) = \sum_{i=1}^{40} \sum_{j=1}^{64} \sum_{l=1}^{16} \min(H_{ijl}^1, H_{ijl}^2)$.

A histogram-type representation considers both the local and global level features and can exhibit robustness to certain variations in the image. Examples of LBP maps generated by this approach is shown in Figure 8.

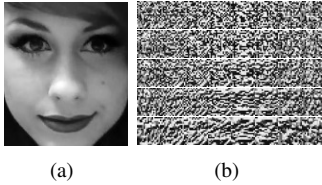


Figure 8. LGBP matcher. (a) Preprocessed sample makeup image from the YMU database. (b) The 40 LBP maps generated after applying LBP to 40 different Gabor magnitude responses.

The LGBP face matcher is applied to both the YMU and VMU databases. See Tables 3 and 4. The results are summarized below:

1. In the YMU database, LGBP results in an EER of 15.89% for the \mathcal{N} vs \mathcal{M} matching case (no make up against makeup) compared to 18.71% by the LBP approach and 21.47% by the Gabor approach.
2. In the VMU database, LGBP results in EERs of 5.44% for \mathcal{N} vs \mathcal{E} , 2.90% for \mathcal{N} vs \mathcal{L} and 5.42% for \mathcal{N} vs \mathcal{F} .

Compared to the LBP, Gabor and VeriLook matchers (see Table 2), the LGBP matcher results in better accuracy.

Table 3. EER (%) of LGBP matcher on the YMU database.

	\mathcal{M} vs \mathcal{M}	\mathcal{N} vs \mathcal{M}	\mathcal{N} vs \mathcal{N}
LGBP	7.58%	15.89%	3.78%

Table 4. EER (%) of LGBP matcher on the VMU database.

	\mathcal{N} vs \mathcal{E}	\mathcal{N} vs \mathcal{L}	\mathcal{N} vs \mathcal{F}
LGBP	5.44%	2.90%	5.42%

The ROC curves of all four algorithms on the YMU database are presented in Figure 9. We can clearly see the performance degradation of the Gabor, LBP and VeriLook matchers in the \mathcal{N} vs \mathcal{M} case. LGBP outperforms the other algorithms; however, with an EER of 15.89% the presence of makeup remains a challenge for face recognition algorithms.

7. Summary and Future Work

In this work we presented preliminary results on the impact of facial makeup on automated face recognition. This is the first work that explicitly establishes the impact of facial makeup on automated biometric systems. We illustrated that non-permanent facial cosmetics can significantly change facial appearance, both locally and globally, by altering color, contrast and texture. Existing face matchers, which rely on such contrast and texture information for establishing a match, can be impacted by the application of facial makeup. We provided clear experimental evidence of the decrease in matching accuracy when makeup is used. Further, we designed a method to mitigate the effect of makeup on matching performance. Future work will involve establishing a more detailed experimental protocol that quantifies the *degree* of makeup applied to a subject's face, and developing algorithms that are robust to changes introduced by facial makeup. While only female subjects were considered in this work, a more elaborate experiment involving both male and female subjects, in the context of spoofing and obfuscation, is currently being designed.

References

- [1] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer. A sparse representation approach to face matching across plastic surgery. In *Proceedings of WACV*, 2012.
- [2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: application to face recognition. *IEEE Trans. on PAMI*, 28:2037–2041, 2006.

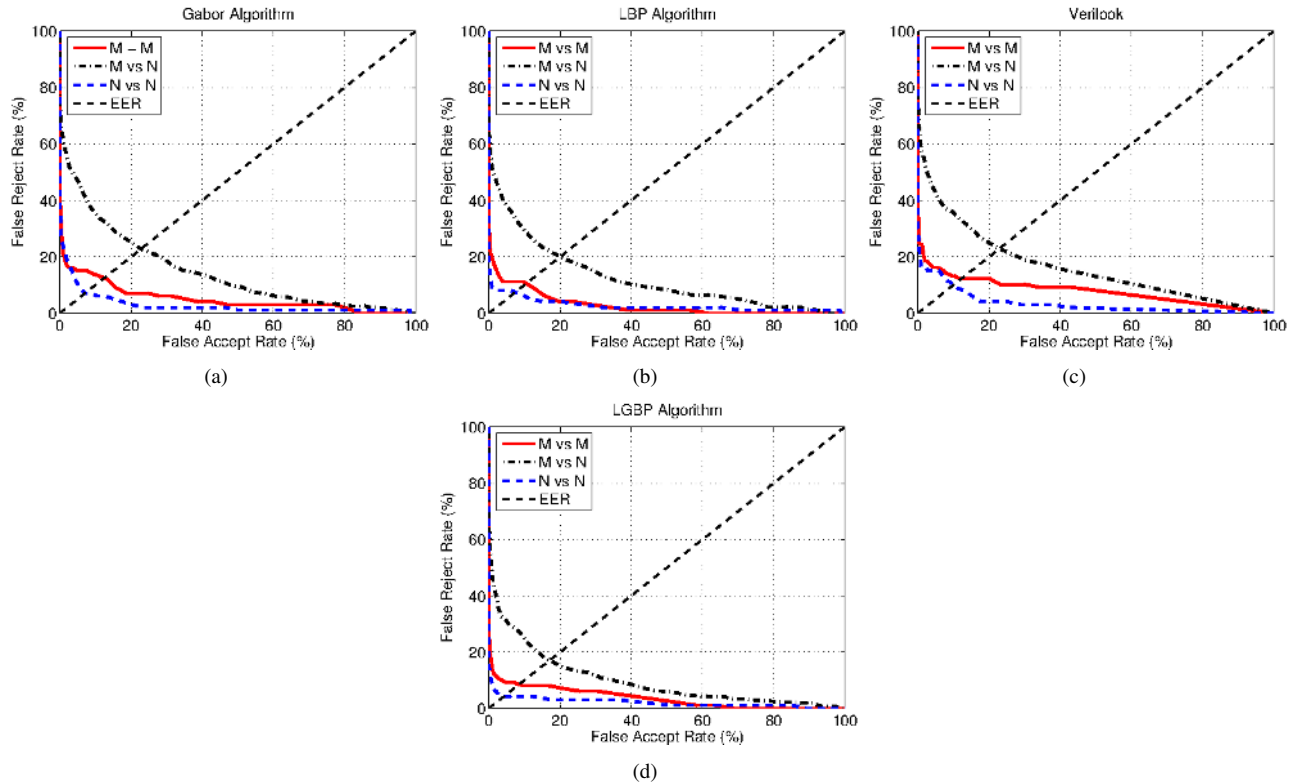


Figure 9. Performance of the four face matchers on the YMU database. (a) Gabor; (b) LBP; (c) Verilook; (d) LGBP. We observe consistent performance decrease in the case \mathcal{M} vs \mathcal{N} for all 4 algorithms. LGBP provides the best results with an EER of 15.89%; this error rate suggests that makeup remains a challenge and requires more research.

- [3] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Evolutionary granular computing approach for recognizing face images altered due to plastic surgery. In *Proceedings of IEEE FG*, 2011.
- [4] A. Dantcheva and J.-L. Dugelay. Female facial aesthetics based on soft biometrics and photo-quality. In *Proceedings of IEEE ICME*, 2011.
- [5] M. Everingham, J. Sivic, and A. Zisserman. Hello! My name is... Buffy” – Automatic naming of characters in TV video. In *Proceedings of BMVC*, pages 899–908, 2006.
- [6] N. Gueguen. The effects of women’s cosmetics on men’s courtship behaviour. *North American Journal of Psychology*, 10:221–228, 2008.
- [7] S. Z. Li and A. K. Jain. *Handbook of Face Recognition*. Springer, New York, 2011.
- [8] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. on Image Processing*, 11(4):467–476, 2002.
- [9] U. Park, R. Jillela, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, 2011.
- [10] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, M. Jaesik, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of CVPR*, 2005.
- [11] G. Rhodes, A. Sumich, and G. Byatt. Are average facial configurations only attractive because of their symmetry? *Psychological Science*, 10:52–58, 1999.
- [12] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz, and H.-P. Seidel. Computer-suggested facial makeup. *Proc. Eurographics*, 30(2), 2011.
- [13] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreydzan. Plastic surgery: A new dimension to face recognition. *IEEE Trans. on Information Forensics and Security*, 5(3):441–448, 2010.
- [14] R. Singh, M. Vatsa, and A. Noore. Effect of plastic surgery on face recognition: a preliminary study. In *Proceedings of IEEE CVPR Workshop on Biometrics*, 2009.
- [15] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. on Image Processing*, 19(6):1635–1650, 2010.
- [16] S. Ueda and T. Koyama. Influence of make-up on facial recognition. *Perception*, 39:260–264, 2010.
- [17] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *Proceedings of ICCV*, pages 786–791, 2005.