Impact of Facial Cosmetics on Automatic Gender and Age Estimation Algorithms

Cunjian Chen[1], Antitza Dantcheva[2] and Arun Ross[2]

[1]*Computer Science and Electrical Engineering, West Virginia University, Morgantown, U.S.A.*

[2]*Computer Science and Engineering, Michigan State University, East Lansing, U.S.A.*

Keywords: Biometrics, Face Recognition, Facial Cosmetics, Makeup, Gender Spoofing, Age Alteration, Automatic Gender Estimation, Automatic Age Estimation.

Abstract: Recent research has established the negative impact of facial cosmetics on the matching accuracy of *automated* face recognition systems. In this paper, we analyze the impact of cosmetics on *automated* gender and age estimation algorithms. In this regard, we consider the use of facial cosmetics for (a) gender spoofing where male subjects attempt to look like females and vice versa, and (b) age alteration where female subjects attempt to look younger or older than they actually are. While such transformations are known to impact *human perception*, their impact on *computer vision* algorithms has not been studied. Our findings suggest that facial cosmetics can potentially be used to confound automated gender and age estimation schemes.

1 INTRODUCTION

Recent studies have demonstrated the negative impact of facial cosmetics on the matching accuracy of automated face recognition systems (Dantcheva et al., 2012; Eckert et al., 2013). Such an impact has been attributed to the ability of makeup to alter the perceived shape, color and size of facial features, and skin appearance in a simple and cost efficient manner (Dantcheva et al., 2012).

The impact of makeup on *human perception* of faces has received considerable attention in the psychology literature. Specifically, the issues of identity obfuscation (Ueda and Koyama, 2010), sexual dimorphism (Russell, 2009), and age perception (Nash et al., 2006) have been analyzed in this context. Amongst other things, these studies show that makeup can lead to higher facial contrast thereby enhancing female-specific traits (Russell, 2009), as well as smoothen and even out the appearance of skin thereby imparting an age defying effect (Russell, 2010). This leads us to ask the following question: can makeup also confound *computer vision* algorithms designed for gender and age estimation from face images? Such a question is warranted for several reasons. Firstly, makeup is widely used and has become a daily necessity for many, as reported in

a recent British poll of 2,000 women[1], and as evidenced by a 3.6 Billion sales volume in 2011 in the United States[2]. Secondly, a number of commercial software have been developed for age and gender estimation[3,4,5]. Thus, it is essential to understand the limitations of these software in the presence of facial makeup. Thirdly, due to the use of such software in surveillance applications (Reid et al., 2013), anonymous customized advertisement systems[6] and image retrieval systems (Bekios-Calfa et al., 2011), it is imperative that they account for the presence of makeup if indeed they are vulnerable to it. Fourthly, gender and age have been proposed as soft biometric traits in automated biometric systems (Jain et al., 2004). Given the widespread use of facial cosmetics, understanding the impact of makeup on these traits would help in accounting for them in biometric systems. Hence, the motivation of this work is to quantify the impact of makeup on gender and age

---

[1]www.superdrug.com/content/ebiz/superdrug/stry/ cgq1300799243/survey release - jp.pdf

[2]www.npd.com/wps/portal/npd/us/news/press-releases/ pr_120301/

[3]www.neurotechnology.com/face-biometrics.html

[4]www.visidon.fi/en/Face_Recognition#3

[5]www.cognitec-systems.de/FaceVACS- VideoScan.20.0.html

[6]articles.latimes.com/2011/aug/21/business/ la-fi-facial-recognition-20110821

estimation algorithms.

However, there is little work establishing the impact of cosmetics on gender and age estimation algorithms. Only one recent publication has considered the effect of makeup on age estimation (Feng and Prabhakaran, 2012), where an age index was used to adjust parameters in order to improve the system's accuracy.

In this work, we seek to answer the following questions:

- Can facial makeup be used to spoof gender with respect to an automated gender estimation algorithm?

- Can the use of facial makeup confound an automated age estimation algorithm?

Towards answering these questions, we first assemble two datasets consisting of a) male subjects applying makeup to look like females and vice-versa, and b) female subjects applying makeup to conceal aging effects. Subsequently, we test gender and age estimation algorithms on these two datasets, respectively. Experimental results suggest that gender and age estimation systems can be impacted by the application of facial makeup. To the best of our knowledge, this is the first work to systematically demonstrate these effects. The results appear intuitive, since humans may have similar difficulties in estimating gender and age after the application of makeup. However, as reported in a recent study in the context of face recognition (Rice et al., 2013), human perception and machine estimation can be significantly different. This becomes especially apparent when only cropped images of the face are considered, without the surrounding hair and body information. In this work, only cropped face images are used for assessing impact of makeup on automated gender and age estimation algorithms.

The rest of the paper is organized as follows. Section 2 introduces the problem of makeup-based gender alteration, presents the assembled dataset in Section 2.1, discusses the employed estimation algorithms in Section 2.3, and reports related results in Section 2.4. Section 3 introduces the problem of makeup induced age alteration, presents the assembled dataset in Section 3.1, discusses the employed age estimation algorithm in Section 3.3, and summarizes the results in Section 3.4. Section 4 discusses the results and Section 5 concludes the paper.

2 MAKEUP INDUCED GENDER ALTERATION

Interviews conducted by Dellinger and Williams (Dellinger and Williams, 1997) suggested that women used makeup for several perceived benefits including revitalized and healthy appearance, as well as increased credibility. However, makeup can also be used to alter the perceived gender, where a male subject uses it to look like a female (Figure 1(a) and Figure 1(b)), or a female subject uses it to look like a male (Figure 1(c) and Figure 1(d)). The makeup in both cases is used to conceal original gender specific cues and enhance opposite gender characteristics. For instance, in the *male-to-female alteration* case, the facial skin is first fully covered by foundation (to conceal facial hair and skin blemishes), and then eye and lip makeup (e.g., eye shadow, eye kohl, mascara and lipstick) are applied in the way females usually do. In the *female-to-male alteration* case, the contrast in the eye and lip areas is decreased using foundation, skin blemishes and contours (e.g., around the nose) are added (e.g., by using brown eye shadow), and male features such as mustache and beard are simulated (e.g., by using eye kohl).

We study the potential of such cosmetic applications to confound automatic face-based gender classification algorithms that typically rely on the texture and structure of the face image to distinguish between males and females (Chen and Ross, 2011). While some algorithms (Li et al., 2012) might also exploit cues from clothing, hair, and other body parts for gender prediction, in this study we consider only the facial region. Therefore, we focus only on the cropped face, which minimizes the inclusion of factors such as hair, clothing and other accessories (see Figure 3).

2.1 Makeup Induced Gender Alteration (MIGA) Dataset

To study makeup induced gender alteration, we searched the Web and assembled a dataset consisting of two subsets:

- *Male subset* consisting of 120 images of 30 subjects (2 before makeup and 2 after makeup images per subject): male subjects apply makeup to look like females,

- *Female subset* consisting of 128 images of 32 subjects (2 before makeup and 2 after makeup images per subject): female subjects apply makeup to look like males.

(a) Original male subject without makeup

(b) Male subject after makeup application

(c) Original female subject without makeup
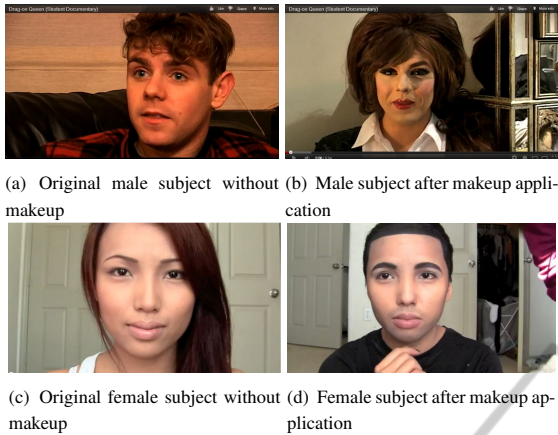
(d) Female subject after makeup application

Figure 1: Examples of subjects applying facial makeup for gender spoofing (from YouTube). Male-to-female (a-b): foundation conceals facial hair and skin blemishes; eye and lip makeup are then applied in the way females usually do. Female-to-male (c-d): dark eye-shadow is used to contour the face shape and the nose; then, thicker eye-brows, mustache, and beard are simulated using special makeup products. Only the facial region is used in this study (see Figure 3).

(a) Male-to-female subset of the MIGA dataset

(b) Female-to-male subset of the MIGA dataset

Figure 2: Example images from the Makeup Induced Gender Alteration (MIGA) dataset: (a) male-to-female subset: male subjects apply makeup to look like females, and (b) female-to-male subset: female subjects apply makeup to look like males. In both (a) and (b), the images in the upper row are before makeup and the ones below are the corresponding images after makeup.

The images were obtained from makeup transformation tutorials posted on YouTube, and the images exhibit differences in illumination and resolution, while subjects exhibit differences in race, facial pose and expression (see Figure 2). Note that the subjects were not trying to deliberately mislead automated systems. Despite the relatively small size of

(a) Male-to-female subset of the MIGA dataset

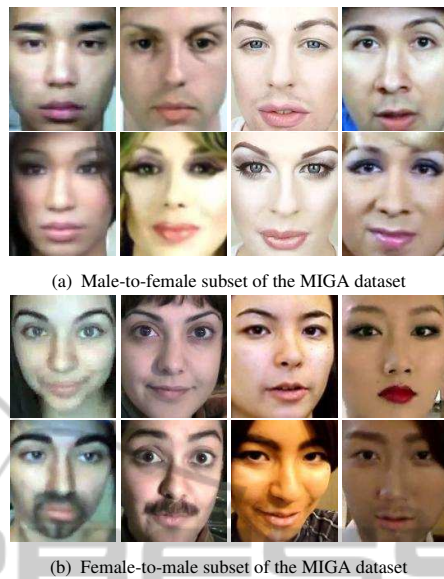(b) Female-to-male subset of the MIGA dataset

Figure 3: The example images from Figure 2 after preprocessing.

the dataset, it enables us to investigate the potential of makeup to confound computer vision-based gender classification systems.

2.2 Gender Classification and Alteration Metrics

For performance evaluation of gender classification systems, we define two classification rates:

- *Male Classification Rate*: the percentage of images (before or after makeup) that are classified as male by the gender classifier.

- *Female Classification Rate*: the percentage of images (before or after makeup) that are classified as female by the gender classifier.

Additionally, we introduce a metric called *gender spoofing index* (*GSI*) that quantifies the success of cosmetic induced gender spoofing. Let $\{S_1, \cdots, S_n\}$ be a set of face images, and let the corresponding label values be $\{v_1, \cdots, v_n\}$, where $v_i \in \{0, 1\}$, with 0 indicating male and 1 indicating female. Let $\{M_1, \ldots M_n\}$ denote the images after the application of makeup. If G denotes the gender classification algorithm, then *GSI* is defined as:

$$ GSI = \frac{\sum_{i=1}^{\ell} I(G(S_i) \neq G(M_i))}{\ell}, \qquad (1) $$

where $G(S_i)$ and $G(M_i)$ are the gender labels as computed by the algorithm for S_i and M_i, respectively, $\ell = \sum_{i=1}^{n} I(G(S_i) = v_i)$ denotes the number of face images before makeup that were correctly classified by

the algorithm and $I(x)$ is the indicator function, where $I(x) = 1$ if x is true and 0 otherwise. In summary, *GSI* represents the percentage of face images whose gender prediction labels were changed after the application of makeup for those face images whose before makeup labels were correctly predicted.

Our hypothesis is that, if makeup can be used for gender spoofing, then the male classification rate will decrease after male-to-female alteration; and the female classification rate will decrease after female-to-male alteration.

2.3 Gender Estimation Algorithms

To study the effectiveness of makeup induced gender spoofing, we annotate the eyes of the subjects, crop the images to highlight the face region only (see Figure 3) and utilize three state-of-the-art gender classification algorithms (academic and commercial).

Commercial Off-the-Shelf (COTS). COTS is a commercial face detection and recognition software, which includes a gender classification routine. While the underlying algorithm and the training dataset that were used are not publicly disclosed, it is known that COTS performs well in the task of gender classification. To validate this, we first perform an experiment on a face dataset[7] consisting of 59 male and 47 female faces that is a subset of the FERET database and which has been used extensively in the literature for evaluating gender classifiers. COTS obtains male and female classification accuracies of 96.61% and 97.87%, respectively, on this dataset. The system does not provide a mechanism to re-train the algorithm based on an external dataset; instead it is a black box that outputs a label (i.e., male or female) along with a confidence value.

Adaboost. The principle of Adaboost (Bekios-Calfa et al., 2011) is to combine multiple weak classifiers to form a single strong classifier as $y(\mathbf{x}) = \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x})$, where $h_t(\mathbf{x})$ refers to the weak classifiers operating on the input feature vector $\mathbf{x}$, T is the number of weak classifiers, α_t is the corresponding weight for each weak classifier and $y(\mathbf{x})$ is the classification output. In this work, feature vector $\mathbf{x}$ consists of pixel values from a 24×24 image of the face. For every pair of feature values (x_i, x_j) in the feature vector $\mathbf{x}$, five types of weak binary classifiers are defined:

$$h_t(\mathbf{x}) \equiv \{g_k(x_i, x_j)\}, \qquad (2)$$

where $i, j = 1 \ldots 24$, $i \neq j$, $k = 1 \ldots 5$, and

$$g_k(x_i, x_j) = 1, \quad if (x_i - x_j) > t_k, \qquad (3)$$

where $t_1 = 0$, $t_2 = 5$, $t_3 = 10$, $t_4 = 25$ and $t_5 = 50$. By changing the inequality sign in (3) from $>$ to $<$, another five types of weak classifiers can be generated, resulting in a total of 10 types of weak classifiers. Since g_k is non-commutative, $g_k(x_i, x_j) \neq g_k(x_j, x_i)$, the total number of weak classifiers for a pair of features x_i and x_j is 20. The total number of weak classifiers selected by the AdaBoost algorithm, T, is 1,000.

In order to utilize the Adaboost method for gender classification, the AR database[8] (350 males and 350 females) was used to train the gender classifier. Each image is rescaled to 24×24 and the column vectors consisting of pixel values are concatenated together to form the feature vector $\mathbf{x}$. Adaboost obtains male and female classification accuracies of 86.44% and 82.98%, respectively, on the FERET subset[7].

OpenBR. OpenBR (Klontz et al., 2013) is a publicly available toolkit for biometric recognition and evaluation. The gender classification algorithm utilized in this toolkit is based on (Klare et al., 2012). An input face image is represented by extracting histograms of local binary pattern (LBP) and scale-invariant feature transform (SIFT) features computed on a dense grid of patches. The histograms from each patch are then projected onto a subspace generated using PCA in order to obtain a feature vector. Support Vector Machine (SVM) is used for classification. The efficacy of the OpenBR gender classification algorithm is again validated on the FERET subset indicated above. It attains accuracies of 96.91% and 82.98% for male and female classification, respectively. The overall true classification rate is 90.57%, which outperforms the other algorithms (Neural Network, Support Vector Machine, etc.) reported in (Makinen and Raisamo, 2008).

2.4 Gender Spoofing Experiments

We conduct two experiments, corresponding to the two subsets in MIGA: (a) male-to-female spoofing; and (b) female-to-male spoofing. For experiment (a) we report the male classification rates, and for experiment (b) we report the female classification rates, as elaborated in Section 2.2. In either case, the accuracy of gender classification before and after the application of makeup are independently determined. Figure 4 presents the output of COTS on some sample images before and after makeup.

a) Male-to-Female Alteration. We report the classification rates of the three gender classification algorithms on the first subset of the MIGA dataset in

---

[7]www.cs.uta.fi/hci/mmig/vision/datasets/

[8]www2.ece.ohio-state.edu/%7Ealeix/ARdatabase.html

Table 1: Male classification rates (%) and *GSI* values (%) corresponding to the three gender classification algorithms on the male-to-female subset of the MIGA dataset.

	Before Makeup	After Makeup	*GSI*
COTS	68.33	6.67	95.12
AdaBoost	78.33	30.0	70.21
OpenBR	55.0	15.0	75.76

Table 2: Female classification rates (%) and *GSI* values (%) corresponding to the three gender recognition algorithms on the female-to-male subset of the MIGA dataset.

	Before Makeup	After Makeup	*GSI*
AdaBoost	53.13	9.37	88.24
COTS	100.0	39.06	60.94
OpenBR	71.88	46.87	47.83

Table 1. COTS obtains a 68.33% male classification rate before makeup and 6.67% after makeup. The *GSI* value is 95.12%. This suggests that for most of the correctly classified male subjects before makeup, the application of makeup alters the gender from the perspective of the commercial system. We observe that AdaBoost has a male classification rate of 78.33% on the before makeup images and 30% on the after makeup images. The *GSI* value for AdaBoost was 70.21%, which suggests that makeup was successful in altering the perceived gender of a good number of face images. OpenBR results in a similar trend where the male classification rate drops from 55.0% to 15.0% after makeup. The corresponding *GSI* value is 75.76%.

b) Female-to-Male Alteration. We report the classification rates of the three gender classification algorithms on the second subset of the MIGA dataset in Table 2. For Adaboost, the female classification rate decreases from 53.13% before makeup to 9.37% after makeup. A GSI of 88.24% indicates that gender alteration, with respect to the classifier, was successful for a good number of subjects whose before makeup images was correctly classified as female. COTS has a 100% female classification rate before makeup, which drops to 39.06% after the application of makeup. OpenBR obtains a 71.88% female classification rate before makeup, which drops to 46.87% after makeup. The *GSI* values for COTS and OpenBR are 60.94% and 47.83%, respectively.

We note from these experiments that some subjects can successfully apply makeup to alter the perceived gender, thereby misleading gender classification systems. Interestingly, we observe that female-to-male alteration is slightly more challenging than its counterpart. The difference can be noticed in the *GSI* values (see Table 1 and Table 2); specifically the male-to-female subset has a higher *GSI* value (e.g.,

COTS and OpenBR). A possible explanation for this observation is that male characteristics are easier to be concealed using makeup (e.g., heavy makeup), than to be created. However, it must also be noted that the three algorithms perform very differently in the gender classification task. The differential performance observed across the three algorithms on male/female classification rates could be due to the implicit differences in the features that they employ and the dataset used to train the individual algorithms.

3 MAKEUP INDUCED AGE ALTERATION

Makeup can also be used to alter the perceived age of a person. This is accomplished by concealing wrinkles and age spots (using light foundation and a concealer), by brightening wrinkle-induced shadows in eye, nose and mouth regions (using concealer and powder), and by highlighting and coloring cheeks (using highlighter and blush). One reason for women to wear makeup is to increase their perceived competence and credibility (Kwan and Trautner, 2009). In other cases, the goal of applying makeup is to make a subject look younger, while in the case of young subjects the opposite effect might be desired (Kwan and Trautner, 2009). Here, we seek to observe the impact of makeup on an automated age estimation algorithm. We minimize other confounding factors (e.g., hair and accessories) by using cropped and aligned faces.

3.1 Description of Datasets

We first conduct experiments on the YMU and VMU datasets (Dantcheva et al., 2012), which were originally used to study the impact of makeup on automatic face *recognition* algorithms[9]. YMU consists of face images of 151 young Caucasian females obtained from YouTube makeup tutorials. For each subject, there are two images before and two images after makeup. VMU contains face images of 51 female Caucasian subjects from the FRGC database, to which makeup was synthetically applied using the Taaz software (Dantcheva et al., 2012). In VMU three types of makeup were virtually applied: lipstick, eye, and full makeup; hence there are four images per subject (one before makeup, one with lipstick only, one with eye makeup only and one with full makeup). In YMU and VMU the application of makeup was primarily to improve facial aesthetics, i.e., they were not

---

[9]www.antitza.com/makeup-datasets.html

(a) Male-to-female subset of the MIGA dataset

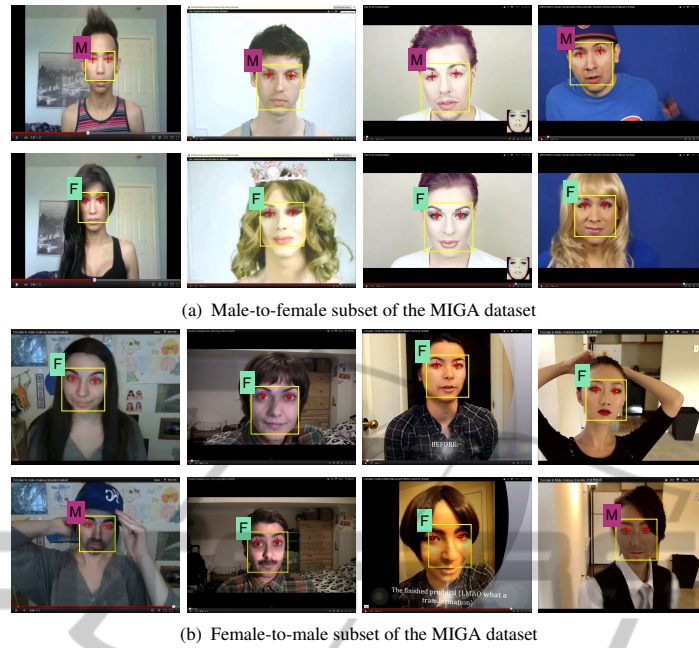(b) Female-to-male subset of the MIGA dataset

Figure 4: The output of the COTS gender classifier on the images shown in Figure 2. M indicates "classified as male" and F indicates "classified as female". While all male-to-female transformations were successful in this example, only the leftmost and rightmost female-to-male transformations were successful.

Figure 5: Example image pairs from the MIAA dataset. Top row: images before makeup; Bottom row: corresponding images after makeup.

applied with the intention of defeating a biometric system.

Additionally, to study makeup induced age alteration, we assembled another dataset (MIAA - Makeup Induced Age Alteration) consisting of images downloaded from the World Wide Web. These images correspond to 53 subjects, with one image before and one after the application of makeup per subject (see Figure 5). While the ground truth of a subject's age is not available, we estimate that the subjects are older than 30 years and that makeup is applied with the goal of both improving aesthetics as well as making subjects look younger. However, for our study, knowledge about the absolute age of the subject is not required, since we are only interested in age *difference* between the before and after makeup images as assessed by the algorithm. Since the before and after makeup images correspond to the same session, the primary confounding covariate between them is the presence or absence of makeup.

In summary, the role of makeup in YMU, VMU and MIAA datasets is different. Subjects in MIAA knowingly apply makeup to appear younger, while in YMU and VMU makeup is not specifically used for anti-aging purpose.

3.2 Age Estimation and Alteration Metrics

The performance of the automated age estimation algorithm is calculated by the *Mean Absolute Error (MAE)*: $MAE = \sum_{k=1}^{N} |g_k - \widehat{g}_k|/N$. Here g_k is the ground-truth-age, $\widehat{g}_k$ the estimated age for the k-th image, and N the number of test images. Age alteration is measured by *Mean Absolute Difference (MAD)*, which is computed as $MAD = \sum_{k=1}^{N} |\widehat{g}_k^b - \widehat{g}_k^a|/N$. Here $\widehat{g}_k^b$ is the estimated age of the before-makeup image and $\widehat{g}_k^a$ is the estimated age of the after-makeup image.

Our hypothesis here is that the estimated image after makeup will be significantly different than the estimated image before makeup. If so, this would indicate that makeup has the ability to impact age estimation algorithms.

3.3 Age Estimation Algorithm

The age estimation software used in this work utilizes the same feature set as the gender classifier in

OpenBR (see Section 2.3), along with SVM regression. We first evaluate the reliability of this algorithm on a large-scale dataset, namely a specific subset of Morph (Ricanek Jr. and Tesafaye, 2006). This dataset contains 10,000 images of subjects in the age range 20 to 75. There are four age groups: 20-40 (2,514 images), 40-50 (5,524 images), 50-60 (1,790 images) and 60-70 (172 images). The MAEs of the OpenBR algorithm on these groups are 5.46, 5.75, 6.47, and 7.88, respectively. We note that the algorithm performs better on the youngest age group (20-40). The algorithm obtains an MAE of 5.84 years on the entire test set (10,000 images). The best reported performance on the entire MORPH database is an MAE of 4.18 years as reported in (Guo and Mu, 2011) based on the kernel-based partial least squares regression method.

3.4 Age Alteration Experiments

First, we conduct age estimation experiments on YMU and VMU in order to a) show the impact of makeup on age estimation (YMU), and b) study this impact based on the type of makeup used (VMU). Towards this, we use the OpenBR software to estimate the age for all images and compute the differences in estimated age, before and after makeup, for each subject:

- Age difference of B versus B (B vs B): Both face images are before makeup.

- Age difference of A versus A (A vs A): Both face images are after makeup.

- Age difference of A versus B (A vs B): One of the face images is after makeup while the other is before makeup.

We observe that age differences between after makeup images and before makeup images (A vs B) are larger than in the other two cases (see Figure 6(a)). This suggests that makeup does have an impact on the performance of automated age estimation. Next, we perform age estimation on the VMU dataset and compute age differences corresponding to (N vs L): a face without makeup versus the same face with lipstick; (N vs E): a face without makeup versus the same face with eye makeup; and (N vs F): a face without makeup versus the same face with full makeup (foundation, eye and lip makeup). We observe that the use of lipstick (N vs L) has a lower impact on age estimation, compared to the application of eye makeup (N vs E) and full makeup (N vs F) (see Figure 6(b)).

Next, we conduct experiments on the MIAA dataset, in order to evaluate the age defying effect of
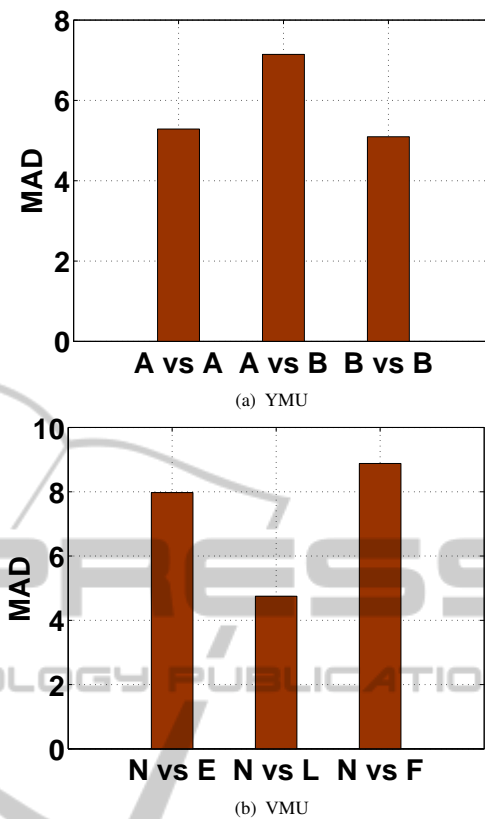

(a) YMU

(b) VMU

Figure 6: MAD results of the OpenBR age estimation algorithm on the (a) YMU and (b) VMU datasets. (a) We observe a higher MAD value for the A vs B case; (b) We observe higher MAD values for eye makeup and full makeup.

makeup on computer vision-based age estimation algorithms. To make this assessment, we use the difference rather than the absolute difference when comparing the before and after makeup images. Results indicate that 56.61% of the subjects are estimated to be younger after the application of makeup. Specifically, 32.08% are estimated as being 5 or more years younger, with a maximum being 20 years younger (see Figure 7). In order to validate the significance of the above result, we perform a one-sided hypothesis test with $H_1 : \mu_b - \mu_a < 0$, where μ_b is the mean age for before makeup images and μ_a is the mean age for after makeup images. The null hypothesis is $H_0 : \mu_b - \mu_a = 0$. The one-sided hypothesis test rejects the null hypothesis at the 5% significance level. It is therefore evident that makeup does have an age-defying effect. Moreover, a MAD of 7.67 is obtained on the MIAA dataset, which is larger than the MAE value (5.84) obtained on the Morph dataset, thus suggesting that the change observed in estimated age is significant even after taking the error tolerance into account. The output of OpenBR on some sample images is presented in Figure 8.
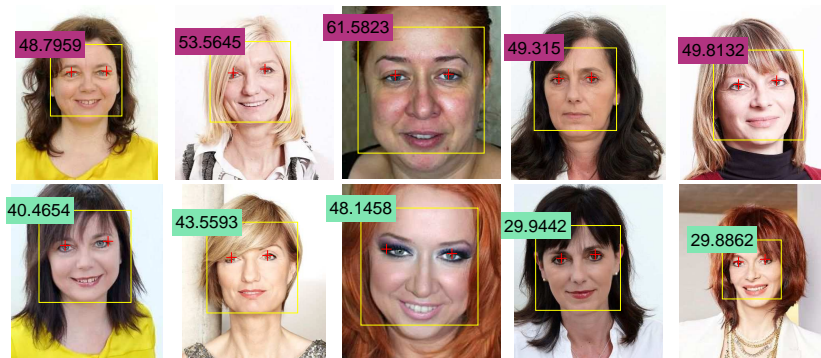
Figure 8: Automatic age estimation results before and after the application of makeup (original images are shown in Figure 5). Top row: images before makeup; Bottom row: corresponding images after makeup.
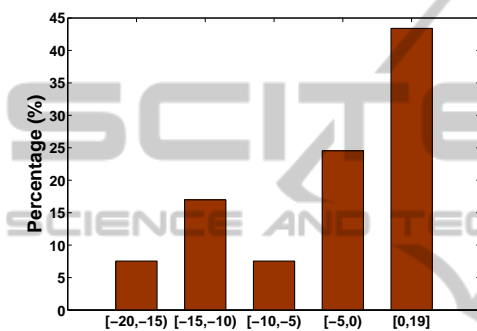

Figure 7: Age defying effect of makeup on the OpenBR algorithm. x-axis values indicate the difference in estimated age in years (after makeup - before makeup), while y-axis values show the percentage of subjects in MIAA.

4 DISCUSSION

We summarize the main observations from the experiments conducted in this work[10]:

- Makeup induced gender spoofing does impact automated gender classification systems. The observation holds for male-to-female, as well as for female-to-male alterations. The female-to-male alteration was observed to be slightly more challenging than its counterpart.

- The application of makeup can impact automatic age estimation algorithms.

These observations point out the need for developing robust gender and age estimation methods that are less impacted by the application of makeup. There are several ways to potentially address this issue:

- Whenever makeup is detected, an image pre-processing scheme can be used to mitigate its ef-

fect, as was shown in the context of face recognition (Chen et al., 2013).

- As demonstrated in the work of Feng and Prabhakaran (Feng and Prabhakaran, 2012), the estimated age can be adjusted accordingly after makeup is detected.

- The accuracy of gender and age estimation algorithms depends on the features used to represent the face, as well as the classifier used to estimate these attributes. Therefore, exploring different types of feature sets and classifiers will be necessary to devise a robust solution.

- The training set used by the learning algorithms can be suitably populated with face images having makeup. This would help the algorithm learn to estimate gender and age in the presence of makeup.

- Age (or gender) can be independently estimated using different components of the face and the independent estimates can be combined to generate a final estimate.

5 CONCLUSIONS

In this work we presented preliminary results on the impact of facial makeup on automated gender and age estimation algorithms. Since automated gender and age estimation schemes are being used in several commercial applications (see Section 1), this research suggests that the issue of makeup has to be accounted for. While a subject may not use makeup to intentionally defeat the system, it is not difficult to envision scenarios where a malicious user may employ commonly-used makeup to deceive the system. Future work will involve developing algorithms that are robust to changes introduced by facial makeup.

---

[10]Details about obtaining the MIGA and MIAA datasets will be posted at www.antitza.com/makeup-datasets.html

ACKNOWLEDGEMENTS

REFERENCES

Bekios-Calfa, J., Buenaposada, J. M., and Baumela, L. (2011). Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):858–864.

Chen, C., Dantcheva, A., and Ross, A. (2013). Automatic facial makeup detection with application in face recognition. In *IAPR International Conference on Biometrics (ICB)*.

Chen, C. and Ross, A. (2011). Evaluation of gender classification methods on thermal and near-infrared face images. In *International Joint Conference on Biometrics (IJCB)*, pages 1–8.

Dantcheva, A., Chen, C., and Ross, A. (2012). Can facial cosmetics affect the matching accuracy of face recognition systems? In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*.

Dellinger, K. and Williams, C. L. (1997). Makeup at work: Negotiating appearance rules in the workplace. *Gender and Society*, 11(2):151–177.

Eckert, M.-L., Kose, N., and Dugelay, J.-L. (2013). Facial cosmetics database and impact analysis on automatic face recognition. In *IEEE International Workshop on Multimedia Signal Processing*.

Feng, R. and Prabhakaran, B. (2012). Quantifying the makeup effect in female faces and its applications for age estimation. In *IEEE International Symposium on Multimedia*, pages 108–115.

Guo, G. and Mu, G. (2011). Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 657–664.

Jain, A. K., Dass, S. C., and Nandakumar, K. (2004). Can soft biometric traits assist user recognition? In *Proceedings of SPIE Defense and Security Symposium*, volume 5404, pages 561–572.

Klare, B., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., and Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, 7(6):1789–1801.

Klontz, J., Klare, B., Klum, S., Taborsky, E., Burge, M., and Jain, A. K. (2013). Open source biometric recognition. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*.

Kwan, S. and Trautner, M. N. (2009). Beauty work: Individual and institutional rewards, the reproduction of gender, and questions of agency. *Sociology Compass*, 3(1):49–71.

Li, B., Lian, X.-C., and Lu, B.-L. (2012). Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing*, 76(1):18–27.

Makinen, E. and Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):541–547.

Nash, R., Fieldman, G., Hussey, T., Leveque, J.-L., and Pineau, P. (2006). Cosmetics: They influence more than caucasian female facial attractiveness. *Journal of Applied Social Psychology*, 36(2):493–504.

Reid, D., Samangooei, S., Chen, C., Nixon, M., and Ross, A. (2013). Soft biometrics for surveillance: An overview. In *Handbook of Statistics*, volume 31.

Ricanek Jr., K. and Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 341–345.

Rice, A., Phillips, P. J., Natu, V., An, X., and O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, (Published Online before Print, September 25).

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, 38(8):1211–1219.

Russell, R. (2010). *Why cosmetics work*. New York: Oxford University Press.

Ueda, S. and Koyama, T. (2010). Influence of make-up on facial recognition. *Perception*, 39:260–264.