A JOINT OPTIMIZATION SCHEME TO COMBINE DIFFERENT LEVELS OF FEATURES FOR FACE RECOGNITION WITH MAKEUP CHANGES

Zhenzhu Zheng[2] *and Guodong Guo*[1,2]

[1] Beijing Advanced Innovation Center for Imaging Technology, Beijing 100048, China
[2] Lane Department of CSEE, West Virginia University, Morgantown, WV 26506, USA

ABSTRACT

We address the problem of makeup face recognition. Our main idea is to incorporate different levels of features into a joint optimization framework. Specifically, we combine both mid-level (e.g. attributes) and low-level features to obtain a new representation for a better matching between makeup and non-makeup faces. Previous studies have discovered the influence of cosmetics on face recognition, and some methods have been developed to address makeup face recognition. However, it is for the first time to integrate different levels of features to address this challenging problem. To minimize the differences between makeup and non-makeup faces, transformations are learned in a joint optimization framework which combines different levels of features. Experimental results demonstrate that our proposed approach outperforms the state-of-the-art methods.

Index Terms— Face recognition, face makeup.

1. INTRODUCTION

It is common for people to ware makeup, especially females. The cosmetic techniques enable them to hide facial flaws and appear more attractive, however, makeup brings challenge to face recognition at the same time, even for humans.

Although a large amount of research work has been conducted to make face recognition more robust and achieve high performance, limited studies have focused on makeup face recognition. In fact, it is a challenging problem since dramatic changes can occur between the same face with and without makeup. Also, these changes vary due to diverse makeup styles when different cosmetic techniques are applied. Fig.1 shows appearances of the same person without and with makeup, respectively. As one can see that significant changes exist between those faces, thus a classifier trained on non-makeup face images does not generalize well to images of makeup faces.

Some studies [1, 2, 3, 4, 5, 6, 7] have been performed on the makeup face recognition problem. In [7], makeup-robust face verification is conducted by measuring correlations between face images in a meta subspace, then a discriminative learning is applied to verify faces based on the low-dimensional features in the learned meta subspace. [3] shows an intense study on the makeup face challenge and proposes a framework incorporating multi-view analysis methods to solve the problem. [2] solves this problem by introducing makeup-related facial attributes and use the intersection of attributes from both makeup and non-makeup faces for matching.

This work is partially supported by the Center for Identification Technology Research (CITeR). Corresponding author: Guodong Guo (email: guodong.guo@mail.wvu.edu).

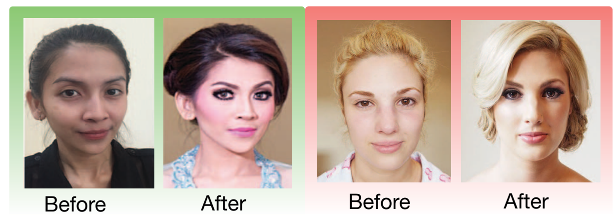

Fig. 1: Some face examples with non-makeup or makeup. The makeup changes brings challenges to face recognition.

We observe that some mid-level facial traits, such as attributes [8, 9, 10, 11, 12], remain unchanged before and after makeup, which might contribute to a good representation for makeup face recognition. However, intuitively a single level representation might not be good enough to fully elaborate a fine representation.

Fig. 2 shows two pairs of faces, non-makeup vs. makeup. The first pair comes from the same identity. It is easy to see that both faces share the same attributes between wearing makeup and not. In this case, the attributes provide an invariant representation and hence might somehow facilitate makeup face recognition. However, attributes are far from being a sufficient representation. As we can see in the second pair of faces, which share identical attributes, are in fact coming from two different identities. Hence, the limited discriminative capacity of attributes might lead to degraded performance for makeup face recognition.

To obtain a good representation, we propose to integrate mid-level and low-level features for makeup face recognition. In this way, shared information between makeup and non-makeup faces, as well as fine details are preserved. Further, a

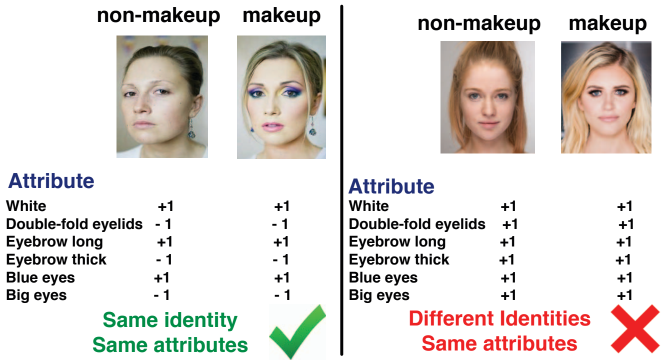

Fig. 2: Attributes: good and bad. Attributes provide an invariant representation which might facilitate makeup face recognition. On the left side, the same identity shares identical attributes between non-makeup and makeup faces. However, attributes have a limited capacity for discrimination. On the right side, different identities can share exactly the same attributes between non-makeup and makeup faces.

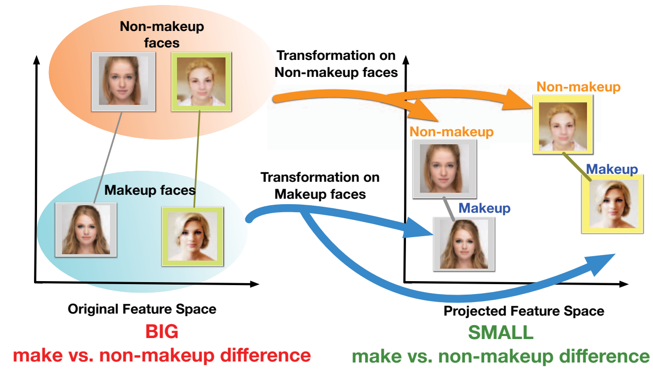

Fig. 3: The Framework. The makeup vs. non-makeup difference is reduced by learning transformations, regarding to makeup and non-makeup faces, respectively, such that the transformed data preserves the maximum discriminability and the minimum makeup vs. non-makeup difference at the same time. Faces within the same color frame indicate the same identity.

joint optimization framework is built, seeking transformations to achieve the maximum discriminability and the minimum differences between makeup vs. non-makeup faces. That is, the transformed data preserves the maximum dissimilarity from faces of different identities, and also minimizes the dissimilarity for faces with the same identity. Finally, coordinate descent method is adopted to solve the joint optimization problem. Fig.3 illustrates the framework visually.

2. METHOD

In this section, we present a joint optimization framework to incorporate the mid-level and low-level representations for makeup face recognition.

Problem Description. We begin with the definitions of terminologies. For clarity, the frequently used notations are summarized in Table. 1. Then the problem can be formulated as follows.

Table 1: Notations

Description	*Notation*
Data (non-makeup or makeup)	$* = \{n, m\}$
Number of samples	k^*
Attribute	$A^* = \{a_1^*, a_2^*, ..., a_k^*\}$
Low-level feature	$B^* = \{b_1^*, b_2^*, ..., b_k^*\}$
Data matrix	$X^* = \{x_1^*, x_2^*, ..., x_k^*\}$
Class label	$L^* = \{l_1^*, l_2^*, ..., l_k^*\}$
Transformation matrix	P^*

Problem 1. *Given labeled non-makeup data* $D^n = \{(x_1, y_1), ..., (x_{k^n}, y_{k^n})\}$ *and labeled makeup data* $D^m = \{(x_1, y_1), ..., (x_{k^m}, y_{k^m})\}$*, learn transformations* P^n *and* P^m *regarding to* X^n *and* X^m*, respectively, such that the transformed data preserves (1) the maximum distance between class, and (2) the minimum distance within class across non-makeup and makeup.*

2.1. Integration of representations at different levels

The integration of different representations seeks a rich representation robust to makeup changes and also powerful for discrimination between different identities. It is designed to minimize the visual differences between makeup and non-makeup faces, based on encoding the facial characteristics at different levels. For example, mid-level representations, e.g. attributes, can characterize the face images with semantic meanings across cosmetic changes. On the other hand, low-level features are still useful to characterize the identity of each individual face image. Thus, mid-level and low-level representations are complementary, and their combinations can deliver a richer encoding for face recognition with makeup changes. In our approach, a rich representation is obtained by integrating mid-level and low-level features. Specifically, raw pixels with PCA for dimensionality reduction are adopted as the low-level features. The mid-level features are composed by attributes, such as "big eyes", "small nose", etc. In this study, we use 28 face attributes as in [2], measured by the distance from the SVM hyperplane instead of binary values. These 28 face attributes are selected specially to encode makeup and non-makeup facial traits. Thus, the augmented data matrix X^* can be written as $X^* = [A^*, B^*]$, with each sample $x_i^* = [a_i^*, b_i^*], i = 1, ..., k^*$.

2.2. Learning transformations regarding to makeup and non-makeup faces

Transformations, regarding to makeup and non-makeup faces respectively, are jointly learned to maximize discriminability and minimize the difference between makeup and non-makeup faces. That is, maximizing between-class dissimilarity and minimizing within-class dissimilarity at the same time.

2.2.1. *Objective function*

We formulate an objective function for our approach as follows:

$$\min_{P^n,P^m}\{D^w(P^n,P^m)-\alpha D^b(P^n,P^m)+\beta R(P^n,P^m)\} \quad (1)$$

where $D^b(P^n, P^m)$ is the item of between-class dissimilarity constraint, while $D^w(P^n, P^m)$ is the within-class dissimilarity constraint. $R(P^n, P^m)$ is transformation constraint. $\alpha > 0$, $\beta > 0$ are trade-off parameters.

Between-class dissimilarity constraint: This cosntraint item is defined in such a way that the distances between classes are maximized across makeup and non-makeup faces. In this paper, we adopt a classical expression of the sum of squared distances, [1]

$$D^b(P^n,P^m) = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} I_{ij}^b \left\|(P^n)^T x_i^n - (P^m)^T x_j^m\right\|^2,$$
$$I_{ij}^b = \begin{cases} 1 & \text{if } l_i^n = l_j^m, \\ 0 & \text{if } l_i^n \neq l_j^m. \end{cases} \quad (2)$$

where I_{ij}^b indicates data across non-makeup face x_i^n and makeup face x_j^m, which are relevant or irrelevant regarding to the class label.

Within-class dissimilarity constraint: This constraint is defined in such a way that the distances between classes are maximized across makeup and non-makeup faces. The sum of squared distances is used:

$$D^w(P^n,P^m) = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} I_{ij}^w \left\|(P^n)^T x_i^n - (P^m)^T x_j^m\right\|^2,$$
$$I_{ij}^w = \begin{cases} 0 & \text{if } l_i^n = l_j^m, \\ 1 & \text{if } l_i^n \neq l_j^m. \end{cases} \quad (3)$$

[1] In this study, we assume the data preserves a pair-wise structure with equal numbers of non-makeup and makeup faces. For simplicity, here we let k be the number of samples.

where I_{ij}^w indicates data across non-makeup faces x_i^n and makeup faces x_j^m, which are relevant or irrelevant regarding to the class label.

Transformation constraint: This constraint ensures that the transformed data satisfy isotropic distribution.

$$R(P^n,P^m) = \frac{1}{2}(\left\|(P^n)^T X^n\right\|_F^2 + \left\|(P^m)^T X^m\right\|_F^2). \quad (4)$$

2.2.2. *Iterative Optimization*

Since the objective function in Eq.(1) is non-convex, it is difficult to get a global optimum. Practically, we propose to use a coordinate descent method [13] to jointly optimize each individual component, and iteratively update P^n or P^m with the other one fixed.

Before going into the iterative process, the objective function in Eq.(1) can be rewritten into a matrix form.

Let $Z^{n,w}$, $Z^{m,w}$, $Z^{n,b}$ and $Z^{m,b}$ be diagonal matrices with $Z_{ii}^{n,w} = \sum_{j=1}^k I_{ij}^w$, $Z_{ii}^{m,w} = \sum_{i=1}^k I_{ij}^w$, $Z_{ii}^{n,b} = \sum_{j=1}^k I_{ij}^b$, $Z_{ii}^{m,b} = \sum_{i=1}^k I_{ij}^b$.

Eq. (2), (3) and (4) can be written as

$$\begin{aligned} D^w(P^n,P^m) = &\frac{1}{2}[(P^n)^T X^n Z^{n,w}(X^n)^T P^n \\ &+ (P^m)^T X^m Z^{m,w}(X^m)^T P^m \\ &- 2(P^n)^T X^n I^w X^m (P^m)^T], \end{aligned} \quad (5)$$

$$\begin{aligned} D^b(P^n,P^m) = &\frac{1}{2}[(P^n)^T X^n Z^{n,b}(X^n)^T P^n \\ &+ (P^m)^T X^m Z^{m,b}(X^m)^T P^m \\ &- 2(P^n)^T X^n I^b X^m (P^m)^T], \end{aligned} \quad (6)$$

$$\begin{aligned} R(P^n,P^m) = &\frac{1}{2}[tr((X^n)^T P^n (P^n)^T X^n) \\ &+ tr((X^m)^T P^m (P^m)^T X^m)], \end{aligned} \quad (7)$$

where T denotes the matrix transpose.

Let $C(P^n, P^m)$ denote the cost funtion from Eq. (1) ,

$$\begin{aligned} C(P^n,P^m) = &D^w(P^n,P^m) - \alpha D^b(P^n,P^m) \\ &+ \beta R(P^n,P^m). \end{aligned} \quad (8)$$

Initialization. We initialize P^n and P^m by maximizing the sum of the between-class distance as

$$\max_{P^n,P^m} D^b(P^n,P^m), \quad \text{s.t. } D^w(P^n,P^m) = 1. \quad (9)$$

Eq. (9) can be solved as a generalized eigenvalue problem.

Algorithm 1 Iterative Optimization

INPUT: Training samples from *non-makeup faces* and *makeup faces*, i.e. data $\{X_i\}_{i=1}^k$, class labels $\{l_i\}_{i=1}^k$, trade-off parameters α and β.
OUTPUT: Learned transformation matrices P^n and P^m for non-makeup and makeup faces, respectively.
PROCEDURE:
1. *Initialization:* Initialize P^n and P^m such that

$$\max_{P^n,P^m} D^b(P^n, P^m),$$
$$\text{s.t. } D^w(P^n, P^m) = 1,$$

which can be optimized by a standard generalized eigenvalue problem.
2. **repeat**
3. *Update step for P^n:* Fix P^m then optimize P^n with Eq.(11)
4. *Update step for P^m:* Fix P^n then optimize P^m with Eq.(12)
5. **until** Convergence

Iterative update P^n and P^m. As there are two variables in the cost function, we adopt the coordinate descendent technique to achieve a local optimum.

(i) Fix P^m to update P^n. Take the derivative of Eq.(8) regarding to P^n and set it to zero, we have the following equation,

$$\begin{aligned}\frac{\partial C(P^n, P^m)}{\partial P^n} =& X^n Z^{n,w} (X^n)^T P^n - X^n I^w (X^m)^T P^m \\ & - \alpha X^n Z^{n,b} (X^n)^T P^n + \alpha X^n I^b (X^m)^T P^m \\ & + \beta X^n (X^n)^T P^n \\ =& 0.\end{aligned} \tag{10}$$

Then we obtain a closed form solution of P^n as [2]

$$\begin{aligned}P^n =& [X^n (Z^{n,w} - \alpha Z^{n,b} + \beta I)(X^n)^T]^{-1} \cdot \\ & X^n (I^w - \alpha I^b)(X^m)^T P^m.\end{aligned} \tag{11}$$

(ii) Fix P^n to update P^m. Similarly, take the derivative of Eq.(8) regarding to P^m and set it to zero, we can get

$$\begin{aligned}P^m =& [X^m (Z^{m,w} - \alpha Z^{m,b} + \beta I)(X^m)^T]^{-1} \cdot \\ & X^m (I^w - \alpha I^b)(X^n)^T P^n.\end{aligned} \tag{12}$$

Perform (i) and (ii) iteratively until convergence. A summary of the method is given in Algorithm 1.

3. EXPERIMENTS

The proposed approach is tested on the dataset collected in [2, 3]. The dataset contains 501 female subjects where each subject has a pair of face images with and without makeup. Four methods are selected for comparisons, namely, a direct match using the cosine distance, Dual-Attribute [2], CCA [3, 14] and PLS [3, 15]. In our implementation, all input images are gray-scale pixels, as skin color might change after makeup. Patch selection is proceeded to extract crucial regions that are helpful for face recognition. Similar to [3], eleven patches are selected from facial regions such as forehead, eyes, nose, cheeks, mouth and jaw. For a fair comparison, we use PCA to reduce the data dimension by preserving 95% of energy for all methods. Parameters are set empirically for the best performance. For our method, we set trade-off parameters as $\alpha = 0.8$, $\beta = 0.1$, and the number of iterations is set to 30. The dataset was split into training and testing sets. The training set contains 802 images from 401 subjects, where each subject contains an image without and with makeup respectively. Then the remaining 100 subjects go to the testing set. No subject overlapping occurs between training and testing set. The same verification protocol is adopted as [3] by generating 100 positive and 100 negative pairs of faces from the testing set. The negative pairs are produced by randomly picking up two different subjects in the testing set, and either the makeup or non-makeup face is selected to ensure that the matching is conducted between a pair of makeup and non-makeup faces. Verification accuracies are reported in Table 2. It shows that our approach outperforms others. We attribute our better performance to two factors: the integration of mid-level and low-level representations and the joint optimization framework. Mid-level representations capture invariant visual traits, while low-level representations provide a detailed description to enhance the discriminative power. Also, a joint optimization framework is built to minimize makeup and non-makeup differences.

Table 2: Verification Accuracy

Method	*Accuracy (%)*
Direct match	69.50
Dual-Attribute [2]	71.00
CCA [3]	72.50
PLS [3]	80.50
Ours	**82.50**

4. CONCLUSIONS

In this paper, we propose a joint optimization framework integrating mid-level and low-level representations for makeup face recognition problem. A good representation can be obtained which is robust across makeup and non-makeup faces, and transformations are learned to further minimize the makeup and non-makeup face difference. Experimental results demonstrate effectiveness of the proposed method.

[2] I indicates the identity matrix.

5. REFERENCES

[1] Cunjian Chen, Antitza Dantcheva, and Arun Ross, "Automatic facial makeup detection with application in face recognition," in *Conference on Biometrics International*. IEEE, 2013, pp. 1–8.

[2] Lingyun Wen and Guodong Guo, "Dual attributes for face verification robust to facial cosmetics," *Journal of Computer Vision and Image Processing*, vol. 3, no. 1, pp. 63–73, 2013.

[3] Guodong Guo, Lingyun Wen, and Shuicheng Yan, "Face authentication with makeup changes," *Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 814–825, 2014.

[4] Ali Moeini, Hossein Moeini, and Karim Faez, "Makeup-insensitive face recognition by facial depth reconstruction and gabor filter bank from women's real-world images," in *International Conference on Image Processing*. IEEE, 2014, pp. 308–312.

[5] Ali Moeini, Hossein Moeini, Fazael Ayatollahi, and Karim Faez, "Makeup-invariant face recognition by 3d face: Modeling and dual-tree complex wavelet transform from women's 2d real-world images," in *International Conference on Pattern Recognition*. IEEE, 2014, pp. 1710–1715.

[6] Hossein Moeini, Saeed Mozaffari, and Ali Moeini, "Makeup-invariant face recognition by combination of local binary pattern and dual-tree complex wavelet transform from women's images," in *International Symposium on Telecommunications*. IEEE, 2014, pp. 497–501.

[7] Junlin Hu, Yongxin Ge, Jiwen Lu, and Xin Feng, "Makeup-robust face verification," in *Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 2342–2346.

[8] Adriana Kovashka and Kristen Grauman, "Attribute adaptation for personalized image search," in *International Conference on Computer Vision*. IEEE, 2013, pp. 3432–3439.

[9] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar, "Attribute and simile classifiers for face verification," in *International Conference on Computer Vision*. IEEE, 2009, pp. 365–372.

[10] Mohammad Rastegari, Ali Farhadi, and David Forsyth, "Attribute discovery via predictable discriminative binary codes," in *Proceeding of the European Conference on Computer Vision*, pp. 876–889. Springer, 2012.

[11] Alireza Farhadi, Ian Endres, Derek Hoiem, and David Forsyth, "Describing objects by their attributes," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1778–1785.

[12] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 951–958.

[13] Paul Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[14] Harold Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[15] Abhishek Sharma and David W Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 593–600.