Neurocomputing 333 (2019) 339–350

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multiview discriminative marginal metric learning for makeup face verification

Lining Zhang[a,*], Hubert P.H. Shum[b], Li Liu[c], Guodong Guo[d], Ling Shao[c]

[a] Institute for Infocomm Research, Agency for Science and Technology Research, 1 Fusionolopis Way, #, 21-01, 138632, Singapore
[b] Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
[c] Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates
[d] Lane Department of Computer Science and Electrical Engineering, West Virgina University, Morgantown, WV 26506, USA

ARTICLE INFO

Article history:
Received 11 August 2018
Revised 1 December 2018
Accepted 8 December 2018
Available online 3 January 2019

Communicated by Dr. Ran He

Keywords:
Makeup face verification
Multiview learning
Distance metric learning
Marginal samples
Face verification

ABSTRACT

Makeup face verification in the wild is an important research problem for its popularization in real-world. However, little effort has been made to tackle it in computer vision. In this research, we first build a new database, i.e., Facial Beauty Database (FBD), which contains paired facial images of 8933 subjects without and with makeup in different real-world scenarios. To the best of our knowledge, FBD is the largest makeup face database to date compared with existing databases for facial makeup research. Moreover, we propose a new discriminative marginal metric learning (DMML) algorithm to deal with this problem in the wild. Inspired by the fact that interclass marginal faces are usually more discriminative than interclass nonmarginal faces in learning the discriminative metric space, we use the interclass marginal faces to depict the discriminative information. Simultaneously, we wish that those interclass marginal faces without makeup relations are separated from each other as far as possible, so that more discriminative information between facial images without and with makeup can be exploited for verification. Furthermore, since multiple features could provide comprehensive information in describing the facial representations from diverse points of view and extract more informative cues from facial images, we also introduce a multiview discriminative marginal metric learning (MDMML) algorithm by effectively learning a robust metric space such that multiple features from different points of view can be integrated to effectively enhance the performance of makeup face verification. Experimental results on two real-world makeup face databases are utilized to show the effectiveness of our method and the possibility of verifying the makeup relations from facial images in real-world.

1. Introduction

Human faces convey a number of important properties, such as emotion, gender, age, identity, expression, and ethnicity. During the past few years, a great number of facial image analysis methods were developed in both computer vision and computer security communities [1–4]. Typical applications include face recognition [1,5–11], facial age estimation [4,8,12–14], facial expression recognition [15–17], facial gender identification [18,19], facial sketch recognition [20,21], and human ethnicity recognition from facial images [22,23]. Although encouraging results have been obtained in the current research, most existing research only focus on facial image analysis under controlled conditions in real-world scenarios, where facial images are usually taken in uncontrolled settings [8,13,15]. In real-world applications, human subjects may wear some cosmetics to hide their facial flaws. Simultaneously, facial makeup can also make humans appear more attractive. The evidence of the effectiveness of using cosmetics for humans [24,25] have shown the improved attractiveness of humans when using cosmetics. As we can notice in Fig. 1, significant facial appearance difference can be observed for human subjects without and with makeup.

To develop effective facial image analysis methods that are robust to makeup changes, the system in real-world should address the influence caused by cosmetics. Wen et al. [26] tried to learn the attributes in makeup faces using the semantic attributes to reduce the influence of makeup on low-level visual features. Moreover, Chen et al. [27] preprocessed facial images with a self-quotient technique and reduced the cosmetic effects before matching two faces. Recently, some works have made efforts on robust

* Corresponding author.
E-mail address: lining.zhang@ieee.org (L. Zhang).

https://doi.org/10.1016/j.neucom.2018.12.003

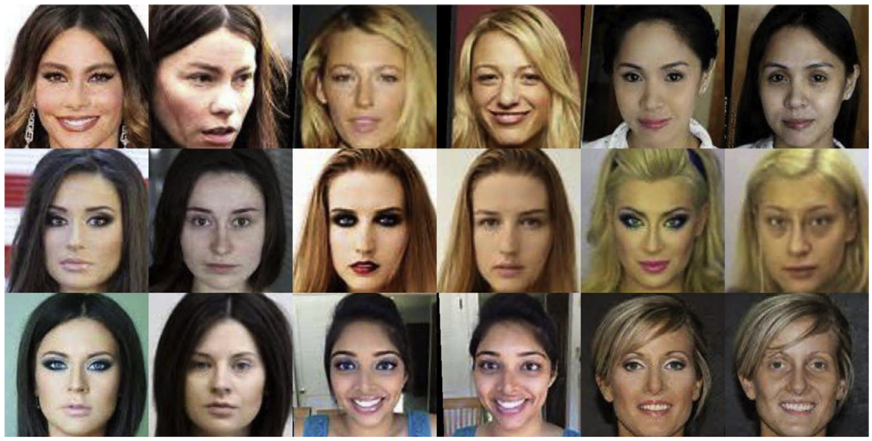


Fig. 1. An illustration of the facial images without and with makeup. It is observed that the facial images with makeup appear significantly different from the facial images without makeup for the same subject. Therefore, it is more useful to perform facial image verification that is robust and efficient to facial images with makeup.

face recognition with makeup changes [28]. However, the number of human subjects in their work is still quite limited (i.e., only a few human subjects in the makeup face databases) and the sizes of facial image databases are also small in terms of the number of human facial images without and with makeup [27–31].

In this paper, we investigate the makeup face verification problem in real-world scenarios. There are only a few efforts in tackling this challenging problem. Given facial images with makeup changes, makeup face verification aims to determine whether these two facial images are from the same subject. In this paper, we define the facial makeup relation as a relationship between two facial images for the same human subject without and with makeup. This new research direction has many potential real-world applications, e.g., social media analysis [32], face recognition [1] and public security [33]. However, limited research efforts have been carried out in this area, mainly due to the lack of such effective makeup face databases and intrinsic challenges of the makeup face verification problem. For this purpose, we build a novel makeup face database containing 17,866 facial images of 8933 subjects without and with makeup under uncontrolled conditions, which we named as Facial Beauty Database (FBD), to evaluate the effectiveness of makeup face verification in the wild. To the best of our knowledge, our FBD is the largest among existing makeup face databases for facial makeup research so far in the computer vision community. Then, we introduce a discriminative marginal metric learning (DMML) method to learn a robust metric space such that facial images with makeup relations are mapped closely and facial images without makeup relations are separated from each other as far as possible. Motivated by the issue that the interclass marginal samples without makeup relations are usually more discriminative than the interclass nonmarginal samples, we use the interclass marginal samples to depict the discriminative information in learning the distance metric space. Simultaneously, we wish those interclass marginal samples are pushed away as far as possible, so that more effective discriminative information can be exploited for verification. Since multiple feature representations could provide comprehensive information in characterizing human faces from different points of view and extract more descriptive features, we present a multiview discriminative marginal metric learning (MDMML) method to obtain a robust distance metric. Moreover, multiple feature representations can be effectively combined to enhance the makeup face verification performance. Experimental results on two real-world makeup face databases are utilized to show the possibility of verifying the makeup relation via facial images and the effectiveness of DMML and MDMML.

This paper is organized as follows: Section 2 briefly reviews the related works. Section 3 presents the proposed methods. Section 4 details the experimental results, and Section 6 concludes the work.

2. Related work

2.1. Makeup face verification

During the past few years, makeup face verification has been studied in both computational neuroscience, computer security and computer vision [26–29], and one interesting finding was noticed: human can easily recognize the makeup relation from facial images even if they are from unknown subjects in different scenarios. Motivated by the fact in computational neuroscience, computer vision researchers are aiming to develop computational approaches to verify the makeup relation from facial images, and there are a few attempts to address this challenging issue recently. Wen et al. proposed to learn facial attributes in facial images without and with makeup separately. In this work, face matching uses the semantic attributes to significantly reduce the influence of facial makeup on low-level features [26]. Moreover, [27] preprocessed facial images with a self-quotient technique and reduced the facial cosmetics efforts before face matching. Recently, some works have made efforts on robust face recognition with makeup changes [28]. However, the number of facial images is still quite limited (i.e., only a few subjects in [26–28]), and the databases are also small in terms of the number of facial makeup and nonmakeup images.

2.2. Makeup face databases

Most of existing makeup face databases contain only a few number of facial images without and with makeup in the computer vision community [27–31,31]. For instance, Guo et al. assembled a facial image database of 1002 faces with 501 pairs of female subjects, which mainly contains adult Asian or Caucasian women [29]. Hu et al. built the FAce Makeup (FAM) database of makeup face images, which are collected from the public figures or celebrities without and with makeup on the Internet [28]. FAM contains 519 subjects, 222 of them are male and the remaining 297 are female. In [34], a YouTube Makeup (YTM) database consisting of 99 subjects, specifically Caucasian females from YouTube makeup tutorials was assembled. The makeup in these facial images varies from subtle to heavy. Moreover, there are also some other makeup face databases (e.g., Virtual Makeup database [34], Makeup In the

Table 1
Comparison of our FBD and other existing makeup face databases. Our FBD offers the largest number of subjects and facial images without and with makeup compared with other existing makeup face databases.

Properties	Guo [29]	FAM [28]	YMU [34]	VMU [34]	MIW [30]	Concordia [31]	FBD
# Images	1002	1038	604	204	154	1290	**17,866**
# Subjects	501	519	151	51	125	21	**8933**
# Male	0	222	0	0	0	0	**112**
# Female	501	297	151	51	125	21	**8821**

Wild (MIW) database [30] and Concordia database [31], etc.) in the current computer vision research. However, most of the existing makeup face databases contain a small number of human subjects, and even the largest one merely provides 1290 makeup face images.

Discriminative metric learning has attracted much attention during the past few years, and there have been many effective methods proposed previously. Typical methods include neighborhood component analysis (NCA) [35], marginal Fisher analysis (MFA) [36], cosine similarity metric learning (CSML) [37], large margin nearest neighbor (LMNN) [38], conjunctive patches subspace learning [39], and information theoretic metric learning (ITML) [40].

Although these algorithms obtained excellent performance in different computer vision applications, they have some intrinsic disadvantages: (1) Interclass marginal samples of different classes are more discriminative than interclass nonmarginal samples in learning the discriminative metric [36], however, most existing discriminative metric learning methods treat the interclass samples of different classes equally and overlook the significantly different contributions in learning the discriminative metric space, and thus the discriminative information conveyed by the interclass samples is ignored. (2) Previous research in computer vision has demonstrated that different features can provide different descriptive information in characterizing the visual information from different points of view. However, most of the existing metric learning methods make an assumption that the samples are extracted from a single view space and will not be able to handle multiview data directly. Therefore, it is urgent to design effective multiview metric learning methods to deal with the data from different views for real-world applications [41,42].

3. Makeup face database

3.1. FBD

Aiming to collect a large-scale and comprehensive makeup face database, in this work, we firstly crawl facial images from various image search engines by using a number of keywords to describe various facial makeup scenarios. To make our large makeup face database more generalized, we do not include keywords referring to specific subjects. Instead, we only use general keywords that can describe the scenarios such as “face before and after makeup”. Besides, we also select several implicit descriptions related to makeup (e.g., “beauty with cosmetics”) as the keywords for image searching. Afterwards, the gathered keywords are used to search for images from several public image search engines including Google Image Search,[1] Yahoo Image Search[2] and Flickr Image Search.[3] To control the facial image quality, we remove images of synthetic faces, tiny faces and unclear faces. Moreover, to obtain cleaner facial images, we further apply VJ face detector to localize the facial areas and crop them out [43]. Finally, the makeup face database contains the cropped 17,866 facial images with 8933 subjects and is termed as FBD. In FBD, we use some similar images to simulate the same persons of different situations and periods in real world applications. FBD offers a superiorly comprehensive database for facial makeup research. In the near future, we plan to release our FBD for further research. The detailed comparison of our FBD and existing makeup face databases is illustrated in Table 1.

3.2. Face representations

To represent the facial images, we used the following four feature representations for makeup face verification, i.e., local binary patterns (LBP) [44], histogram of oriented gradients (HOG) [45], scale-invariant feature transform (SIFT) [46], three-patch LBP (TPLBP) [47]. The reason we selected these features is that they have shown good performance in recent kinship verification research [48–50]. In this work, we followed the same parameter settings for the features in [48] so that a fair comparison can be obtained. In the following, we will detail each feature representations of facial images.

LBP: For each facial image, we partition it into 4×4 nonoverlapped blocks with the size 16×16. For each individual block, we extract a 256-dimensional histogram feature to describe the block. Finally, all of the feature representations are concatenated into a 4096-dimensional features to represent each face image.

HOG: Each facial image is participated into 16×16 blocks with size 4×4. Then, each individual block was evenly divided into 8×8 blocks with size 8×8 again. Finally, we extract a 9-dimensional HOG feature descriptor for each block and concatenate them to form 2880-dimensional features to represent each face image.

SIFT: We partition each facial image into several overlapping blocks, and then extracted the features from each individual block. In this work, the block size is set as 16×16 and the overlapping scale is 8. Finally, there are 49 blocks for the whole facial image and each facial image can be represented as a 6272-dimensional feature in the high-dimensional space.

TPLBP: Each facial image has been partitioned into 4×4 nonoverlapping blocks with scale 16×16. The 3×3 block centered on the pixel and 8 blocks located uniformly in the ring of radium around this block were considered in experiments. For each block, we can extract 256-dimensional histogram feature and each facial image can be represented as a 4096-dimensional feature vector.

3.2.1. Data preparations

In experiments, we first applied the VJ face detector to well localize the facial areas [43] of FBD, and cropped and aligned the facial areas into 64×64 size according to Shan et al. [51]. Finally, the nonfacial areas were carefully removed and the facial areas will be utilized for makeup face verification. For each facial image, we used the histogram equalization technique to smooth the aligned images of FBD. We converted the color facial images into grey ones. Some aligned facial images without and with makeup of FBD are illustrated in Fig. 3.

[1] http://images.google.com.
[2] http://images.search.yahoo.com.
[3] http://www.flickr.com.

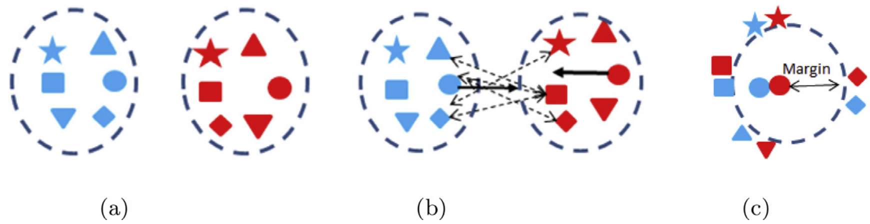


Fig. 2. An illustration of DMML. (a) The facial images without and with makeup in the original high-dimensional visual feature space. The blue data in left denote the facial images without makeup, and the red data in right denote the facial images with makeup. The objective of DMML is to learn a robust metric space so that two facial images with makeup relations are pulled closely and those facial images without makeup relations can be separated as far as possible. (b) The basic principle of DMML. (c) The expected distributions of facial images without and with makeup in the new metric space, where the distances of makeup face images of same subjects are decreased and those facial images of different subjects are separated as much as possible, respectively.

Fig. 3. An illustration of the facial images of FBD and two neighboring images along each row are facial images with makeup and without makeup of the same subject, respectively.

4. The proposed approach

4.1. Basic principle

To well describe our methods, Fig. 2 illustrates the basic principle of DMML. In Fig. 2(a), there are two sets of facial samples, where the samples in the left denote the facial images without makeup and those in the right denote the facial images with makeup, respectively. The facial images without and with makeup of different subjects are indicated by corresponding circles, five-pointed star, squares and triangles in Fig. 2. In the high-dimensional space, there is always a large difference between facial images without makeup and facial images with makeup due to the color and texture variations. Hence, the facial images of different subjects are usually misclassified in real-world applications. From the classification point of view, the interclass marginal samples are more discriminative than those interclass nonmarginal samples in learning the discriminative metric space. Motivated by Yan et al. [36], we are interested in learning an effective metric space such that the facial images with makeup relations are pulled closely and the interclass marginal samples without makeup relations can be separated as far as possible, as indicated in Fig. 2(b) and (c). That is to say, the similarities of makeup face images of different subjects should be significantly decreased so that the interclass margin between different subjects in the new metric space will be increased and more discriminative information could be utilized for makeup face verification.

4.2. DMML

Let $F = \{(x_i, y_i)|i = 1, 2, \ldots, n\}$ denote the training set of n pairs of facial images,ã;;where x_i and y_i are the ith the facial image without makeup and the facial image with makeup in a h-dimensional space. Our DMML method aims to find an effective Mahalanobis distance metric M so that the distances between the facial samples x_i and $y_j (i = j)$ are as small as possible, and those facial images between x_i and $y_j(i \neq j)$ are as large as possible simultaneously, and

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^T M (x_i - y_j)}, \tag{1}$$

where M is an $h \times h$ square matrix and $1 \leq i,\ j \leq n$. Moreover, the distance metric d should be symmetric, nonnegative, and triangularly unequal.

To learn an effective distance metric, we formulate our proposed DMML method as follows:

$$\begin{aligned}\min_M f(M) &= f_1(M) - f_2(M) - f_3(M) \\ &= \frac{1}{n}\sum_{i=1}^{n} d^2(x_i, y_i) - \frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k} d^2(x_i, y_{i_j}) \\ &\quad - \frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k} d^2(x_{i_j}, y_i)\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^T M(x_i - y_i)$$
$$- \frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_i - y_{i_j})^T M(x_i - y_{i_j})$$
$$- \frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{i_j} - y_i)^T M(x_{i_j} - y_i), \quad (2)$$

where y_{i_j} represents the jth interclass marginal samples of x_i and x_{i_j} denotes the jth interclass marginal samples of y_i, respectively. The objective of f_1 is to ensure that x_i and y_i are pulled closely in the new metric space since the two facial images have the makeup relation. f_2 aims at ensuring that if y_{i_j} is one of the interclass marginal samples of x_i, then they will be pushed away from each other as much as possible in the new metric space. In the same way, f_3 ensures that if x_{i_j} is one of the interclass marginal samples of y_i, they should also be separated from each other as much as possible.

Since the distance metric M is usually symmetric and positive semidefinite, we instead find a non-square transformation matrix W with size $h \times l (l \le h)$, thus

$$M = WW^T, \quad (3)$$

Then, we can reformulate Eq. (1) as the following optimization problem, i.e.,

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^T M(x_i - y_j)}$$
$$= \sqrt{(x_i - y_j)^T WW^T(x_i - y_j)}$$
$$= \sqrt{(s_i - t_j)^T(s_i - t_j)}, \quad (4)$$

where $s_i = W^T x_i$ and $t_j = W^T y_j$.

Thus, we can simplify $f_1(M)$ as the following form as

$$f_1(M) = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^T M(x_i - y_i)$$
$$= tr\left(W^T \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)(x_i - y_i)^T W\right)$$
$$= tr\left(W^T D_1 W\right), \quad (5)$$

where $D_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)(x_i - y_i)^T$, W means the transformation matrix. Similarly, $f_2(M)$ and $f_3(M)$ can be simplified as

$$f_2(M) = tr\left(W^T \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_i - y_{i_j})(x_i - y_{i_j})^T W\right)$$
$$= tr\left(W^T D_2 W\right), \quad (6)$$

$$f_3(M) = tr\left(W^T \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{i_j} - y_i)(x_{i_j} - y_i)^T W\right)$$
$$= tr\left(W^T D_3 W\right), \quad (7)$$

where D_2 and D_3 are defined as $\frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_i - y_{i_j})(x_i - y_{i_j})^T$ and $\frac{1}{n_k}\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{i_j} - y_i)(x_{i_j} - y_i)^T$, respectively.

Thus, we can simplify the DMML method as follows:

$$\min_W f(W) = tr\left(W^T(D_1 - D_2 - D_3)W\right)$$
$$s.t. W^T W = I, \quad (8)$$

where the constraint $W^T W = I$ is used to remove an arbitrary scaling factor in the projection. Therefore, W can be effectively solved as a standard eigenvalue decomposition problem, i.e.,

$$(D_1 - D_2 - D_3)w = \lambda w. \quad (9)$$

Let us define $w_1, w_2, \ldots, w_l$ as the eigenvectors corresponding to the l smallest eigenvalues ordered by $\lambda_1 \le \lambda_2 \le \lambda_3 \ldots \le \lambda_l$. The $h \times l$ transformation matrix W can be obtained to project the facial images x_i without makeup and the facial images with makeup y_i into a low-dimensional feature space s_i and t_i as follows:

$$s_i = W^T x_i, t_i = W^T y_i, i = 1, 2, \ldots, n. \quad (10)$$

4.3. MDMML

In computer vision, previous research efforts have indicated that different feature representations can provide comprehensive information in describing the facial features from different points of view. Therefore, we expect to utilize multiple feature representations from different points of view for the makeup face verification problem [52,53]. The problem is that, most conventional metric learning approaches cannot be directly applied to multiview data in real-world applications [35,37,38,40] due to the complexity of the visual features from different points of view. One possible solution to this problem is to concatenate multiple feature representations together as a single feature vector in the high dimensional space and then apply conventional metric learning approaches for real-world applications. However, the concatenation of multiple features is usually not reasonable since different feature representations usually carry different statistical characteristics. This operation ignores the diversity of different features from different points of view, which thus cannot efficiently utilize the comprehensive information conveyed by different feature representations. Therefore, we introduce a new multiview DMML method to learn a robust metric space for measuring the similarity of multiple feature representations of facial images.

Here, suppose we have v different views of feature representations, and $F^t = \{(x_i^t, y_i^t) | i = 1, 2, \ldots, n\}$ is the feature representation of the tth view of facial images with n pairs, where $x_i^t \in R^h$ and $y_i^t \in R^h$ are the ith facial images without makeup and the facial images with makeup from the tth view, respectively, and $t = 1, 2, \ldots, v$. MDMML aims to find a distance metric d such that the distances between the facial images x_i^t and $y_i^t (i = j)$ are as small as possible, and those between x_i^t and $y_j^t (i \ne j)$ are as large as possible.

Aiming to exploit the complementary information of different feature representations, a number of nonnegative parameters $\alpha = [\alpha_1, \ldots, \alpha_v]$ are imposed on the objective function of DMML for each view. Generally, the larger α_i is, the more important role the view x_i^t plays in learning the low-dimensional transformation matrix W. Here, we can generally formulate MDMML as the following constrained optimization problem:

$$\min_{W,\alpha} \sum_{t=1}^{v} \alpha_t tr\left(W^T(D_1^t - D_2^t - D_3^t)W\right)$$
$$s.t. W^T W = I, \sum_{t=1}^{v} \alpha_t = 1, \alpha_t \ge 0. \quad (11)$$

The solution to Eq. (11) is $\alpha_t = 1$ corresponding to $\min tr\left(W^T(D_1^t - D_2^t - D_3^t)W\right)$ over different views, and $\alpha_t = 0$ otherwise. This indicates that only one kind of feature representation from one view can be selected by using this solution. Thus, the performance of this solution can be equivalent to using the one from the best view, in which different information of facial image feature representations from diverse views has not been exploited. It is not appropriate to only select the best view in real-world applications. Motivated by the authors in [52,53], we

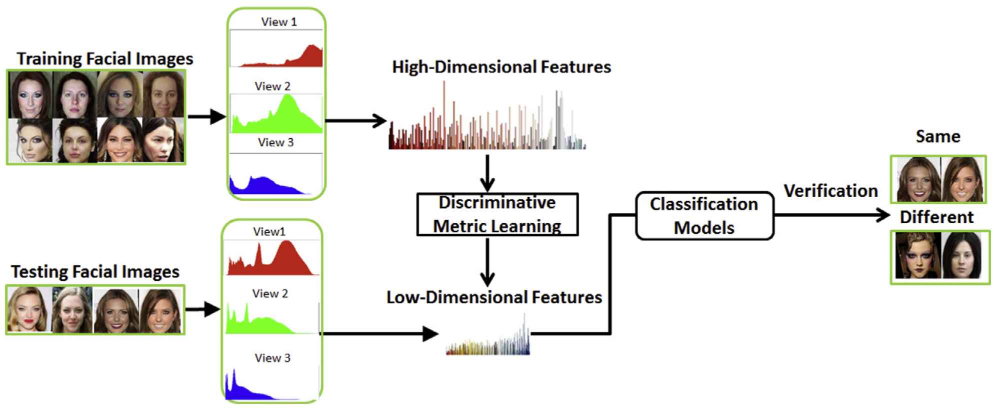


Fig. 4. The framework of our makeup face verification system. Suppose we have a number of training facial images, our system can extract different visual feature representations from different views and then learn a discriminative metric such that the facial images with makeup relations can be pulled closely and the facial images without makeup relations are separated from each other as far as possible. For each testing facial image pair, the system can also extract the same visual feature representations from different views and then project the high-dimensional features into the low-dimensional space. A conventional classification model can be used to verify whether the testing facial image pairs are from the same subject or not.

modify α_t to be α_t^r with $r > 1$, so that each feature representation from different views will have an individual contribution to the final transformation matrix W. Finally, this problem can be reformulated as follows:

$$\min_{W,\alpha} \sum_{t=1}^{v} \alpha_t^r tr\left(W^T (D_1^t - D_2^t - D_3^t) W\right)$$
$$s.t. W^T W = I, \sum_{t=1}^{v} \alpha_t = 1, \alpha_t \geq 0. \quad (12)$$

To the best of our knowledge, there is no effective method to find the optimal solution to Eq. (12), which is a nonlinear constrained nonconvex problem. Motivated by the authors in [52,53], here, we introduce an iterative algorithm by using the alternating computation to calculate a locally optimal solution. The algorithm can update W and α iteratively.

Firstly, we need to fix W and update α. By introducing a Lagrange multiplier λ to formulate the constraint $\sum_{t=1}^{v} \alpha_t = 1$ together, we can get the new objective function as follows:

$$L(\alpha, \lambda) = \sum_{t=1}^{v} \alpha_t^r tr\left(W^T \left(D_1^t - D_2^t - D_3^t\right) W\right) - \lambda \left(\sum_{t=1}^{v} \alpha_t - 1\right). \quad (13)$$

Let $\frac{\partial L(\alpha,\lambda)}{\partial \alpha_t} = 0$ and $\frac{\partial L(\alpha,\lambda)}{\partial \lambda} = 0$, we have

$$\begin{cases} r\alpha_t^{r-1} tr\left(W^T \left(D_1^t - D_2^t - D_3^t\right) W\right) - \lambda = 0 \\ \sum_{t=1}^{v} \alpha_t - 1 = 0. \end{cases} \quad (14)$$

Finally, we can obtain α_t as follows:

$$\alpha_t = \frac{\left(1/tr\left(W^T \left(D_1^t - D_2^t - D_3^t\right) W\right)\right)^{1/(r-1)}}{\sum_{t=1}^{v} \left(1/tr\left(W^T \left(D_1^t - D_2^t - D_3^t\right) W\right)\right)^{1/(r-1)}}. \quad (15)$$

By using the new α, we can update W. The optimization problem in Eq. (11) can be reformulated as:

$$\max_{W} tr\left(W^T \left(\sum_{t=1}^{v} \alpha_t^r \left(D_1^t - D_2^t - D_3^t\right)\right) W\right)$$
$$s.t. W^T W = I. \quad (16)$$

Thus, W can be easily calculated by solving the eigenvalue decomposition problem in the following:

$$\left(\sum_{t=1}^{v} \alpha_t^r \left(D_1^t - D_2^t - D_3^t\right)\right) w = \lambda w. \quad (17)$$

The proposed MDMML algorithm can be summarized as Algorithm 1.

Algorithm 1 MDMML.

Input: The tth view of n pairs of facial images without and with makeup; the number of interclass samples k; the maximum number of iteration T, and the error of convergence ε.
Output: The final metric W.
Step 1: Set $\alpha = [1/v, 1/v, \ldots, 1/v]$ and calculate W^0 by using Eq. (17).
Step 2: Checking procedure to find the metric space W.
For $i = 1, \ldots, T$
2.1 Compute α as in Eq.(15);
2.2 Calculate W^i by using Eq. (17);
2.3 If $\left|W^i - W^{i-1}\right| < \varepsilon$, then Step 3.
End
Step 3: Output the best metric $W = W^i$.

5. Experiments

In this section, we first design a makeup face verification system for makeup face verification based on two makeup face databases (i.e., FAM and FBD) and provide the baseline results for other researchers to compare their methods with ours. Moreover, we have also evaluated DMML and MDMML by conducting comprehensive makeup face verification experiments based on these two databases.

5.1. Makeup face verification system

In this subsection, we give an overview of our makeup face verification system. As shown in Fig. 4, given a set of training facial images without and with makeup, the visual feature representations from different views (i.e., LBP, HOG, SIFT, etc.) are first extracted and constructed in the high-dimensional feature space. Then, a discriminative metric space can be learned in which the facial images with makeup relations are pulled closely. Meanwhile, the facial images without makeup relations are separated from each other further. Finally, a classification model is used to divide the feature space into two classes, i.e., one for the same subject pairs and the other is for different subject pairs.

Table 2
The classification accuracy (Percent), AUC (Percent) and EER (Percent) of four feature representations (i.e., LBP, HOG, SIFT, TPLBP) on FAM.

Features	Dimension	Accuracy	AUC	EER
LBP	3776	75.8	76.3	30.5
HOG	1764	70.6	71.5	32.6
SIFT	6272	78.0	79.2	27.3
TPLBP	4096	76.2	77.1	29.8

Table 3
The classification accuracy (Percent), AUC (Percent) and EER (Percent) of four feature representations (i.e., LBP, HOG, SIFT, TPLBP) on FBD.

Features	Dimension	Accuracy	AUC	EER
LBP	3776	70.3	71.4	35.3
HOG	1764	65.2	66.3	37.5
SIFT	6272	72.8	73.7	30.7
TPLBP	4096	71.8	72.9	32.9

Table 4
The verification accuracy (*percent*) of different metric learning methods on FAM.

Features	NCA	CSML	LMNN	DMML	MDMML
LBP	73.2	72.5	74.6	75.3	79.5
HOG	68.7	67.6	69.8	70.5	
SIFT	76.4	77.5	77.9	78.2	
TPLBP	74.3	73.5	75.8	77.2	

Table 5
The verification accuracy (*percent*) of different metric learning methods on FBD.

Features	NCA	CSML	LMNN	DMML	MDMML
LBP	68.5	67.4	69.7	71.5	73.2
HOG	63.7	62.9	64.6	65.3	
SIFT	70.5	71.4	72.1	72.8	
TPLBP	69.8	67.4	70.5	71.6	

5.2. Experimental setup

5.2.1. The classification model

Since the makeup face verification is a binary classification problem, we apply the conventional support vector machine (SVM) for verification. In experiments, we use the conventional linear kernel as the similarity metric of each pair of facial images for its effectiveness in high dimensional feature space [54]. We apply five-fold cross validation on training facial images to find the optimal parameters. Specifically, we divide the training facial images into five folds, and each one will have 20% of the facial images with makeup relations. Moreover, we use four folds to train the SVM classification model, and then we utilize the remaining fold to determine the parameters of SVM.

5.2.2. Experimental protocol

In real-world scenarios, we expect our makeup face verification system can effectively verify whether there are makeup relations for new pairs of facial images without redesigning the system. Thus, we introduce the open set protocol for facial makeup relation [55], which is widely used in face verification experiments. In experiments, we used five-fold cross validation to evaluate the system on the makeup face databases. Specifically, each subset of FAM and FBD were equally partitioned into five small folds and each fold has around 20% of facial images with makeup relations. Here, we consider all face pairs with makeup relations as positive face pairs, and facial images without makeup relations as negative face pairs. In the experiments, the positive face pairs are true pairs of facial images (i.e., two images of the same subject: one without makeup and the other one with makeup), and the negative face pairs are false pairs of facial images (i.e., two images of different subjects: one without makeup and the other one with makeup). The size of positive face pairs is usually much smaller than that of negative face pairs. In the experiments, the facial image without makeup was randomly paired with a facial image with makeup to construct the negative face pairs. Moreover, we have to make such that each face image without and with makeup appears only once in negative face pairs. Therefore, the size of positive face pairs and that of negative face pairs will be equal to train a classifier.

5.3. Experimental results

5.3.1. Analysis on different feature representations

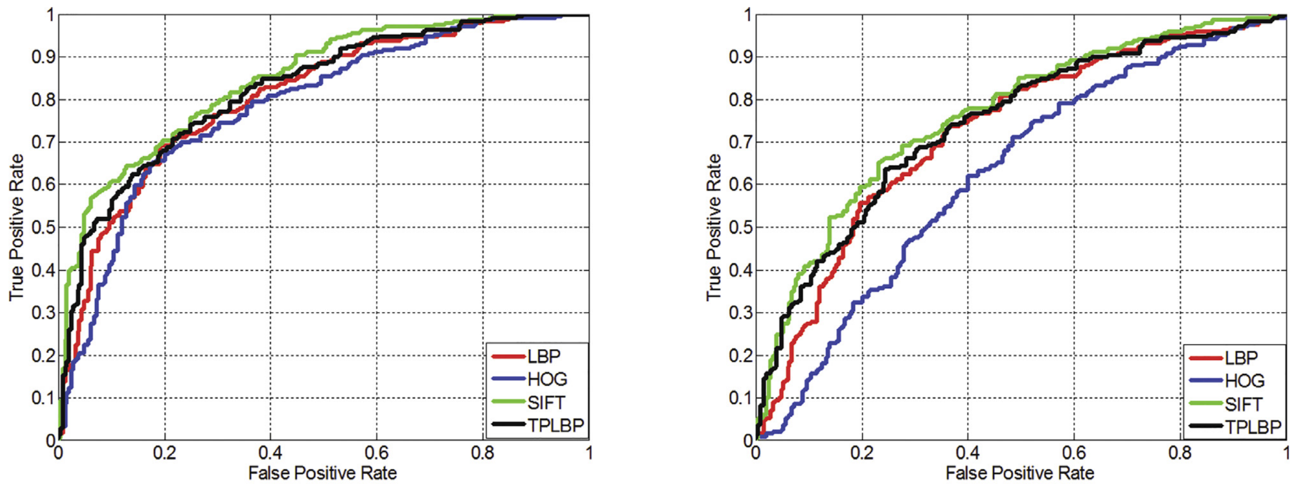
For testing face pairs, we aim to verify whether they are the same subject or not by using the SVM classification model. We use the classification accuracy rate, the area under the ROC curve (AUC), the equal error rate (EER) to evaluate the performance of the system. The accuracy rate is defined as n_c/n_t, where n_t is the size of the whole testing face pairs and n_c is the size of testing facial image pairs with correct classification. The classification accuracy rate, AUC and EER of different feature representations on the two makeup face databases are shown in Tables 2 and 3. These two tables illustrate the best feature representations for makeup face verification tasks. As shown in Tables 2 and 3, the best feature representations for makeup face verification on FAM and FBD is SIFT, which can significantly outperform the other feature representations, i.e., LBP, HOG and TPLBP.

Moreover, to visualize the performance difference of the four feature representations, we also show the receiver operating characteristic (ROC) curves of the four feature representations in Fig. 5, and Figs. 5(a) and 5 (b) plot the ROC curves of these feature representations on FAM and FBD, respectively. We notice from the experimental results that the SIFT feature representation can outperform other visual features on both FAM and FBD in terms of the ROC curves.

5.3.2. Comparisons with existing discriminative metric learning methods

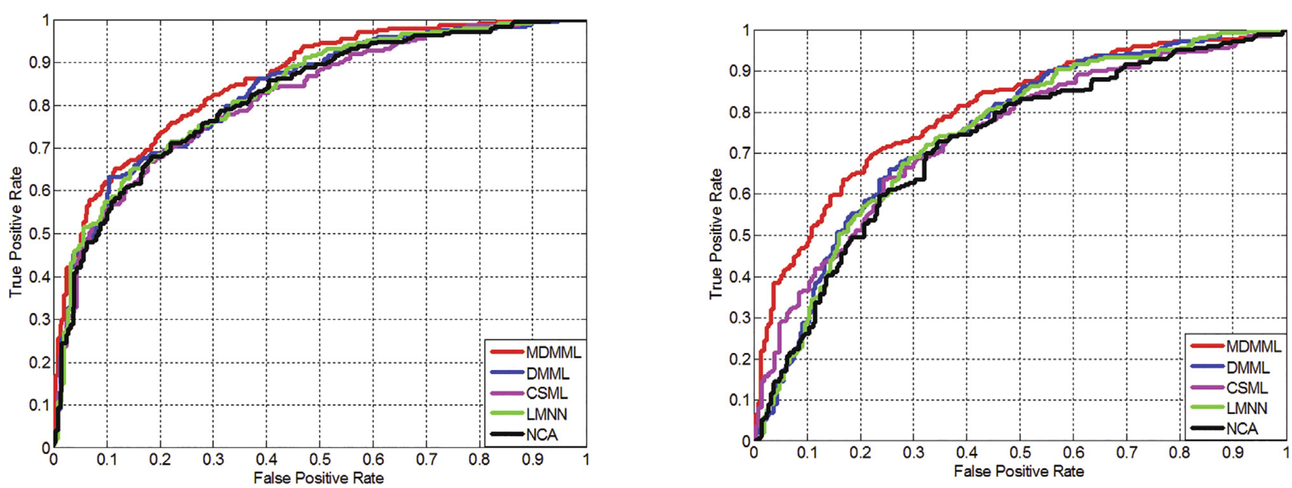
In this subsection, we have compared the proposed methods with three representative metric learning methods, which could also be used for makeup face verification in the wild, i.e., CSML [37], NCA [56], LMNN [38]. The number of nearest neighbor samples is set as 5 for these methods. Here, we empirically set the feature dimension of our proposed methods as 30 and 40, respectively.

The makeup face verification rate of compared methods with the different feature representations on FAM and FBD are shown in Tables 4 and 5, respectively. As shown in the two tables, our proposed DMML and MDMML methods can constantly outperform the other compared methods, i.e., CSML, NCA, and LMNN. We notice that DMML can outperform the other compared algorithms in all experiments on the two databases, which means that learning a distance metric by using the interclass marginal samples can give more discriminative information for makeup face verification tasks. In DMML, the interclass marginal samples with makeup relations are fully utilized to exploit the discriminative information within different classes for the verification tasks. CSML uses cosine metric to measure the similarity between two samples, which may not work well for LBP and TPLBP features in face verification tasks [37]. Although NCA and LMNN also utilize the neighborhood samples to describe the discriminative information of different classes, their



(a) (b)

Fig. 5. The ROC curves of the four feature representations (i.e., LBP, HOG, SIFT, TPLBP) obtained on (a) FAM and (b) FBD, respectively.



(a) (b)

Fig. 6. The ROC curves of different metric learning methods obtained on FAM and FBD, respectively.

solutions usually do not perform well in a high dimensional feature space [38,56].

Moreover, MDMML can obtain better performance than DMML for makeup face verification as shown in Tables 4 and 5. This is mainly because MDMML can effectively utilize multiple feature representations of facial images in a commonly metric space so that more describable information can be exploited for makeup face verification tasks.

The experimental results on the FAM database are generally higher than those obtained on FBD, which means the makeup face verification relation on FBD is more difficult than that on FAM. This is mainly because that the facial images in the FAM database are captured under controlled nature and posed restrictions on the variations other than makeup. On the other hand, the facial images in FBD are collected from different scenarios in real-world Internet search. Moreover, the size of FBD is much larger than that of FAM.

Moreover, we have also plotted the ROC curves of compared methods in Fig. 6, where Figs. 6(a) and (b) plotted the ROC curves of experimental results on FAM and FBD, respectively. It should be noted that in experiments, the CSML, NCA, LMNN and DMML methods use the SIFT feature because it can achieve better performance compared with other feature representations. As shown in Fig. 6, our methods show a much better performance than other compared methods in terms of the ROC curves.

5.3.3. Comparisons with existing multiview learning methods

In this subsection, we will compare the MDMML method with two popular methods for multiview learning. One is the multiview learning method for dimension reduction, i.e., multiview spectral embedding (MSE), which has extended the conventional spectral embedding method for multiview data [52]; the other one is the multiple kernel learning (MKL), which constructs multiple kernels for multiple features to describe the data complementarily [57].

Fig. 7 illustrates the mean accuracy of compared methods on FAM and FBD. We notice that our MDMML method can obtain a much better performance than the other two multiview learning methods. The reason is that our method uses the interclass

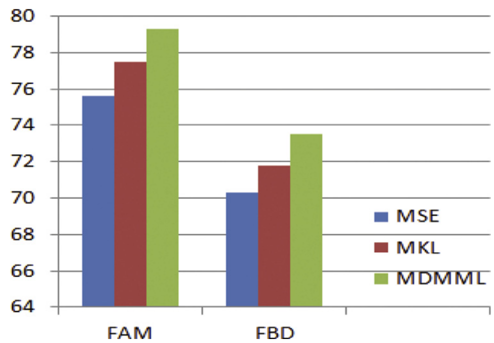


Fig. 7. The verification accuracy (*Percent*) of compared Methods on FAM and FBD.

Table 6
The verification accuracy (*Percent*) of different classification models on FAM.

Methods	Features	NN	KNN	SVM
DMML	LBP	73.5	74.7	75.8
	HOG	69.5	71.3	70.6
	SIFT	77.5	75.9	78.0
	TPLBP	76.1	75.4	76.2
MDMML	ALL	78.7	77.5	79.6

marginal samples to learn the robust metric space while other methods do not sufficiently utilize this discriminative information.

5.3.4. Comparisons with different classification models

To further evaluate the effectiveness of DMML and MDMML, we also compare our proposed methods with different classification models in makeup face verification task. Generally, besides the conventional SVM model, we also employ two widely used classification models, i.e., the nearest neighbor (NN) classifier and the k-nearest neighbor (kNN) classifier. These classification models are widely used in previous face recognition and verification tasks [5,36,48,49]. In experiments, we empirically set the number of the nearest neighborhood in the KNN classifier as 7. In NN, the new sample is classified by calculating the distance to the nearest training samples. The label of the new sample is determined by the nearest training data. The mean accuracy of DMML and MDMML for makeup face verification with different classifiers on FAM and FBD are shown in Tables 6 and 7, respectively. In experiments, we can notice that the classification models have some effect on

Table 7
The verification accuracy (*Percent*) of different classification models on FBD.

Methods	Features	SVM	NN	KNN
DMML	LBP	69.5	70.1	70.3
	LE	63.4	64.8	65.2
	SIFT	70.5	71.3	72.8
	TPLBP	69.8	70.6	71.8
MDMML	ALL	72.6	72.5	73.7

the experimental results and different classification models can get similar performance, which can effectively show the robustness of DMML and MDMML for makeup face verification tasks.

5.3.5. The sensitivity of parameters

In this subsection, we investigate the effect of the number of interclass marginal samples k in DMML and MDMML. Fig. 8 gives the mean verification accuracy of DMML and MDMML in experiments, where Fig. 8(a) and (b) are the experimental results acquired on FAM and FBD, respectively. Here, one can see our DMML and MDMML can achieve the best performance when k is determined as 5 for both DMML and MDMML. Moreover, we can also observe that DMML and MDMML can show stable verification performance for different numbers of nearest neighborhood samples. Hence, it is easy to select an appropriate number of nearest samples for DMML and MDMML to obtain good performance for makeup face verification.

To evaluate the effect of different feature dimensions of both DMML and MDMML, Fig. 9 shows the mean verification accuracy of DMML and MDMML versus the feature dimensions of our methods, where Fig. 9(a) and (b) are the experimental results achieved on FAM and FBD, respectively. It is clearly shown that both DMML and MDMML get stable performance when the dimension of the feature is larger than 30 for DMML and 35 for MDMML, respectively.

5.3.6. Visualization of makeup face verification

To gain further insight into the challenges of FBD and the limitations of our methods, we illustrate the most confident predictions on FBD made by our MDMML method. Fig. 10 presents the most confident incorrect matching. These images demonstrate the challenges and complexities of makeup face verification under different scenarios. Many mistakes result from the misleading context

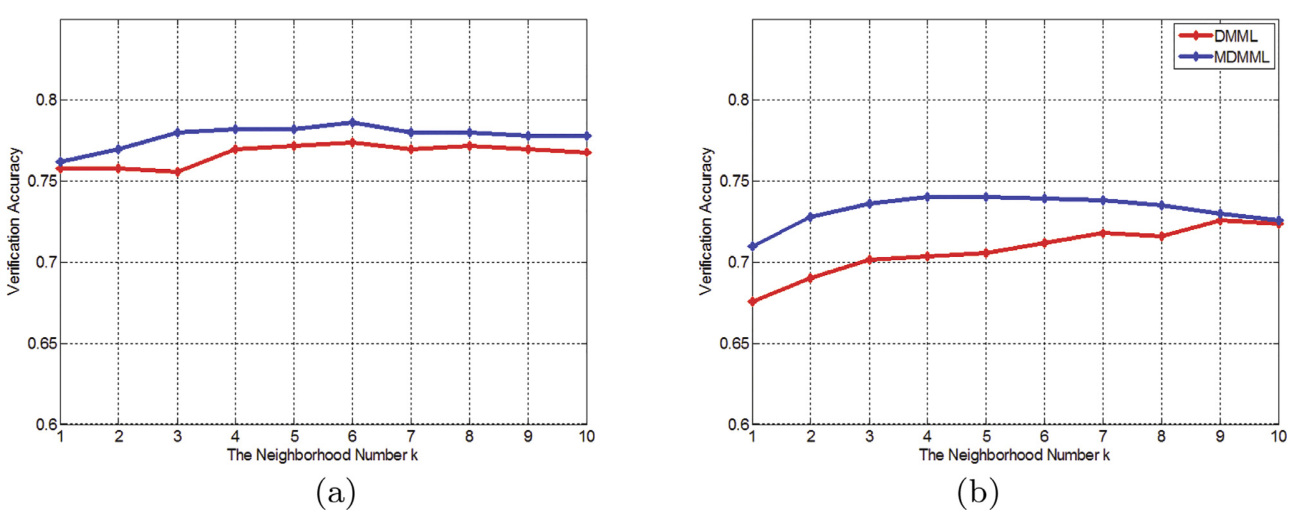


Fig. 8. The mean verification accuracy of DMML and MDMML versus the numbers of nearest neighborhood samples k on (a) FAM and (b) FBD.

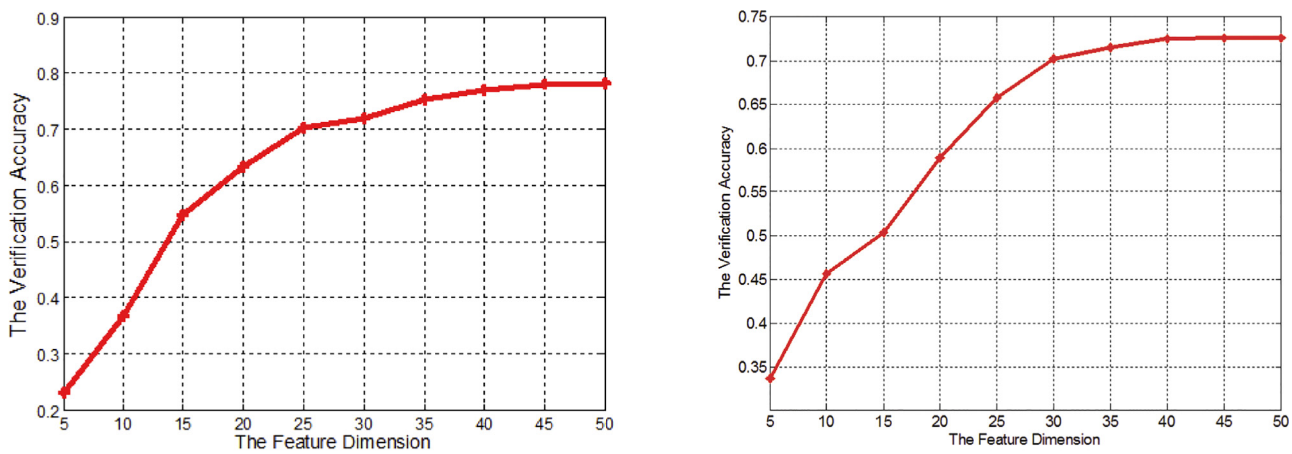


Fig. 9. The mean verification accuracy of DMML and MDMML versus the feature dimensions (a) FAM and (b) FBD.

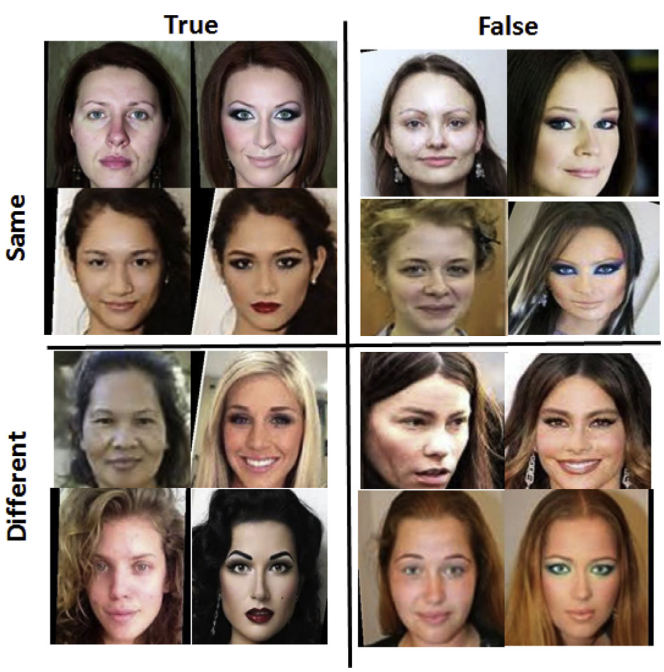


Fig. 10. The same and different labels are the ground truth labels of facial image pairs, and the true and false labels indicate whether the proposed method predict correctly. The experimental results show that the challenges of the makeup face verification problem.

and background. For instance, same makeup face pairs are misclassified because of pose ambiguity and different viewpoints. Meanwhile, different facial image pairs are always incorrectly classified as the same pairs due to the similar background, skin color and hair style.

5.3.7. Discussions

Here, we discuss some possible applications of the makeup face verification system in real-world scenarios. One of the most representative applications is that our makeup face verification can be used for social media analysis. There are usually billions of facial images on various social websites (i.e., Facebook, Flickr, etc.), and millions of images are added to the websites everyday. One key problem is how to automatically manage such large-scale images. In this problem, there are two challenges to be tackled: (1) who the people in images are and (2) how to recognize the facial images with cosmetics. Previous face recognition techniques may be effective to tackle the first problem, and makeup face verification should be a useful technique to alleviate the second challenge. When the facial makeup relation is known, it is possible for us to automatically organize the images according to the subject identities. Currently, our makeup face verification method has achieved around 70% accuracy when two facial images were captured under different conditions. It has provided us useful information to analyze the relation of two subjects since our methods can get much higher performance than a random guess.

Another important application of the makeup face verification is public security. Nowadays, face recognition techniques are the dominant approaches to recognize humans. However, conventional face recognition methods are usually hindered by glasses, cosmetics and thus these methods cannot be directly applied for public security. Makeup face verification techniques can provide useful tools to verify whether two facial images with cosmetics are from the same subject or not.

6. Conclusions and future work

In this paper, we have studied the makeup face verification problem in the wild. A real database of 17,866 facial makeup images with 8933 subjects was collected from Internet search for our study, which is named as Facial Beauty Database (FBD). To the best of our knowledge, our FBD is the largest database in the world, which can be used for facial makeup research in computer vision. Moreover, we have proposed a discriminative marginal metric learning (DMML) method for makeup face verification in the wild. Inspired by the fact that interclass marginal samples without makeup relations are always more discriminative than interclass nonmarginal samples in learning the discriminative metric space, we use the interclass marginal samples to depict the discriminative information and expect those interclass marginal samples are separated from each other as far as possible, such that more discriminative information can be exploited for verification. Since multiple feature representations could provide comprehensive information in describing the facial information from different points of view and capture more descriptive information, we further introduce a multiview discriminative marginal metric learning (MDMML) method by learning a robust metric space such that different feature representations can be effectively integrated to further improve the performance of makeup face verification. Experimental results on two real-world databases are used to show the possibility of verifying the identity via facial image analysis and the effectiveness of the new methods.

In the future, we are interested in exploring more discriminative feature representations (e.g., deep features) and combining them with MDMML to further improve the performance of makeup face verification. Moreover, we are also interested in applying our makeup face verification approaches to some interesting potential applications, e.g., social media analysis and public security.

Acknowledgment

This project was supported by the Engineering and Physical Sciences Research Council (EPSRC) (Ref: EP/M002632/1).

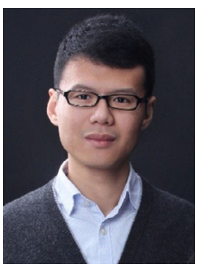
References

[1] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[2] C. Ding, J. Choi, D. Tao, L.S. Davis, Multi-directional multi-level dual-cross patterns for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 518–531.
[3] C. Ding, D. Tao, A comprehensive survey on pose-invariant face recognition, ACM Trans. Intell. Syst Technol. 7 (3) (2016) 37.
[4] G. Guo, Y. Fu, C.R. Dyer, T.S. Huang, Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 17 (7) (2008) 1178–1188.
[5] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach Intell. 27 (3) (2005) 328–340, doi:10.1109/TPAMI.2005.55.
[6] E. Kokiopoulou, Y. Saad, Orthogonal neighborhood preserving projections, in: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05), IEEE, 2005, p. 8.
[7] W. Yu, X. Teng, C. Liu, Face recognition using discriminant locality preserving projections, Image Vis. Comput. 24 (3) (2006) 239–248.
[8] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[9] Q. Feng, C. Yuan, J.S. Pan, J.F. Yang, Y.T. Chou, Y. Zhou, W. Li, Superimposed sparse parameter classifiers for face recognition, IEEE Trans. Cybern. 47 (2) (2017) 378–390.
[10] S. Yang, J. Liu, Y. Fang, Z. Guo, Joint-feature guided depth map super-resolution with face priors, IEEE Trans. Cybern. 48 (1) (2018) 399–411.
[11] C. Ren, Z. Lei, D. Dai, S. Li, Enhanced local gradient order features and discriminant analysis for face recognition, IEEE Trans. Cybern. 46 (11) (2016) 2656–2669.
[12] Z. Li, D. Gong, X. Li, D. Tao, Aging face recognition: a hierarchical learning model based on local patterns selection, IEEE Trans. Image Process. 25 (5) (2016) 2146–2154.
[13] X. Geng, Z. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2234–2240.
[14] E. Bondi, P. Pala, S. Berretti, A. Del Bimbo, Reconstructing high-resolution face models from kinect depth sequences, IEEE Trans. Inf. Forensics Secur. 11 (12) (2016) 2843–2853.
[15] I. Cohen, N. Sebe, A. Garg, M.S. Lew, T.S. Huang, Facial expression recognition from video sequences, in: Proceedings of the 2002 IEEE International Conference on Multimedia and Expo, ICME'02, 2, IEEE, 2002, pp. 121–124.
[16] W. Zheng, X. Zhou, C. Zou, L. Zhao, Facial expression recognition using kernel canonical correlation analysis (KCCA), IEEE Trans. Neural Netw. 17 (1) (2006) 233–238.
[17] A. Majumder, L. Behera, V.K. Subramanian, Automatic facial expression recognition system using deep network-based data fusion, IEEE Trans. Cybern. 48 (1) (2018) 103–114.
[18] B. Moghaddam, M. Yang, Learning gender with support faces, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 707–711.
[19] B. Moghaddam, M. Yang, Gender classification with support vector machines, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 2000, pp. 306–311.
[20] S. Zhang, X. Gao, N. Wang, J. Li, Robust face sketch style synthesis, IEEE Trans. Image Process. 25 (1) (2016) 220–232.
[21] S. Zhang, X. Gao, N. Wang, J. Li, M. Zhang, Face sketch synthesis via sparse representation-based greedy search, IEEE Trans. Image Process. 24 (8) (2015) 2466–2477.
[22] S. Gutta, J.R. Huang, P. Jonathon, H. Wechsler, Mixture of experts for classification of gender, ethnic origin, and pose of human faces, IEEE Trans. Neural Netw. 11 (4) (2000) 948–960.
[23] Y. Ou, X. Wu, H. Qian, Y. Xu, A real time race classification system, in: Proceedings of the 2005 IEEE International Conference on Information Acquisition, IEEE, 2005, pp. 6–pp.
[24] B. Burlando, L. Verotta, L. Cornara, E. Bottini-Massa, Herbal Principles in Cosmetics: Properties and Mechanisms of Action, CRC Press, 2010.
[25] A. Dantcheva, J. Dugelay, Female facial aesthetics based on soft biometrics and photo-quality, in: Proceedings of ICME, 2011.
[26] L. Wen, G. Guo, et al., Dual attributes for face verification robust to facial cosmetics, J. Comput. Vis. Image Process. 3 (1) (2013) 63–73.
[27] C. Chen, A. Dantcheva, A. Ross, Automatic facial makeup detection with application in face recognition, in: Proceedings of the 2013 International Conference on Biometrics (ICB), IEEE, 2013, pp. 1–8.
[28] J. Hu, Y. Ge, J. Lu, X. Feng, Makeup-robust face verification, in: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 2342–2346.
[29] G. Guo, L. Wen, S. Yan, Face authentication with makeup changes, IEEE Trans. Circuits Syst. Video Technol. 24 (5) (2014) 814–825.
[30] R. Ohtsuki, S. Tominaga, R. Hikima, Appearance analysis of human skin with cosmetic foundation, in: Proceedings of the IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2012, p. 82920.
[31] S. Varshovi, Facial Makeup Detection using HSV Color Space and Texture Analysis, Citeseer, 2012 Ph.D. thesis.
[32] D. Zeng, H. Chen, R. Lusch, S. Li, Social media analytics and intelligence, IEEE Intell. Syst. 25 (6) (2010) 13–16.
[33] Y. Tian, M. Lu, A. Hampapur, Robust and efficient foreground analysis for real-time video surveillance, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, IEEE, 2005, pp. 1182–1187.
[34] A. Dantcheva, C. Chen, A. Ross, Can facial cosmetics affect the matching accuracy of face recognition systems? in: Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), IEEE, 2012, pp. 391–398.
[35] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Netw. 10 (3) (1999) 626–634.
[36] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51, doi:10.1109/TPAMI.2007.250598.
[37] H.V. Nguyen, L. Bai, Cosine similarity metric learning for face verification, in: Proceedings of Asian Conference on Computer Vision, Springer, 2010, pp. 709–720.
[38] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.
[39] L. Zhang, L. Wang, W. Lin, Conjunctive patches subspace learning with side information for colaborative image retrieval, IEEE Trans. Image Process. 21 (8) (2012) 3707–3720, doi:10.1109/TIP.2011.2165335.
[40] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 209–216.
[41] C. Xu, D. Tao, C. Xu, Multi-view intact space learning, IEEE Trans. Pattern Anal. Mach. Intell. 37 (12) (2015) 2531–2544.
[42] M. Kan, S. Shan, H. Zhang, S. Lao, X. Chen, Multi-view discriminant analysis, in: Proceedings of European Conference on Computer Vision, Springer, 2012, pp. 808–821.
[43] P. Viola, M. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.
[44] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[45] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, IEEE, 2005, pp. 886–893.
[46] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[47] L. Wolf, T. Hassner, Y. Taigman, Descriptor based methods in the wild, in: Proceedings of Workshop on Faces in 'real-life' Images: Detection, Alignment, and Recognition, 2008.
[48] H. Yan, J. Lu, X. Zhou, Prototype-based discriminative feature learning for kinship verification, IEEE Trans. Cybern. 45 (11) (2015) 2535–2545.
[49] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 331–345.
[50] X. Zhou, J. Hu, J. Lu, Y. Shang, Y. Guan, Kinship verification from facial images under uncontrolled conditions, in: Proceedings of the 19th ACM International Conference on Multimedia, ACM, 2011, pp. 953–956.
[51] S. Shan, Y. Chang, W. Gao, B. Cao, P. Yang, Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution, in: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 2004, pp. 314–320.
[52] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 40 (6) (2010) 1438–1446.
[53] M. Wang, X. Hua, X. Yuan, Y. Song, L. Dai, Optimizing multi-graph learning: towards a unified video annotation scheme, in: Proceedings of the 15th ACM International Conference on Multimedia, ACM, 2007, pp. 862–871.
[54] B. Schölkopf, C. Burges, A. Smola, Advances in Kernel Methods: Support Vector Learning, MIT Press, 1998.
[55] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, University of Massachusetts, Amherst, 2007. Technical Report 07–49
[56] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov, Neighbourhood components analysis, in: Proceedings of Advances in Neural Information Processing Systems, 2004.

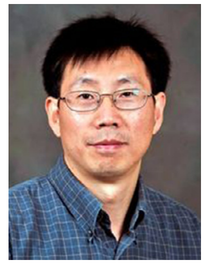
[57] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, J. Mach. Learn. Res. 12 (Jul) (2011) 2211–2268.

Lining Zhang (S'11-M'14) received the B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is a Senior Scientist with the Institute for Infocomm Research, Singapore. He has published extensively in top venues, such as the IEEE Transactions on Image Processing, the IEEE Transactions on Circuits and Systems for Video Technology, and the IEEE Transactions on Cybernetics. His research interests include artificial intelligence, computer vision, video/image processing, medical image analysis, machine learning, and computational intelligence. He is a member of the IEEE.

Hubert P. H. Shum is an Associate Professor (Reader) at Northumbria University. Before joining the university, he worked as a Lecturer at the University of Worcester, a post-doctoral researcher at RIKEN Japan, as well as a research assistant at the City University of Hong Kong. He received his Ph.D. degree from the School of Informatics at the University of Edinburgh, as well as his MSc and BEng degrees from the City University of Hong Kong. His research interests include character animation, machine learning, human motion analysis and computer vision.



Li Liu received the B.Eng. degree in electronic information engineering from Xian Jiaotong University, Xian, China, in 2011, and the Ph.D. degree from the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, U.K., in 2014. He is currently with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates. His current research interests include computer vision, machine learning, and data mining.



Guodong Guo is an Associate Professor in the Department of Computer Science and Electrical Engineering at West Virginia University. He is also the Director and Founder of the Computer Vision Laboratory (CVL) at WVU, and affiliated with the Center for Identification Technology Research (CITeR), a unique national Biometric research center funded by NSF. Dr. Guo received his Bachelor degree from Tsinghua University in Beijing, and Ph.D. In Computer Science at the Univsersity of Wisconsin – Madison.

Ling Shao (M '09-SM '10) is a professor with the School of Computing Sciences at the University of East Anglia, Norwich, UK. Previously, he was a professor (2014–2016) with Northumbria University, a senior lecturer (2009–2014) with the University of Sheffield and a senior scientist (2005–2009) with Philips Research, The Netherlands. His research interests include computer vision, image/video processing and machine learning. He is an associate editor of IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems and several other journals. He is a Fellow of the British Computer Society and the Institution of Engineering and Technology. He is a senior member of the IEEE.