# MAKEUP-ROBUST FACE VERIFICATION

*Junlin Hu*[1], *Yongxin Ge*[2], *Jiwen Lu*[3], *and Xin Feng*[4]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]School of Software Engineering, Chongqing University, China
[3]Advanced Digital Sciences Center, Singapore
[4]School of Computer Science and Engineering, Chongqing University of Technology, China

## ABSTRACT

We investigate in this paper the problem of face verification in the presence of face makeups. To our knowledge, this problem has less formally addressed in the literature. A key challenge is how to increase the measured similarity between face images of the same person without and with makeups. In this paper, we propose a novel approach for makeup-robust face verification, by measuring correlations between face images in a meta subspace. The meta subspace is learned using canonical correlation analysis (CCA), with the objective that intra-personal sample correlations are maximized. Subsequently, discriminative learning with the support vector machine (SVM) classifier is applied to verify faces based on the low-dimensional features in the learned meta subspace. Experimental results on our dataset are presented to demonstrate the efficacy of our approach.

***Index Terms*—** Makeup, face verification, canonical correlation analysis.

## 1. INTRODUCTION

Over the past two decades, a large amount of face recognition research has been done to pursue robustness to different intra-subject variations, including variations in pose [1], illumination [2], [3], and expression [4], [5]. Recently, several age-invariant face recognition algorithms [6], [7] have also been presented in the literature. However, to the best of our knowledge, the problem of face recognition in the presence of makeup has not been formally addressed in the literature. In many real world applications such as visual surveillance and web face image retrieval, there are usually some makeups on human faces, especially for females. Hence, a face recognition system which is robust to face makeup could be especially valuable to practical applications.

Different from face recognition, face verification aims to determine whether two face images come from the same person or not. We model face verification in the presence of face makeups as a two-class classification problem. Given an input image pair $I_1$ and $I_2$, the task is to assign the pair as either intra-personal ($I_1$ and $I_2$ from the same person) or inter-personal ($I_1$ and $I_2$ from different persons). Due to the makeup variation, the difference between two intra-personal face images taken without and with makeups is usually large. To address this, we propose learning a meta subspace in which the difference of each pair of intra-personal samples is reduced as much as possible. Fig. 1 illustrates the basic idea of our proposed approach. Let $X$ and $Y$ denote the face images without and with makeups, $W_x$ and $W_y$ are projection matrices learned by our approach on $X$ and $Y$, respectively. Then, each pair of intra-personal face samples without and with makeups are projected as close as possible in the learned meta feature subspace, such that more discriminative information can be exploited for verification. Experimental results show the efficacy of our approach.

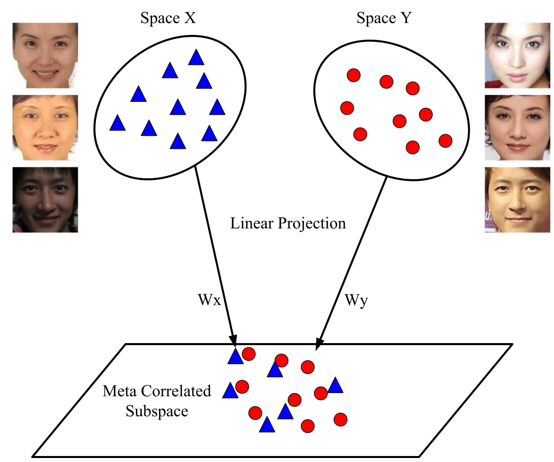


**Fig. 1**. Illustration of how maximizing intra-individual correlations leads to makeup invariance.

## 2. PROPOSED APPROACH

### 2.1. Correlation Maximized by CCA

We adopt canonical correlation analysis (CCA) [1], [8] to learn a meta feature subspace to project both face samples without and with makeups into a common feature space to achieve the makeup-invariance.

Let $(X, Y)$ be the training set consisting of $n$ pairs of feature vectors from the images taken without and with makeups, respectively, where $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^{p \times n}$, $Y = \{y_1, \ldots, y_n\} \in \mathbb{R}^{q \times n}$, $p$ and $q$ are dimensions of $X$ and $Y$, respectively. Here, both $X$ and $Y$ have been normalized to zero mean.

CCA aims to learn pairs of canonical components, i.e., projective directions $(w_x, w_y)$ which maximize the correlation between intra-personal feature vectors $(x, y)$ in the common space. More formally, the objective function of CCA$(X, Y)$ can be expressed as

$$
\begin{aligned}
(w_x, w_y) &= \arg\max_{w_x, w_y} \frac{\sum_{i=1}^{n} w_x^T x_i y_i^T w_y}{\sqrt{\sum_{i=1}^{n} w_x^T x_i x_i^T w_x}\sqrt{\sum_{i=1}^{n} w_y^T y_i y_i^T w_y}} \\
&= \arg\max_{w_x, w_y} \frac{w_x^T X Y^T w_y}{\sqrt{w_x^T X X^T w_x}\sqrt{w_y^T Y Y^T w_y}} \qquad (1)
\end{aligned}
$$

Since the means of $X$ and $Y$ are zero, $C_{xx} = XX^T$ and $C_{yy} = YY^T$ are the covariance matrices of $X$ and $Y$, respectively, and $C_{xy} = C_{yx}^T = XY^T$ is the within-person cross-covariance matrix between $X$ and $Y$.

To this end, the optimization problem of Eq. (1) can be rewritten as

$$
\arg\max_{w_x, w_y} w_x^T C_{xy} w_y \qquad (2)
$$

$$
\begin{aligned}
s.t. \quad & w_x^T C_{xx} w_x = 1 \\
& w_y^T C_{yy} w_y = 1 \qquad (3)
\end{aligned}
$$

Using the Lagrangian multiplier method [8], we can obtain $w_x$ and $w_y$ by solving the following generalized eigenvalue equation:

$$
\begin{bmatrix} & C_{xy} \\ C_{yx} & \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \qquad (4)
$$

where the generalized eigenvalue $\lambda$ is the correlation between $x$ and $y$. The eigenvectors $(w_{xi}, w_{yi})$, $i = 1, \ldots, d$, corresponding to the first $d$ largest non-zero generalized eigenvalues are the pairs of projective directions for CCA [8]. In fact, $C_{xx}$ and $C_{yy}$ could be singular in many applications such as face verification, which stems from the fact that the number of training images is smaller than the dimension of each image, a deficiency that is generally known as small sample size (SSS) problem. One possible way to address this SSS problem is to perform PCA projection to reduce the dimension of the original face feature space. It is to be noted that some useful discriminative information could be compromised in the intermediate PCA stage [9], [10]. Hence, we alternatively applied the regularization technique proposed in [11] to add a small perturbation $\mu I$ to $C_{xx}$ and $C_{yy}$ of CCA. In our experiments, $\mu$ is empirically set to be 0.000001.

**Table 1**. Comparisons of the average sum of squared difference (ASSD) before and after CCA.

| | Before CCA | After CCA |
|---|---|---|
| positive pairs | 0.0068 | $1.62 \times 10^{-10}$ |
| negative pairs | 0.0071 | 0.0071 |

Let $W_x = [w_{x1}, \ldots, w_{xd}]$ and $W_y = [w_{y1}, \ldots, w_{yd}]$, we project a pair of face images without and with makeups $(x, y)$ into a common feature space by $W_x$ and $W_y$ as follows:

$$
\hat{x} = W_x^T x \in \mathbb{R}^d \qquad (5)
$$

$$
\hat{y} = W_y^T y \in \mathbb{R}^d \qquad (6)
$$

where $(\hat{x}, \hat{y})$ is a pair of low-dimensional feature vectors in the learned common feature space.

We select 200 positive and 200 negative image pairs from our dataset, and compute the average sum of squared difference (ASSD) of these image pairs before and after CCA. The positive samples are the pair of images from the same person and the negative samples are the pair of images from different persons. Generally, the number of negative pairs is larger than that of positive pairs. Here, both 200 positive and negative pairs are from the same 200 subjects and the face images without and with makeups of each subject occur once in both the positive and negative pairs. Table 1 tabulates the ASSD of these 200 positive and 200 negative images, where ASSD is calculated as ASSD = $\frac{1}{200}\|\hat{x} - \hat{y}\|^2$. For a fair comparison, the feature dimension of each image is set to 100. For image pairs before CCA, PCA is used to reduce the feature dimensions, and for those after CCA, CCA is applied to reduce the feature dimensions. We can clearly see that CCA can significantly reduce the difference between intra-personal face image pairs without and with makeups.

### 2.2. Verification

Having obtained the feature representations of each pair of face images, we apply support vector machine (SVM) [12] to determine whether they are from the same or different persons. Specifically, given a pair of low-dimensional feature vectors in a meta feature subspace $(\hat{x}, \hat{y})$, they are first mapped into a feature space as

$$
z = F(\hat{x}, \hat{y}) \qquad (7)
$$

where $z$ is the feature vector extracted from $(\hat{x}, \hat{y})$ through a feature extraction function $F$.

In this study, we adopt the following conventional feature extraction function:

$$
F : z = \hat{x} - \hat{y} \in \mathbb{R}^d \qquad (8)
$$

**Fig. 2**. Several examples of our FAM database. From top to bottom are the cropped face images for females without makeup, females with makeup, males without makeup, and males with makeup, respectively. Each column of the first two rows are from the same female person, and each column from the bottom two rows are from the same male person.

where $z$ is a feature vector applied to face verification.

Lastly, SVM is used to divide the feature space into two classes, one for intra-personal pairs and the other for inter-personal pairs. In our implementations, we use the libsvm toolbox [12] for face verification.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Data Set

To advance the makeup-robust face verification research and show the effectiveness of our proposed approach, we build the FAce Makeup (FAM) Database from face images of public figures or celebrities without and with makeups available on the Internet. Our dataset contains 519 subjects, 222 of them are male and the remaining 297 are female. Each subject has two face images in our dataset, one is with makeup and the other is not. Due to the uncontrolled nature in capturing these face images, we pose no restrictions on the variations other than makeup. However, the key variations of these collected face images is the makeup factor because our key objective here is to investigate the problem of face verification in the presence of face makeups. In our experiments, the images are converted to gray-scale and normalized to $64 \times 64$ pixels according to the manually labeled eyes positions. Some examples of the cropped facial images are shown in Fig. 2.

### 3.2. Experimental Setup

We adopt a 5-fold cross-validation strategy in our experiments. Specifically, one group is used as testing data while the other four groups as training data, and the process is repeated five times for each group in turn to be used for testing. In the following experiments, we report the average verification rate with the standard deviation.

**Table 2**. Verification accuracy (%) obtained by different feature descriptors on our dataset (mean±std).

| Feature | Dim | Positive | Negative | Mean |
|---|---|---|---|---|
| LBP | 256 | $52.90 \pm 3.38$ | $49.60 \pm 5.79$ | $51.25 \pm 3.16$ |
| TPLBP | 256 | $50.80 \pm 2.63$ | $46.80 \pm 6.61$ | $48.80 \pm 2.20$ |
| FPLBP | 16 | $52.90 \pm 3.38$ | $49.60 \pm 5.79$ | $51.25 \pm 3.16$ |
| LE | 200 | $51.40 \pm 3.61$ | $48.50 \pm 6.12$ | $49.95 \pm 2.21$ |
| SIFT | 200 | $53.70 \pm 6.58$ | $51.30 \pm 5.80$ | $52.50 \pm 1.47$ |
| HOG | 9 | $50.08 \pm 6.19$ | $48.20 \pm 6.15$ | $49.50 \pm 2.75$ |
| Ours | 9 | $63.10 \pm 2.55$ | $61.70 \pm 5.39$ | $\mathbf{62.40 \pm 3.18}$ |

**Table 3**. Verification accuracy (%) obtained by different face verification methods on our dataset (mean±std).

| Method | Dim. | Positive | Negative | Mean |
|---|---|---|---|---|
| CSML | 50 | $60.30 \pm 3.56$ | $58.86 \pm 3.68$ | $59.58 \pm 3.43$ |
| MEML | 49 | $61.50 \pm 2.35$ | $60.50 \pm 5.39$ | $61.00 \pm 3.43$ |
| Ours | 9 | $63.10 \pm 2.55$ | $61.70 \pm 5.39$ | $\mathbf{62.40 \pm 3.18}$ |

### 3.3. Results and Analysis

**Experiment 1: Comparisons with Existing Face Feature Descriptors:** We compare our approach with several state-of-the-art face feature descriptors methods: Local Binary Patterns (LBP) [13], Three-Patch LBP (TPLBP) [14], Four-Patch LBP (FPLBP) [14], LEarning-based (LE) [15], and HOG [14]. For the LBP feature, we used 256 bins rather than 59 bins to describe each face image because we found such parameter setting achieved better performance than that used in [13]. For the LE method, we followed the parameter setting in [15] and used 200 bins to encode a histogram feature for each image. For the SIFT feature, we densely sampled and computed the SIFT descriptors of $16 \times 16$ patches over a grid with spacing of 8 pixels. For the TPLBP feature, we followed the parameter setting in [14] and used 256 bins to encode a histogram feature for each image. For details on these feature descriptors, refer to [13], [15], [16], [14]. For our method, the initial image features used are the raw pixel intensity values. Table 2 records the verification rate with the standard deviation obtained from different feature descriptor methods on our dataset. As shown in this table, our approach outperforms LBP, TPLBP, FPLBP, LE, SIFT and HOG with advantages in average verification accuracy of 11.15, 13.60, 11.15, 12.45, 9.90, and 12.90 percentage points, respectively.

**Experiment 2: Comparisons with State-of-the-art Face Verification Algorithms:** We compare our method with two state-of-the-art face verification methods: Cosine similarity metric learning (CSML) [17] and margin emphasized metric learning (MEML) [18]. Since the SIFT feature performs the best, we select it for feature representation.

**Table 4**. Verification accuracy (%) obtained by different feature descriptors on the male subset (mean±std).

| Feature | Dim. | Positive | Negative | Mean |
|---|---|---|---|---|
| LBP | 256 | $54.20 \pm 7.62$ | $53.40 \pm 8.48$ | $53.80 \pm 3.25$ |
| TPLBP | 256 | $53.20 \pm 9.04$ | $47.20 \pm 7.02$ | $50.20 \pm 3.19$ |
| FPLBP | 16 | $53.40 \pm 6.22$ | $48.20 \pm 6.40$ | $50.80 \pm 2.52$ |
| LE | 200 | $58.60 \pm 7.41$ | $45.40 \pm 6.60$ | $52.00 \pm 2.98$ |
| SIFT | 200 | $60.80 \pm 6.50$ | $62.80 \pm 5.26$ | $61.80 \pm 3.96$ |
| HOG | 9 | $50.80 \pm 5.67$ | $52.40 \pm 5.64$ | $51.60 \pm 2.17$ |
| Ours | 20 | $57.80 \pm 4.03$ | $57.60 \pm 4.82$ | $\mathbf{57.70 \pm 3.46}$ |

**Table 5**. Verification accuracy (%) obtained by different feature descriptors on the female subset (mean±std).

| Feature | Dim. | Positive | Negative | Mean |
|---|---|---|---|---|
| LBP | 256 | $54.00 \pm 6.11$ | $53.20 \pm 8.39$ | $53.60 \pm 5.89$ |
| TPLBP | 256 | $55.40 \pm 6.93$ | $52.40 \pm 7.58$ | $53.90 \pm 6.04$ |
| FPLBP | 16 | $53.80 \pm 6.26$ | $46.80 \pm 7.08$ | $50.30 \pm 5.00$ |
| LE | 200 | $54.60 \pm 4.90$ | $51.20 \pm 8.01$ | $52.90 \pm 5.98$ |
| SIFT | 200 | $58.00 \pm 5.57$ | $58.00 \pm 7.20$ | $58.00 \pm 4.88$ |
| HOG | 9 | $54.00 \pm 5.49$ | $55.40 \pm 6.80$ | $54.70 \pm 4.49$ |
| Ours | 10 | $58.80 \pm 4.65$ | $57.40 \pm 4.20$ | $\mathbf{58.10 \pm 4.01}$ |

For CSML and MEML, we follow the parameter settings in [17] and [18], respectively. Table 3 tabulates the verification rate with the standard deviation obtained from different face verification methods. We can observe that our approach is comparable to the state-of-the-art face verification methods.

**Experiment 3: Comparisons across Genders:** Since female makeups often alter appearances to greater extends than the male counterparts, it is interesting to investigate how the face verification performance differs between males and females. For a fair comparison, we select the same number of face image pairs for both males and females. Specifically, we use all the 222 image pairs of males in our dataset and randomly select the same number of female image pairs to construct the male and female subset, respectively. We adopt a 4-fold cross-validation strategy for experiments. Tables 4 and 5 record the verification rate with the standard deviation obtained from different feature descriptor methods on the male and female subset, respectively. As shown in these two tables, our approach is the best method on the female subset and the second best on the male subset, respectively. The reason why the SIFT feature descriptor is better than our approach is that under this scenario our approach is a learning-based method and there are only about 165 image pairs for learning the meta subspace and such limited number of samples may not effectively discover the relations between the face samples without and with makeups.

**Experiment 4: Comparisons with Human Observers in Face Verification Across Makeup Variance:** As an important baseline, the human ability in face verification across makeup variance was also tested. We selected 30 pairs (15 females and 15 males) of interpersonal images and presented them to 10 human observers (5 males and 5 females) with age of 20 to 30 years old. To make a fair comparison with our computer algorithm, the 30 pair images with size of $64 \times 64$ each were presented to human observers. None of them received training on the task before the experiment. Table 6 shows the accuracy of human ability on facial verification. We can observe that our proposed approach can achieve comparable performance with human observers on facial verification in presence of makeup for females, and performs worse than human for males. The reason is that face makeup for females tends to more significantly change the appearance of faces, making it more challenging for both computer and humans to correctly verify females without and with face makeups.

**Table 6**. Accuracy (%) of human ability on face verification across makeup variance.

| Method | Male | Female |
|---|---|---|
| Human | 90.00 | 63.33 |
| Ours | 63.30 | 61.50 |

## 4. CONCLUSION

In this paper, we have investigated the problem of face verification in the presence of makeup. We have built a dataset containing paired images of individuals without and with makeups. We have further proposed a canonical correlation analysis-based approach to increase the similarity of two face images as measured by correlations in a meta subspace between face images without and with makeups, such that more discriminative information can be exploited for verification. Experimental results on our face makeup dataset have demonstrated the efficiency of our approach.

## 5. ACKNOWLEDGEMENT

This work is partially supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center from the Agency for Science, Technology and Research (A*STAR) of Singapore, National Natural Science Foundation of China (61202348), and Chongqing Natural Science Foundation (CSTC2012jjA1549).

## 6. REFERENCES

[1] A. Li, S. Shan, X. Chen, and W. Gao, "Maximizing intra-individual correlations for face recognition across pose differences," in *CVPR*, 2009, pp. 605–611.

[2] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *PAMI*, vol. 23, no. 6, pp. 643–660, 2001.

[3] S. Z. Li, R. F. Chu, S. C. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *PAMI*, vol. 29, no. 4, pp. 627–639, 2007.

[4] PH Tsai and T. Jan, "Expression-invariant face recognition system using subspace model analysis," in *SMC*, 2005, pp. 1712–1717.

[5] A. Bronstein, M. Bronstein, and R. Kimmel, "Robust expression-invariant face recognition from partially missing data," in *ECCV*, 2006, pp. 396–408.

[6] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *PAMI*, vol. 24, no. 4, pp. 442–455, 2002.

[7] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," *PAMI*, vol. 32, no. 5, pp. 947–954, 2010.

[8] T. Sun, S. Chen, J. Yang, and P. Shi, "A novel method of combined feature extraction for recognition," in *ICDM*, 2008, pp. 1043–1048.

[9] J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognition*, vol. 36, no. 2, pp. 563–566, 2003.

[10] J. Yang, A. F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," *PAMI*, vol. 27, no. 2, pp. 230–244, 2005.

[11] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 445, pp. 165–175, 1989.

[12] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, Software available at: http://www.csie.ntu.edu.tw/ jlin/libsvm.

[13] T. Ahonen, A. Hadid, et al., "Face description with local binary patterns: application to face recognition," *PAMI*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *ECCVW*, 2008.

[15] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *CVPR*, 2010, pp. 2707–2714.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[17] H. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," *ACCV*, pp. 709–720, 2011.

[18] S. Li and S. Shan, "Margin emphasized metric learning and its application to gabor feature based face recognition," in *FG*, 2011, pp. 579–584.