

# Identifying Gender from Unaligned Facial Images by Set Classification

Wen-Sheng Chu<sup>†</sup>, Chun-Rong Huang<sup>†</sup> and Chu-Song Chen<sup>†‡¶</sup>

<sup>†</sup>*Institute of Information Science, Academia Sinica, Taipei, Taiwan*

<sup>‡</sup>*Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan*

<sup>¶</sup>*Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan.*

{wschu, nckuos, song}@iis.sinica.edu.tw

**Abstract**—Rough face alignments lead to suboptimal performance of face identification systems. In this study, we present a novel approach for identifying genders from facial images without proper face alignments. Instead of using only one input for test, we generate an image set by randomly cropping out a set of image patches from a neighborhood of the face detection region. Each image set is represented as a subspace and compared with other image sets by measuring the canonical correlation between two associated subspaces. By finding an optimal discriminative transformation for all training subspaces, the proposed approach with unaligned facial images is shown to outperform the state-of-the-art methods with face alignment.

**Keywords**—gender identification; set classification; face alignment; subspace learning; discriminative analysis

## I. INTRODUCTION

Gender identification, i.e., inferring female or male from facial images, is a useful preprocessing step for face applications, such as facial expression recognition and age estimation. Previous studies on gender identification have relied on careful alignment of a facial image into a standard template. The studies in [1], [2] conduct experiments on many combination of the state-of-the-art face alignment and pattern classification methods for gender identification. The authors found that the classification rates increase with the accuracy of face alignment because such alignment could reduce the variability during building the model in the training phase.

Since face detection can provide a rough face region for each subject, most works of gender identification regard the detected image patches as desired facial images. This approach, however, suffers from the problem that the detector could not specify a robust result of face location. The performance of gender identification, therefore, is somewhat limited to the accuracy of face alignment. In practice, false recognition may occur frequently because of variations of human poses or difficult alignments. Two widely used methods for aligning faces are affine warping for geometric shape alignment [3], and a statistical model built for facial shape and appearance [4]. Nevertheless, these alignment processes require feature detection or manual efforts of labeling specific facial features, e.g. eyes and mouth for affine warping or plenty facial landmark points for the AAM method [4]. When applied to a new test face set under

different variation settings, such as illumination or pose, the face model must be rebuilt from a newly collected training set to identify the variation. Moreover, rough alignment of a single input is likely to downgrade the results.

In this paper, we aim to discover the gender correlations from unaligned facial images to avoid alignment processes. We generate a facial image set by randomly cropping facial image patches around the detected face position of a single image. Rather than using the scattered distribution of these unaligned facial images, a subspace is constructed for each image set to extract the joint information from these randomly cropped facial patches. Because the facial image set contains more facial appearance of the same subject, a subspace is capable of representing more integral facial information than a single input. Thus, we employ sets of multiple unaligned facial images and represent them as subspaces to achieve better results for gender identification.

After each facial image set is represented as a subspace to express non-aligned facial variations of the same subject, we measure the similarity between two subspaces by canonical correlation, which has been successfully applied to several image-set-based applications such as object and face recognition [5], [6], [7], [8]. It is a notable property that the intersection between two subspaces of image sets are less sensitive to variations compared to that between a vector and a subspace [7]. Namely, a subspace constructed by unaligned facial images is a considerable representation to reduce the variation caused by face alignment. In addition, we find an optimal gender discriminative transformation for all training subspaces by exploiting the method of discriminant-analysis of canonical correlation (DCC) [6]. The transformation maximizes the correlation within subspaces of the same gender and simultaneously minimizes the correlation between subspaces of different genders. Finally, our approach is demonstrated to perform better gender identification results without face alignment over the approaches that assume the face is well aligned.

## II. UNALIGNED FACIAL IMAGE SETS

Due to the fact that face detection as well as face alignment algorithms does not align a face perfectly, we intend to discover the gender correlations among sets of unaligned facial images. To capture more facial variations for each

subject  $i$ , we collect an image set  $\mathbf{X}_i$  by randomly cropping out a set of facial image patches around the position initiated by face detection. An image set  $\mathbf{X}_i$  can be generated from image patches  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  locating at random positions, scales and rotations around the detected face region. As shown in Figure 1, red and blue rectangles indicate the detected face region and randomly cropped image patches, respectively. Here, we resize and reshape each image patch  $\mathbf{x}_i$  to an  $n$ -dimensional column vector.

Given image sets  $\{\mathbf{X}_1, \dots, \mathbf{X}_N\}$  for  $N$  subjects, we gather the common information of these unaligned facial images by describing each  $\mathbf{X}_i$  as its own linear subspace  $\mathcal{S}_i$  spanned by a set of bases. Similar to [5], [6], [8], an  $n \times d$  unitary basis matrix  $\mathbf{P}_i$  for representing a linear subspace associated with  $\mathbf{X}_i$  is computed by

$$\mathbf{X}_i \mathbf{X}_i^T \simeq \mathbf{P}_i \Lambda_i \mathbf{P}_i^T, \quad (1)$$

where  $\Lambda_i$  is the top- $d$  largest eigenvalue and  $\mathbf{P}_i$  is the corresponding eigenvectors forming unitary orthogonal bases for linear subspace  $\mathcal{S}_i$ . After compressing the general information of each facial image set into its subspace, we then find an optimal gender discriminative transformation  $\mathbf{T}$  for all subspaces  $\{\mathcal{S}_1, \dots, \mathcal{S}_N\}$  such that the correlation between transformed subspaces of the same gender are maximized and meanwhile those of different genders are minimized.

### III. GENDER DISCRIMINANT TRANSFORMATION

Let the  $n \times d$  orthonormal basis matrix  $\mathbf{P}_i$  represent the subspace  $\mathcal{S}_i$  of subject  $i$ . We measure the similarity between a pair of subspaces  $\mathcal{S}_i$  and  $\mathcal{S}_j$  by the canonical correlation. A standard algorithm [9] to compute the canonical correlation for two subspaces is the solution using singular value decomposition (SVD) of  $\mathbf{P}_i^T \mathbf{P}_j \in \mathbb{R}^{d \times d}$ :

$$\mathbf{P}_i^T \mathbf{P}_j = \mathbf{U}_{ij} \Lambda \mathbf{U}_{ji}^T \quad s.t. \quad \Lambda = \text{diag}(\sigma_1, \dots, \sigma_d), \quad (2)$$

where  $\mathbf{U}_{ij}$ ,  $\mathbf{U}_{ji}$  are orthonormal vectors, and the sum of the singular values  $\Lambda$  (or cosines of principal angles) is the canonical correlation associated with the two subspaces. The smaller the principal angles are, the larger the canonical correlation would be. In other words, the more a pair of subspaces intersect with each other, the more likely they belong to the same gender. We denote  $\mathbf{C}_{ij} = \mathbf{P}_i \mathbf{U}_{ij}$  and  $\mathbf{C}_{ji} = \mathbf{P}_j \mathbf{U}_{ji}$  as the canonical vectors (or principal vectors [9]). According to [5], [6], we measure the canonical correlation between two subspaces  $\mathcal{S}_i$  and  $\mathcal{S}_j$  as follows:

$$\text{corr}(\mathcal{S}_i, \mathcal{S}_j) \approx -\text{tr}((\mathbf{C}_{ij} - \mathbf{C}_{ji})^T (\mathbf{C}_{ij} - \mathbf{C}_{ji})). \quad (3)$$

One could view the correlation from a geometric perspective that the angle between two unitary bases is proportional to the length of difference between them.

Having defined the canonical correlation between two subspaces, our next goal is to obtain an optimal transformation  $\mathbf{T}$  that transforms all subspaces of image sets to another space that contains maximal correlation within the same

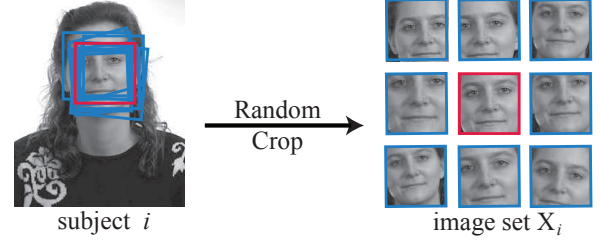


Figure 1. An illustration of our idea of generating an image set by randomly cropping out a set of facial patches (blue rectangles) around the face detection region (red rectangles).

gender and minimal correlation between different genders. To find  $\mathbf{T}$ , we employ the discriminant-analysis of canonical correlations (DCC) [6] by solving

$$\mathbf{T} = \arg \max_{\mathbf{T}} \frac{\text{tr}(\mathbf{T}^T \mathbf{S}_b \mathbf{T})}{\text{tr}(\mathbf{T}^T \mathbf{S}_w \mathbf{T})}. \quad (4)$$

$\mathbf{S}_b$  and  $\mathbf{S}_w$  are the between-gender and within-gender scatter matrices defined according to Eq. (3) and are given by

$$\mathbf{S}_b = \sum_{i=1}^N \sum_{k, c_i \neq c_k} (\mathbf{C}'_{ik} - \mathbf{C}'_{ki})(\mathbf{C}'_{ik} - \mathbf{C}'_{ki})^T \quad (5)$$

$$\mathbf{S}_w = \sum_{i=1}^N \sum_{j, c_i = c_j} (\mathbf{C}'_{ij} - \mathbf{C}'_{ji})(\mathbf{C}'_{ij} - \mathbf{C}'_{ji})^T \quad (6)$$

The optimal transformation matrix  $\mathbf{T}$  is then obtained by solving Eq. (4) as a eigen-decomposition of  $\mathbf{S}_w^{-1} \mathbf{S}_b$ . Note that the non-singularity of  $\mathbf{S}_w$  can be guaranteed in several ways, such as pseudo-inverse [10] or pre-applying PCA to reduce the rank of data similarity [11].

The transformed canonical vectors  $\mathbf{C}'_{ij} = \mathbf{Q}_i \mathbf{U}'_{ij}$  and  $\mathbf{C}'_{ji} = \mathbf{Q}_j \mathbf{U}'_{ji}$  used in Eq. (5) and (6) are obtained by calculating canonical correlation between two transformed subspaces  $\mathcal{S}'_i$  and  $\mathcal{S}'_j$  through the SVD of  $\mathbf{Q}_i^T \mathbf{Q}_j$ :

$$\mathbf{Q}_i^T \mathbf{Q}_j = \mathbf{U}'_{ij} \Lambda \mathbf{U}'_{ji}^T. \quad (7)$$

Recalling Eq. (2) that canonical correlation is defined for orthonormal inputs of subspaces,  $\mathbf{Q}_i$  is obtained as an orthonormal basis matrix for the associated subspace  $\mathcal{S}'_i$  by computing a preliminary  $QR$ -factorization of  $\mathbf{T}^T \mathbf{P}_i$  and can be written as follows:

$$\mathbf{Q}_i = \mathbf{T}^T (\mathbf{P}_i \mathbf{R}_i^{-1}), \quad (8)$$

where  $\mathbf{R}_i$  is the  $d \times d$  invertible upper-diagonal matrix.

### IV. EXPERIMENTAL RESULTS

In our study,  $24 \times 24$ -pixel resolution facial image patches were pre-located using the cascaded face detector [12] without further face alignment processes. We perform histogram equalization on each facial image patch for normalizing ambient lighting variations. Two public available databases,

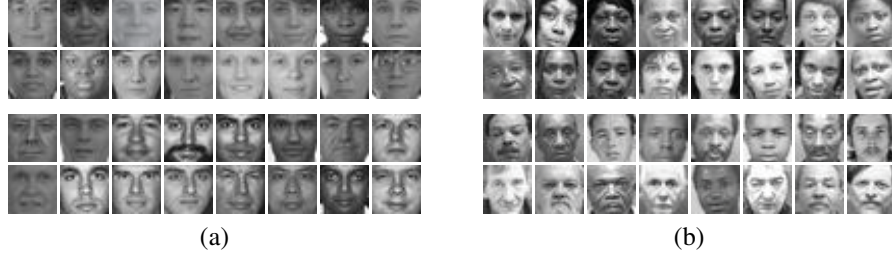


Figure 2. Example facial patches of female (top two rows) and male (bottom two rows) from (a) the FERET database and (b) the MORPH database, where the later database can be regarded as a more challenging one because of more variation in aging and ethnicity.

the FERET database [13] and the MORPH database [14], were used for the experiments. While FERET is one of the mostly used database with good quality facial images for gender identification, MORPH is a notable illustration of much larger number of facial images with more challenging variations in terms of aging and ethnicity.

In order to compare our results with those reported in [1], we followed their experimental setups as follows. We used fa- and fb-subsets from the FERET database by removing duplicate images of the same subject; at the end there are 450 facial images for both females and males. For the MORPH database, which is not used in [1], we collect 8,033 female and 47,810 male images; there are totally 418 subjects of age ranging from 18-69 and of races of Caucasian, African-American and Asian. To the best of our knowledge, we are the first tackling the gender identification problem over such variations with wide range of age and races. Example images from both databases are shown in Figure 2.

We collected 75 facial image patches as the image set for each subject. A subspace is constructed for each subject using Eq. (1) with 98 percentage of data energy preserved. Being proved a stable method against the dimensionality of transformation  $\mathbf{T}$  [6], the dimension is fixed at 160 for the following experiments. The solution of  $\mathbf{T}$  was ensured nonsingular by computing the pseudo-inverse of the within-gender scatter matrix in Eq. (6).

We compared the performance of the proposed subspace-based approach to discriminative sample-based methods, i.e., k-Nearest Neighbor (kNN) of images transformed by “PCA plus LDA” [11]. The number of nearest neighbors is set to 1 and 10. Mutual subspace method (MSM) [8], a simpler accession of canonical correlations, was also implemented for verifying the discriminative power of the transformation  $\mathbf{T}$ . Our performance was reported by classification accuracy versus the number of training subspaces. Note that each subspace represents a subject; the same number of female and male subjects for both training and test were chosen from different image sets.

As shown in Figure 3, it is easier to obtain a better result for the FERET than the MORPH database, because the facial images are frontal and easier to classify even though alignment processes are ignored. The accuracy increases

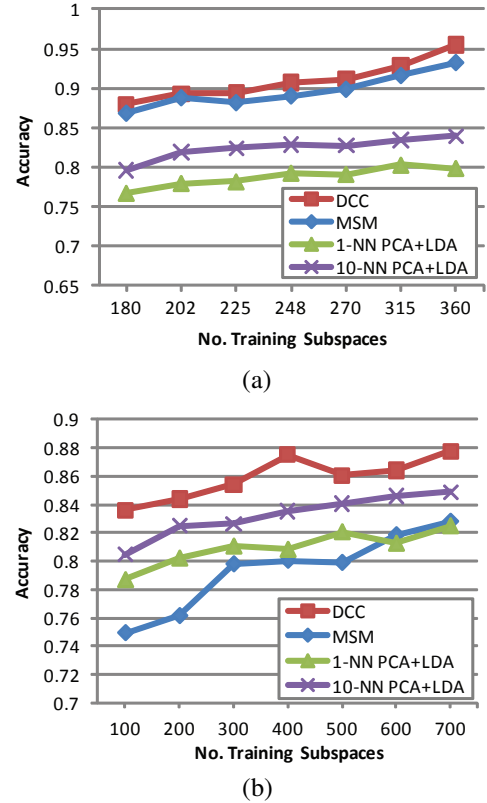


Figure 3. Classification accuracy between different set classification methods of (a) the FERET database and (b) the MORPH database over different numbers of training subspaces.

with the number of training subspaces in both databases since the larger number of training subspaces are selected, the more discriminatory gender information from different subjects can be learned. It can be seen from Figure 3 (a) that both subspace-based methods, DCC and MSM, outperform sampled-based methods. Although the input facial images are not aligned in advance, subspaces are capable of providing more general information of a subject than a single image regardless of alignment processes. In Figure 3 (b), however, the performance of MSM is inferior to 10-NN PCA+LDA. Even 1-NN PCA+LDA performs better

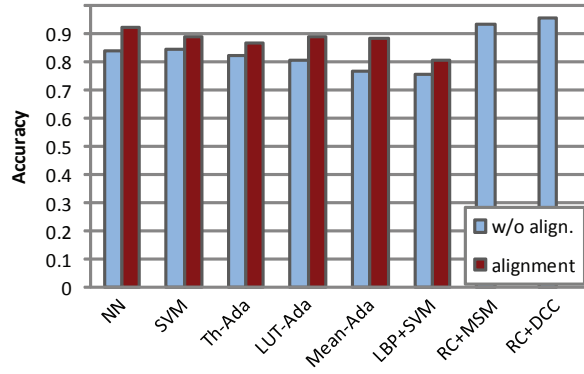


Figure 4. Comparison of classification rates of the FERET database between our method and those discussed in [1].

than MSM when the number of training subspaces is not large enough. Because the MORPH database contains more diverse range of variations and less controlled facial poses than the FERET database, training subspaces are likely to spread too exclusively for test subspaces to match the correct gender. The DCC method, on the other hand, discovers an optimal transformation that projects these subspaces to a more discriminative space. Thus, as shown in both figures in Figure 3, DCC delivers a the most robust performance with less difference of accuracy compared to other methods.

Finally, we compare our results to those reported in [1]. These methods include Neural Network (NN), Support Vector Machine (SVM), Threshold Adaboost (Th-Ada), LUT Adaboost (LUT-Ada), Mean Adaboost (Mean-Ada) and Local Binary Patterns combined with SVM (LBP+SVM). Note that the experiments in [1] were conducted only on one public available database, FERET. Results for FERET images with and without alignment are shown in Figure 4 (colored dark red and light blue, respectively); RC is abbreviated from “random cropping” for the proposed method. We found that the performance of the proposed method is better than the methods without alignments, and still slightly better than those with proper alignments. The benefits of combining image sets with subspace matching methods were confirmed through this experiment.

## V. CONCLUSION

In summary, we have presented an approach for identifying genders from unaligned facial images. Rather than using only one facial image for test, we generate a set of unaligned images form a single input and represent each image set as a subspace. All subspaces are transformed by considering discriminating set information between different genders, and have been shown to improve the performance of gender identification using unaligned facial images.

## ACKNOWLEDGMENT

This research was supported in part by the National Science Council of Taiwan, under the Grants NSC99-2631-H-001-020 and NSC98-2221-E-001-012-MY3.

## REFERENCES

- [1] E. Mäkinen and R. Raisamo, “An experimental comparison of gender classification methods,” *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544–1556, 2008.
- [2] —, “Evaluation of gender classification methods with automatically detected and aligned faces,” *TPAMI*, vol. 30, no. 3, pp. 541–547, 2008.
- [3] B. Moghaddam and M.-H. Yang, “Learning gender with support faces,” *TPAMI*, vol. 24, pp. 707–711, 2002.
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” in *European Conference on Computer Vision* (2), 1998, pp. 484–498.
- [5] W.-S. Chu, J.-C. Chen, and J.-J. Lien, “Kernel discriminant analysis based on canonical differences for face recognition in image sets,” *Asian Conference on Computer Vision*, vol. 4844, pp. 700–711, 2007.
- [6] T.-K. Kim, J. Kittler, and R. Cipolla, “Discriminative learning and recognition of image set classes using canonical correlations,” *TPAMI*, vol. 29, no. 6, pp. 1005–1018, 2007.
- [7] M. Nishiyama, M. Yuasa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi, “Recognizing faces of moving people by hierarchical image-set matching,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [8] O. Yamaguchi, K. Fukui, and K. I. Maeda, “Face recognition using temporal image sequence,” in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 318–323.
- [9] G. H. Golub, “Numerical methods for computing angles between linear subspaces,” *Mathematics of Computation*, vol. 27, no. 123, pp. 579–594, 1973.
- [10] G. H. Golub and W. Kahan, “Calculating the singular values and pseudo-Inverse of a matrix,” *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.
- [11] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman *et al.*, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” *TPAMI*, vol. 19, no. 7, pp. 711–720, 1997.
- [12] P. Viola and M. Jones, “Robust real-time object detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2002.
- [13] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, “The FERET database and evaluation procedure for face-recognition algorithms,” *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.
- [14] M. Minear and D. C. Park, “A lifespan database of adult facial stimuli,” *Behavior Research Methods Instruments and Computers*, vol. 36, no. 4, pp. 630–633, 2004.