Two-dimensional Linear Discriminant Analysis for Low-Resolution Face Recognition

Di Zhao, Zhenxue Chen*, Chengyun Liu and Yanan Peng
School of Control Science & Engineering
Shandong University
Jinan, China
chenzhenxue@sdu.edu.cn

Abstract—**Low-resolution (LR) is a challenging problem in the real world. In order to obtain better performance for low-resolution face recognition (LRFR), this paper employs a novel approach for matching low-resolution images with high resolution (HR) images based on two-dimensional linear discriminant analysis (2D-LDA) and metric learning method. The LR and HR images are transformed into a common space via the 2D-LDA method in which the most discriminative information between them is preserved. Also, it overcomes the singularity and loss of the spatial information problem because of its matrix representation. To further improve the recognition performance, metric learning method is used based on neighborhood component analysis (NCA) which aims to maximize the leave-one-out (LOO) classification accuracy. Experiments on the ORL database with a wide range of resolutions illustrate the usefulness of the proposed method.**

Keywords—Low Resolution; Two-Dimensional Linear Discriminant Analysis; Discriminative Information; Metric Learning; Neighborhood Component Analysis

I. INTRODUCTION

Face recognition achieved impressive success under controlled conditions in recent years. However, with the wide range of surveillance cameras, LR issue may arise in unconstrained environments, causing the discriminatory information lost. Thus, the LRFR is still a tough problem.

Existing LRFR methods can be classified into the following categories [1]. One category is face hallucination [2] which reconstructs the HR facial images from the LR ones. Most of these methods pay more attention to the visual quality of the reconstructed images rather than the recognition performance. Furthermore, it has the high computational complexity. The second category is to extract resolution-robust feature [3][4]. Shortcomings of these methods are the lack of relation information between the HR and LR sample pairs.

Recently, a lot of works have been focused on projection-based methods [5], where HR and LR images are projected into a common space with a specific objective. Coupled mappings are learned to project the different resolutions images into a common feature space by Li et al. [6]. In addition, the local relationship is preserved due to the penalty weighting matrix. Ren et al. [7] proposed coupled kernel embedding (CKE) method. The locality between neighborhood in the kernel space is preserved because of this method. Furthermore, the dissimilarities obtained by the Gram matrices are minimized in the LR and HR spaces. Zhou et al. [8] introduced the simultaneous discriminant analysis approach which aims to learn two mappings that project LR and HR images into a unified space. Thus, the most discriminative information is preserved. The multidimensional scaling (MDS) method is introduced to the LRFR domain by Biswas et al. [9]. It embeds LR probe images and HR gallery images to a distance-preserving common space.

This paper proposes a novel LRFR algorithm based on 2D-LDA, where a mapping is introduced to transform the images from HR and LR into a common space. Metric learning based on NCA is further extended the algorithm to maximize the performance of nearest neighbor classification. The main contributions of this paper are summarized below.

(1) In the training stage, a discriminant transformation matrix based on 2D-LDA is learned. With the learned transformation matrix, the between-class and within-class information between the HR and LR training images which is essential for recognition performance is explored. In addition, the spatial information is preserved because the 2D-LDA algorithm works with images in matrix representation.

(2) A distance metric learning method based on NCA is proposed. This method finds the optimal weights that are used for LRFR with the objective of maximizing the LOO accuracy in the transformed space.

The remainder of this paper is organized as follows. Section II elaborates the proposed discriminant transformation matrix learning and metric learning for LRFR. Section III presents the experimental results and analysis. Finally, the conclusion is drawn in Section IV.

II. THE PROPOSED METHOD

A. *Discriminant Transformation Matrix Learning*

Linear discriminant analysis (LDA) is a classic algorithm which projects the data into a space in order to maximize the

This work is supported by National Natural Science Foundation of China (61203261), China Postdoctoral Science Foundation funded project (2012M521335), Jiangsu Key Laboratory of Big Data Analysis Technology/B-DAT (Nanjing University of Information Science & Technology, Grant No.: KXK1404) and The Fundamental Research Funds of Shandong University (2015JC014).

978-1-5386-3524-7/17/$31.00 ©2017 IEEE

ratio of the between-class scatter and within-class scatter. Thus, it obtains maximum discriminative information which is the key insight to improve LRFR performance. However, the within-class scatter matrix may be singular in a real application and LDA works with vectorized representation that may lose spatial information. To overcome the limitations of LDA, the training HR and LR images are used to learn the discriminant transformation matrix using a 2D-LDA [10] approach in this section. In particular, the raw images used for the 2D-LDA algorithm have the form of 2D matrices instead of 1D vectors, which is the main difference between 2D-LDA and LDA. Therefore, the facial spatial structure information which is important for face recognition can be better preserved.

Given a training set with M face HR-LR pairs, each image is represented as a raw pixel matrix. Assume there are c classes, and M_i is the number of training subset from the ith class. The between-class scatter matrix and within-class scatter matrix which reflect the relationship between HR-LR pairs can be formulated as

$$S_w^{HL} = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{M_i}(H_{ij} - \overline{L}_i)^{\mathrm{T}}(H_{ij} - \overline{L}_i) \tag{1}$$

$$S_w^{LH} = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{M_i}(L_{ij} - \overline{H}_i)^{\mathrm{T}}(L_{ij} - \overline{H}_i) \tag{2}$$

$$S_b^{HL} = \frac{1}{M}\sum_{i=1}^{c} M_i(\overline{H}_i - \overline{L})^{\mathrm{T}}(\overline{H}_i - \overline{L}) \tag{3}$$

$$S_b^{LH} = \frac{1}{M}\sum_{i=1}^{c} M_i(\overline{L}_i - \overline{H})^{\mathrm{T}}(\overline{L}_i - \overline{H}) \tag{4}$$

where $\overline{H}$ and $\overline{L}$ denotes the mean image of the whole HR and LR training samples respectively, $\overline{H}_i$ is the mean matrix on HR training images from the ith class and $\overline{L}_i$ is that on LR training images. Here, H_{ij} and L_{ij} respectively are the jth HR and LR image from class i. S_w^{HL} represents the within-class scatter matrix by establishing the difference from each HR image to the mean LR image from the same class. S_w^{LH}, S_b^{HL}, S_b^{LH} are defined similarly as S_w^{HL}. Based on the above formulations, the class relationship between the LR and HR images can be written as

$$S_w = S_w^{HL} + S_w^{LH} \tag{5}$$

$$S_b = S_b^{HL} + S_b^{LH} \tag{6}$$

Let W denotes the transformation matrix. 2D-LDA aims to learn a discriminant transformation matrix W which maximizes the between-class distance and minimizes the within-class distance. Thus, we introduce the following criterion [11] [12]

$$J(W) = \frac{W^{\mathrm{T}}S_bW}{W^{\mathrm{T}}S_wW} \tag{7}$$

When the criterion is maximized, the optimal transformation matrix $W^* = [w_1, w_2, \cdots, w_d]$ is obtained. In order to choose W^*, the optimization problem is equivalent to the following generalized eigenvalue problem: $S_bw = \lambda S_ww$. Hence, the eigenvectors of $S_w^{-1}S_b$ corresponding to the first d largest eigenvalues form the optimal transformation matrix W^*.

B. Metric Learning Based on NCA

Based on above analysis, the training HR samples and testing LR samples are first projected to the common space using the acquired transformation matrix W^* in the training phase. If $\hat{H}$ and $\hat{L}_{test}$ represent the feature matrix corresponding to HR training and LR testing images, the transformed feature matrix are given by

$$\hat{H} = HW^* \quad ; \quad \hat{L}_{test} = L_{test}W^* \tag{8}$$

For subsequent processing, we convert $\hat{H}$ and $\hat{L}_{test}$ matrix to L-dimension vectors $\hat{h}$ and $\hat{l}_{test}$ by concatenating each column pixels successively.

Let $D = \{(\hat{h}_1, y_1), (\hat{h}_2, y_2), \cdots (\hat{h}_i, y_i), \cdots (\hat{h}_M, y_M)\}$ be HR training set, where $\hat{h}_i$ is treated as L-dimension feature vectors while $y_i \in \{1, 2, \cdots, c\}$. Aiming to improve the performance of nearest neighbor classifier, this paper introduce an appropriate distance metric learning method based on NCA [13][14] feature selection with regularization to learn feature weights γ on the HR training set. We represent the weighted distance between two HR training samples $\hat{h}_i$ and $\hat{h}_j$ by

$$D(\hat{h}_i, \hat{h}_j) = \sum_{l=1}^{L}\gamma_l^2\left|\hat{h}_{il} - \hat{h}_{jl}\right| \tag{9}$$

where γ_l is the lth feature weight.

Suppose LOO classification accuracy is the evaluation index of the nearest neighbor classifier. Our goal is to maximize LOO accuracy on the training set D. Hence, each sample $\hat{h}_i$ selects another sample $\hat{h}_j$ as its neighbor with some probability p_{ij}, and its class label is inherited from the point it selects. The p_{ij} is defined as follows

$$p_{ij} = \begin{cases} \dfrac{\exp(-D(\hat{h}_i, \hat{h}_j))}{\sum_{k\neq i}\exp(-D(\hat{h}_i, \hat{h}_j))}, & if\ \mathrm{i} \neq j \\ 0, & if\ \ i = j \end{cases} \tag{10}$$

Hence the probability of the HR training sample $\hat{h}_i$ being correctly classified is given by

$$p_i = \sum_{j\in\Omega_i} p_{ij} \tag{11}$$

Here, Ω_i is considered as the set of samples in the same class as $\hat{h}_i$. Then the LOO accuracy over the sample set can be written as

$$F(\gamma)=\sum_{i=1}^{M}\sum_{j\in\Omega_i}p_{ij} \tag{12}$$

In addition, a regularization term can be added to the (12) to avoid overfitting, resulting in the following objective function

$$f(\gamma)=-(\sum_{i=1}^{M}\sum_{j\in\Omega_i}p_{ij}-\lambda\sum_{l=1}^{L}\gamma_l^2) \tag{13}$$

where the regularization parameter λ can be tuned via five-fold cross validation. Finding the weight vector γ can be further calculated as the following problem.

$$\lambda^{*}=\arg\min f(\gamma) \tag{14}$$

The stochastic gradient descent (SGD) is used to minimize the objective function solving for the feature weights γ. Next, we select more important features for classification using the feature weights γ and a relative threshold β which is used to extract features with feature weights . We choose the feature weights which satisfy $\gamma_i \geq \beta\times\max(1,\max(\gamma))$. Therefore, the final feature weights γ^{*} is obtained.

C. Classification

Based on γ^{*}, the final HR training feature vectors and LR testing feature vectors of each image are obtained.

$$\tilde{h}=\hat{h}\gamma^{*} \quad ; \quad \tilde{l}=\hat{l}_{test}\gamma^{*} \tag{15}$$

Then, the paper used Euclidean distance between them to measure the feature vectors.

$$d=\sqrt{\sum_{i=1}^{l'}\left|\tilde{h}_i-\tilde{l}_i\right|^2} \tag{16}$$

where l' is the length of the feature vectors. Finally, we apply a nearest neighbor (NN) classifier using the selected features to the final HR training feature vectors and LR testing feature vectors.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, experiments are conducted on ORL database to evaluate the LRFR performance of the proposed method.

A. *Algorithm Analysis for a Certain Resolution*

The ORL database consists of 400 face images of 10 individuals. The resolution of HR training images is kept to be the original resolution 112×92 in the following experiments. Notably, the images below 112×92 are first down-sampled into specified resolution and up-sampled using bicubic interpolation algorithm in order to be recognized. We perform experiments with resolution of 12×10 pixels (corresponding scale factor of 10) as an example to conduct experiments and analysis. Fig. 1 shows some examples of HR and LR images on the ORL database.

Fig. 1. Examples of HR and LR images. (a) HR images of size 112×92. (b) LR images of size12×10. (c) Images using bicubic interpolation.

We select six of the HR and LR images for training and the other LR images for testing. We now evaluate the effect of varying dimension of the optimal transformation matrix d and the regularization parameter λ.

We first set the dimension of the optimal transformation matrix d to 3 and vary the regularization parameter λ. Then λ is tuned using five-fold cross-validation with the objective of producing the minimum classification loss on the HR training set. Fig. 2 shows the classification loss corresponding to different regularization parameter λ.

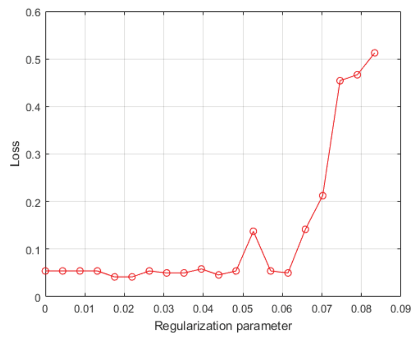

Fig. 2. The loss function values corresponding to different regularization parameters.

The results show the best regularization parameter that corresponds to the minimum loss function values is 0.0175. Then we use the solver SGD on all HR training data using the best λ to obtain the feature weights γ. In order to further improve the performance of classification, we set a relative threshold β=0.02 to obtain the final feature weights γ^{*}. The feature weights are shown in Fig. 3.

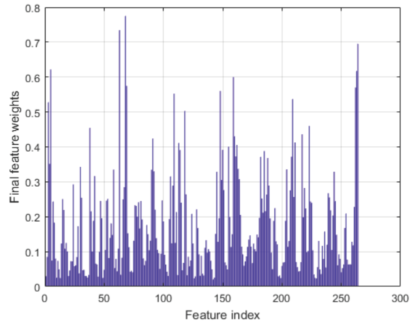

Fig. 3. Final feature weights on all HR training data

We further test the LRFR face recognition performance by changing the values of the dimension of the optimal transformation matrix d. Each dimension corresponds to its best λ value. NN classifier is used to achieve the recognition rate. Fig. 4 illustrates the recognition accuracy with corresponding to different values of the dimension of the optimal transformation matrix.

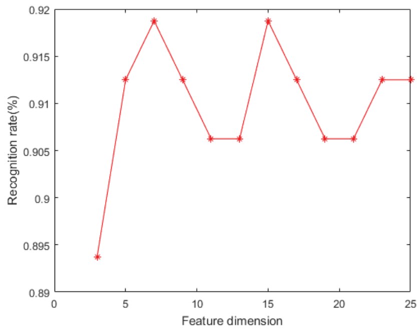

Fig. 4. The recognition accuracy corresponding to different dimensions of the optimal transformation matrix.

As shown in Fig. 4, the recognition accuracy vary with feature dimensions, and there is not a smooth relationship between them. For subsequent comparisons, we select the feature dimension corresponding to the best recognition accuracy which is 91.88%.

B. *Performance Comparisons with Different Resolutions*

In this section, the experiments are conducted over a wide range of resolutions the ORL database. In order to verify the effectiveness of the proposed method, we compare the proposed method with other projection-based approaches such as two-dimensional kernel principal component analysis (KPCA), principal component analysis (2D-PCA), marginal fisher analysis (MFA) and linear discriminant analysis (LDA). In order to avoid singular problem, we apply PCA algorithm before LDA and MFA.

We perform experiments on the ORL database with five different resolutions, namely, 56×46, 28×23, 19×16, 14×12 and 12×10 (corresponding scale factor of 2, 4, 6, 8, 10). Notably, the experimental procedure is the same as the previous work in A section except the resolution. The recognition performance is illustrated in Fig. 5.

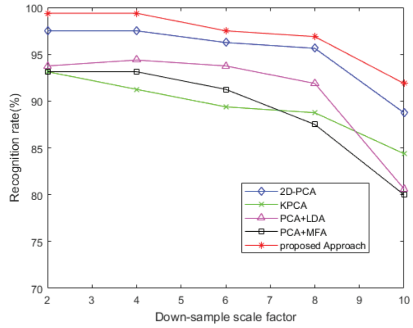

Fig. 5. Recognition rates on the ORL database

Fig. 5 shows that the proposed method achieves a relative high recognition accuracy for a wide range of resolutions.

IV. CONCLUSIONS

This paper proposes a novel method to address the problem of LRFR. The main idea is to learn a discriminant transformation matrix based on 2D-LDA. With the learned transformation matrix, the relationship of the LR and HR images and the most discriminative information are preserved. In order to further improve the recognition performance, we introduce an appropriate distance metric learning method based on NCA which selects more important features. Experimental results on the ORL database illustrate the effectiveness of the proposed method. In future work, we will extend this method and use it in the surveillance videos.

REFERENCES

[1] Z. Wang, Z. Miao, Q. M. J. Wu, Y. Wan, and Z. Tang, "Low-resolution face recognition: A review," The Visual Computer, vol. 30, no. 4, pp. 359-386, 2014.

[2] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A Comprehensive Survey to Face Hallucination," International Journal of Computer Vision (IJCV), vol.106, no. 1, pp. 9-30, 2014.

[3] Z. Lei, T. Ahonen, M. Pietikainen, and S.Z. Li, "Local frequency descriptor for low-resolution face recognition," In: Proc. IEEE 9th Int. Conf. on Automatic Face and Gesture Recognition (FG), Santa Barbara, CA, USA, pp. 161–166, 2011.

[4] J. Li, Z. Chen, and C. Liu, "Low-Resolution Face Recognition of Multi-Scale Blocking CS-LBP and Weighted PCA," International Journal of Pattern Recognition and Artificial Intelligence, vol. 30, no. 8, 2016.

[5] S. Ouyang, T. Hospedales, Y. Song, et al. "A Survey on Heterogeneous Face Recognition: Sketch, Infra-red, 3D and Low-resolution *," Image & Vision Computing, vol. 56, pp. 28-48, 2014.

[6] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," IEEE Signal Process. Letters, vol. 17, no. 1, pp. 20–23, 2010.

[7] C.X. Ren, D.Q. Dai, and H. Yan, "Coupled kernel embedding for low-resolution face image recognition," IEEE Transactions on Image PRocessing, vol. 21, no. 8, pp. 3770-3783, 2012.

[8] C. Zhou, Z. Zhang, D. Yi, Z. Lei, and S. Li, "Low-resolution face recognition via simultaneous discriminant analysis," The International Joint Conference on Biometrics (IJCB), pp. 1-6, 2011.

[9] S. Biswas, K.W. Bowyer, and P.J. Flynn, "Multidimensional scaling for matching low-resolution face images," TPAMI, vol. 34, no. 10, pp. 2019-2030, 2012.

[10] M. Li, and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," Pattern Recognition Letters, vol. 26, no. 5, pp. 527-532, 2005.

[11] J. Ye, R. Janardan, and Q. Li, "Two-Dimensional Linear Discriminant Analysis," Photogrammetric Engineering & Remote Sensing, vol. 5, no. 6, pp. 1431-1441, 2004.

[12] J. Yang, D. Zhang, X. Yong, and J.Y. Yang, "Two-dimensional discriminant transform for face recognition," Pattern Recognition, vol. 38, no. 7, pp. 1125-1129, 2005.

[13] M. Butman, and J. Goldberger, "Face Recognition Using Classification-Based Linear Projections," EURASIP Journal on Advances in Signal Processing, 2008.

[14] W. Yang, K. Wang, and W. Zuo, "Neighborhood Component Feature Selection for High-Dimensional Data," Journal of Computers, vol. 7, no. 1, pp. 161-168, 2012.