Low-Resolution Face Recognition via Simultaneous Discriminant Analysis

Changtao Zhou   Zhiwei Zhang   Dong Yi   Zhen Lei   Stan Z. Li*
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
Beijing, China
zct.bjtu@gmail.com

Abstract

Low resolution (LR) is an important issue when handling real world face recognition problems. The performance of traditional recognition algorithms will drop drastically due to the loss of facial texture information in original high resolution (HR) images. To address this problem, in this paper we propose an effective approach named Simultaneous Discriminant Analysis (SDA). SDA learns two mappings from LR and HR images respectively to a common subspace where discrimination property is maximized. In SDA, (1) the data gap between LR and HR is reduced by mapping into a common space; and (2) the mapping is designed for preserving most discriminative information. After that, the conventional classification method is applied in the common space for final decision. Extensive experiments are conducted on both FERET and Multi-PIE, and the results clearly show the superiority of the proposed SDA over state-of-the-art methods.

1. Introduction

During the past decades, face recognition has received great interest and achieved impressive success, especially in applications under controlled environment. However, the performance of face recognition in uncontrolled environment is still far from being satisfactory. In this paper we focus on the low resolution face recognition problem. Low resolution issue is common in real world applications, for instance, in surveillance based face recognition faces are frequently captured at a distance, making the images relative very small and unclear. Face recognition based on LR images is a very challenging problem because compared with high-resolution (HR) images, LR ones lose much facial information (e.g. skin texture) which is crucial for high recognition accuracy.

To deal with this problem, there are usually three existing

---

*Stan Z. Li is the corresponding author

categories. One category is to extract the invariant discriminant information from low-resolution images directly. Su *et al.* [16] utilized FFT technique to extract the global structure information of face image for recognition. Li *et al.* [10] proposed coupled metric learning to transform the high and low resolution images into a feature space to be classified. These methods usually require the images to be well aligned and their performance may decline due to misalignment. Some approaches are based on low resolution invariant local features, such as, local phase quantization(LPQ)[14] and local frequency descriptor(LFD)[7], *etc*. In [1], LPQ is designed based on an assumption on point spread function (PSF) that in low frequency the phase value of LR and HR is unchanged. However, such assumption is not always the case in real world applications. Inspired by LPQ, Zhen Lei *et al.* [7] proposed LFD by quantizing the relative local pixel relationship in low-frequency domain instead of absolute value so that the assumption on PSF is no longer needed. However, these blur-invariant descriptors cannot achieve very high accuracy when the images are of very low resolution.

The second category to deal with this problem is "two-step" based methods which first synthesize the HR images from a single or multiple LR ones by super-resolution(SR) techniques and then the traditional algorithms could be applied consequently. Freeman *et al.* [3] proposed a patch-wise Markov network learned from the training set as the SR prediction model. Liu *et al.* [13] proposed to integrate a holistic model and a local model for SR reconstruction. We refer readers to [17] for other methods. Unfortunately, most of these SR algorithms are not designed for recognition, but for visual enhancement of images.

Recently, a series of works [18, 6, 11] have been proposed to avoid explicit SR procedure in the image domain but to learn a mapping from LR image to HR image in training set and then apply it to testing set. The approach which performed SR reconstruction in the eigenface domain has been investigated in [5]. P. Hennings-Yeomans *et al.* [18] propose a joint objective function that integrates the aims

1

of super-resolution and face recognition. H. Huang and H.t. He [6] proposed to perform a nonlinear mapping between HR/LR PCA features via radial basis functions (RBFs) in a coherent space obtained by Canonical correlation analysis (CCA). B. Li [11] developed an objective function that agrees with CCA by introducing a penalty weighting matrix to maintain locality. Compared with two-step methods, these approaches improved the recognition rate. However, most of these algorithms are unsupervised, discarding the sample labels which are crucial for classification.

In this paper, we propose Simultaneous Discriminant Analysis to learn a couple of mappings from LR and HR face images respectively to a common subspace, where the mapping function is designed on the idea of the classical discriminant analysis [2]. Then the conventional classification methods be applied. By mapping into the common space, the data gap between LR and HR faces are reduced, and therefore can achieve better classification results.

The remainder of this paper is organized as follows. Section 2 describes the problem formulation and the proposed algorithm. Section 3 discusses the relations to the most related methods. Experimental results on some databases are given in Section 4. Finally in Section 5, we conclude the paper.

2. Simultaneous Discriminant Analysis(SDA)

2.1. Problem Formulation

In this work, we suppose the working scenario as follows. HR face images are enrolled, and LR ones are used to test. Our approach firstly attempts to project the data points in the original HR and LR image subspaces into a third common subspace by a couple of mappings: one for LR vectors, $f_L : \Re^m \mapsto \Re^d$, the other for HR ones, $f_H : \Re^M \mapsto \Re^d$. m, M and d represents the dimensions of the original LR, HR and the third subspaces respectively. Then some distance metric can be computed in the new feature space by

$$d_{ij} = D(f_L(l_i), f_H(h_j)) \qquad (1)$$

where l_i and h_j respectively denotes the i th and j th sample of LR and HR face image sets. Then classification can be performed consequently. Here we prefer linear mapping functions. So we assume that two matrices P_L and P_H with sizes of $m \times d$ and $M \times d$ specify the mapping functions as $f_L(x_L) = P_L^T x_L$ and $f_H(x_H) = P_H^T x_H$, respectively.

The similarity measure of equation (1) indicates that the LR and HR samples should be treated equally after they are projected into the common subspace though they are endowed with quite different discriminative information in the original spaces. Therefor, to enhance the discriminative power of the corresponding projection functions, it is natural to integrate the label information into the new common space. Based on the principle of discriminant analysis, we

propose the objective to be

$$J(\theta_H, \theta_L) = \frac{J_b(\theta_H, \theta_L)}{J_w(\theta_H, \theta_L)} \qquad (2)$$

where $J_b(\theta_H, \theta_L)$ and $J_w(\theta_H, \theta_L)$ are measures of between-class and within-class scatters in the common space which are obtained by means of class membership of samples regardless of their original identities of LR or HR. θ_H and θ_L denote the parameters to be learned. By maximizing the objective, the learning of projection functions can be converted to a generalized eigen problem that can be solved efficiently, and is illustrated in the following subsection.

2.2. Algorithm

For convenience, we denote the two involved mappings from LR and HR face image sets to the common subspace as $[l_1^1, \cdots, l_{N_{L_1}}^1, \cdots, l_{N_{L_c}}^C] \mapsto [y_1^1, \cdots, y_{N_{L_1}}^1, \cdots, y_{N_{L_c}}^C]$ and $[h_1^1, \cdots, h_{N_{H_1}}^1, \cdots, h_{N_{H_c}}^C] \mapsto [z_1^1, \cdots, z_{N_{H_1}}^1, \cdots, z_{N_{H_c}}^C]$. where N_{L_i} and N_{H_j} respectively denote the number of samples in the i th class of LR and the j th class of HR face subsets; l_i^c is the i th LR sample of class c, and $h_j^{c'}$ is the j th HR sample of class c'. The above mapping relationship can be concisely formulated as $Y = P_L^T L$ and $Z = P_H^T H$.

To facilitate the subsequent derivation, we firstly compute the means of samples in each class of LR and HR face subsets as

$$\mu_{L_i} = \frac{1}{N_{L_i}} \sum_{j=1}^{N_{L_i}} l_j^i, \quad \mu_{H_i} = \frac{1}{N_{H_i}} \sum_{j=1}^{N_{H_i}} h_j^i \qquad (3)$$

and the total means of LR and HR faces as

$$\mu_L = \frac{1}{N_L} \sum_{i=1}^{C} \sum_{j=1}^{N_{L_i}} l_j^i, \quad \mu_H = \frac{1}{N_H} \sum_{i=1}^{C} \sum_{j=1}^{N_{H_i}} h_j^i \qquad (4)$$

where $N_L = N_{L_1} + N_{L_2} + \cdots + N_{L_C}$ and $N_H = N_{H_1} + N_{H_2} + \cdots + N_{H_C}$ are the total numbers of samples over all classes corresponding to LR and HR face image sets, respectively.

The final aim of the algorithm is to find the optimal projective matrixes (or directions) respectively and simultaneously for LR and HR images that gather the samples from the same class and meanwhile disperse the samples from different ones. To avoid to be of verbosity, we now consider the case when the projection is of one dimension which means that P_L and P_H are vectors. We also use P_L and P_H to denote the one-dimensional projection directions separately for LR and HR face images.

Under the projections of P_L and P_H, the mean of each class and the total mean of all samples in the common subspace can be given by

$$\mu_i = \frac{1}{N_i} \left(\sum_{j=1}^{N_{Hi}} z_j^i + \sum_{j=1}^{N_{Li}} y_j^i \right) = \alpha_i P_H^T \mu_{H_i} + \beta_i P_L^T \mu_{L_i} \qquad (5)$$

$$\mu = \frac{1}{N}(\sum_{i=1}^{C}\sum_{j=1}^{N_{Hi}} z_j^i + \sum_{i=1}^{C}\sum_{j=1}^{N_{Li}} y_j^i) = \alpha P_H^T \mu_H + \beta P_L^T \mu_L \quad (6)$$

Where $N_i = N_{H_i} + N_{L_i}$ is the number of samples in the i th class including both LR and HR ones projected to the common space by P_L and P_H, and $N = \sum_{i=1}^{C} N_i = N_H + N_L$ represent the total number of samples.

If we define the mutual ratios of the numbers of LR and HR samples belonging to the i th class as α_i and β_i and the ones of the total numbers of LR and HR samples as α and β which are the following

$$\alpha_i = \frac{N_{H_i}}{N_{H_i} + N_{L_i}}, \quad \beta_i = \frac{N_{L_i}}{N_{H_i} + N_{L_i}} \quad (7)$$

$$\alpha = \frac{N_H}{N_H + N_L}, \quad \beta = \frac{N_L}{N_H + N_L} \quad (8)$$

and note equation (3), (5) and (7), the within-class scatter can be easily computed by

$$J_{w_i}(\theta_H, \theta_L) = \sum_{j=1}^{N_{H_i}}(z_j^i - \mu_i)^2 + \sum_{j=1}^{N_{L_i}}(y_j^i - \mu_i)^2$$
$$= \begin{bmatrix} P_H^T & P_L^T \end{bmatrix} \begin{bmatrix} C_{HH}^i & C_{HL}^i \\ C_{LH}^i & C_{LL}^i \end{bmatrix} \begin{bmatrix} P_H \\ P_L \end{bmatrix}$$
$$= P^T C^i P$$
$$(9)$$

where $P = [P_H^T P_L^T]^T$ is the concatenation of the projectors P_H and P_L. The blocks of matrix C^i are

$$C_{HH}^i = \sum_{j=1}^{N_{H_i}}(h_j^i - \alpha_i \mu_{H_i})(h_j^i - \alpha_i \mu_{H_i})^T + \alpha_i^2 N_{L_i}\mu_{H_i}\mu_{H_i}^T$$
$$C_{HL}^i = -2\beta_i(1-\alpha_i)N_{H_i}\mu_{H_i}\mu_{L_i}^T$$
$$C_{LH}^i = -2\alpha_i(1-\beta_i)N_{L_i}\mu_{L_i}\mu_{H_i}^T$$
$$C_{LL}^i = \sum_{j=1}^{N_{L_i}}(l_j^i - \beta_i \mu_{L_i})(l_j^i - \beta_i \mu_{L_i})^T + \beta_i^2 N_{H_i}\mu_{L_i}\mu_{L_i}^T$$
$$(10)$$

Then summing within-class scatters over all classes, we get the total within-class scatter

$$J_w(\theta_H, \theta_L) = \frac{1}{N}\sum_{i=1}^{C} J_{w_i} = P^T(\frac{1}{N}\sum_{i=1}^{C} C^i)P = P^T S_w P \quad (11)$$

among which $S_w = \sum_{i=1}^{C} C^i$, is the counterpart of **Within-Class Scatter Matrix** in discriminant analysis[2].

Similarly, by substituting equations(3), (4), (5), (6), (7) and (8) into the following, the between-class scatter is formulated as

$$J_b(\theta_H, \theta_L) = \frac{1}{N}\sum_{i=1}^{C} N_i(\mu_i - \mu)^2$$
$$= \begin{bmatrix} P_H^T & P_L^T \end{bmatrix} \begin{bmatrix} S_{HH} & S_{HL} \\ S_{LH} & S_{LL} \end{bmatrix} \begin{bmatrix} P_H \\ P_L \end{bmatrix} \quad (12)$$
$$= P^T S_b P$$

Where, the sub-matrixes of S_b are

$$S_{HH} = \frac{1}{N}\sum_{i=1}^{C} N_i(\alpha_i \mu_{H_i} - \alpha\mu_H)(\alpha_i \mu_{H_i} - \alpha\mu_H)^T$$

$$S_{HL} = \frac{1}{N}\sum_{i=1}^{C} N_i(\alpha_i \mu_{H_i} - \alpha\mu_H)(\beta_i \mu_{L_i} - \beta\mu_L)^T$$

$$S_{LH} = \frac{1}{N}\sum_{i=1}^{C} N_i(\beta_i \mu_{L_i} - \beta\mu_L)(\alpha_i \mu_{H_i} - \alpha\mu_H)^T$$

$$S_{LL} = \frac{1}{N}\sum_{i=1}^{C} N_i(\beta_i \mu_{L_i} - \beta\mu_L)(\beta_i \mu_{L_i} - \beta\mu_L)^T$$

$$(13)$$

By comparing with LDA[2], we can see that S_b is equivalent to the so-called **Between-Class Scatter Matrix**.

Substituting (11) and (12) into (2), the objective becomes to

$$\arg\max_P J(\theta_H, \theta_L) = \frac{J_b(\theta_H, \theta_L)}{J_w(\theta_H, \theta_L)} = \frac{P^T S_b P}{P^T S_w P} \quad (14)$$

Which can be reduced to solve the generalized eigen problem

$$S_w^{-1} S_b P = P\Lambda \quad (15)$$

And the eigenvectors corresponding to the first d largest eigenvalues of $S_w^{-1} S_b$ could constitute the expected transformation matrixes, e.g., P_L and P_H(here, P_L and P_H are all projection matrixes). Additionally, from equation (10) we can deduce that α_i should be equal to β_i to ensure C^i to be symmetric so that the symmetry of S_w be guaranteed. This means that each class contains equal number of LR and HR face samples in training sets which is usually satisfiable in real applications because LR training samples can often be obtained by blurring and down-sampling HR ones.

2.3. Regularization

Over-fitting can be avoided to some extent by regularizing the objective in equation (14). In the framework proposed above, it is quite an easy and natural thing to incorporate some prior information into the objective function which could boost the algorithm's performance further.

Firstly, we consider the constraint of **Local consistency**[19]. In fact, a regularization of local consistency can be abstracted as minimizing a weighted cost when the corresponding objective is being pursued, which is computed over all projected samples. Here we give a generic form of risk as

$$Re = \frac{\gamma}{2}\sum_{i=1}^{N_H}\sum_{j=1}^{N_H}(P_H^T h_i - P_H^T h_j)^2 v_{ij}$$
$$+ \frac{\eta}{2}\sum_{i=1}^{N_L}\sum_{j=1}^{N_L}(P_L^T l_i - P_L^T l_j)^2 m_{ij} \quad (16)$$
$$+ \psi\sum_{i=1}^{N_H}\sum_{j=1}^{N_L}(P_H^T h_i - P_L^T l_j)^2 w_{ij}$$

to be minimized. With a little of manipulation of matrix, the above can be rewritten as

$$
\begin{aligned}
Re &= \gamma P_H^T H (D_V - V) H^T P_H + \eta P_L^T L (D_M - M) L^T P_L \\
&\quad + \psi (P_H^T H W_R H^T P_H + P_L^T L W_C L^T P_L - 2 P_H^T H W L^T P_L) \\
&= \begin{bmatrix} P_H^T & P_L^T \end{bmatrix} \begin{bmatrix} R_{HH} & R_{HL} \\ R_{LH} & R_{LL} \end{bmatrix} \begin{bmatrix} P_H \\ P_L \end{bmatrix} \\
&= P^T R P
\end{aligned}
\tag{17}
$$

Where $V(v_{ij})$ and $M(m_{ij})$ can encode some constraints like *local smoothing* in HR and LR image sets respectively. And constraints that resorts to labels from inter-sets can be absorbed by $W(w_{ij})$. Specifically, D_V is a diagonal matrix with diagonal items $D_{V_{ii}} = \sum_{j=1}^{N_H} v_{ij}$. D_M has the similar meaning in terms of M. And W_R and W_C are also diagonal matrixes but $W_{R_{ii}} = \sum_{j=1}^{N_L} w_{ij}$ and $W_{C_{ii}} = \sum_{i=1}^{N_H} w_{ij}$. W is usually not symmetric unless the cost enforced on h_i and l_j is the same as the one enforced on l_i and h_j. The sub-matrixes of R are computed as follows

$$
\begin{aligned}
R_{HH} &= \gamma H (D_V - V) H^T + \psi H W_R H^T \\
R_{HL} &= -\psi H W L^T = R_{LH}^T \\
R_{LL} &= \eta L (D_M - M) L^T + \psi L W_C L^T
\end{aligned}
\tag{18}
$$

On the other hand, if we let alone the specific definitions of the sub-matrixes of R and replace R_{HH} and R_{LL} with some Laplacian matrixes and let R_{HL} and R_{LH} equal to zero, the above cost will enforce some kind of constraints on the projectors when it is being minimized, just like the contextual one in [9]. Specially, when R_{HH} and R_{LL} are some scalar matrixes, say $R_{HH} = \gamma I$ and $R_{LL} = \eta I$, the cost will convert to the conventional l_2-norm regularization on P. And this is also the form adopted in all the experiments below.

Combing equation (14) and (17), we get the regularized version of Simultaneous Discriminant Analysis (SDA)

$$
\arg\max_P J(\theta_H, \theta_L) = \frac{P^T S_b P}{P^T (S_w + R) P}
\tag{19}
$$

It can also be solved by a generalized eigen problem

$$
(S_w + R)^{-1} S_b P = P \Lambda
\tag{20}
$$

3. Difference with Most Related Methods

The most related methods to the proposed algorithm include Coupled Spectral Regression (CSR)[8] and Common Discriminant Feature Extraction (CDFE) [12]. CSR was introduced to deal with the problem of maching heterogeneous faces (NIR versus VIS). It's motivation is also to collaboratively learn a couple of mapping functions that are of discriminant. But the assumption incorporated into CSR is that the directions of the expected projectors should be similar to each other which is dubious in many problems. And

another rigid limit resulted from its assumption is that CSR can not be used to tackle problems among which the two involed sets are of different dimension (e.g., LR vs. HR) directly. But in the proposed method, there is no such requirement and limit. And in CDFE, class membership is only exploited by a means of crossing sets, whereas in the proposed method labels of samples are completely transferred to the common subspace and fully explored.

4. Experiments

We conduct our experiments on FERET face database[15] and Multi-PIE face database. In order to demonstrate the effectiveness of the proposed algorithm, we compare our method with a series of baselines in standard single-frame face recognition and the state-of-the-art algorithm CM/CLPM[11]. The baselines comprise of HR-LDA/PCA, which perform LDA/PCA in HR face image domain; BL-LDA/PCA, which recognizes the feature extracted from the interpolated image that is obtained by scaling a LR face to the size of HR one by bi-linear interpolation.CM bases on coupled mappings which project face images with different resolutions into a unified feature space that favors of classification.And CLPM is an extension of CM which introduces a locality-preserving constraint into its objective.

In all experiment face images are aligned by the coordinates of two eyes and normalized to zero mean and unit variance.And the algorithm of CM/CLPM used in below experiments is downloaded from the web-site of the author of [11].The parameters are set: $\alpha = 11$ and $N(i) = 1$.

4.1. FERET

The training set of our experiment on FERET contains 1002 frontal face images from 429 persons. The experimental results are obtained on the standard gallery (1196 images) and the probe set "*fafb*" (1195 images).The first five samples used are shown in Fig.1. In all experiment, the HR face images are generated by scaling the original ones to the size of 72×72. and then the LR images of size 12×12 are acquired by smoothing and down-sampling their corresponding HR ones. The regularization coefficients are empirically assigned to $\gamma = \eta = 10^{-5}$.

The recognition rates with different feature dimensions are shown in Fig.2(a). Our method (SDA) with 69-D features achieves the recognition rate of 92.7%. Whereas, the highest recognition rates obtained respectively by the compared algorithms are: HR-LDA 90.6%, CLPM 89.5%, CM 78.1%, BL-LDA 71.6%, HR-PCA 68.1%, BL-PCA 57.9%, with feature dimensions of 100, 80, 144, 150, 457 and 176. Which indicates that the proposed method outperforms the state-of-the-art algorithm CLPM as well as HR-LDA. This is really a surprise. However, the result is also quite reasonable because SDA combines the discriminant information

Figure 1. Face images in FERET face database.(a) HR trainging faces with size 72×72.(b)LR face images for training with size 12×12.(c)LR testing face images with size 12×12.
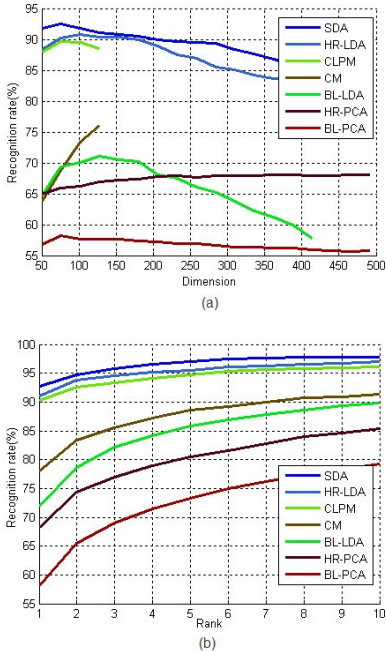

Figure 2. Rank-1 and cumulative recognition rates on FERET face database.

from both HR and LR sets optimally to some extent which more favor of classification. And another explanation to this results perhaps lies in the property of FERET database itself as discussed in the next section.The cumulative recognition rates are illustrated in Fig.2(b).

4.2. Multi-PIE

In this experiment, influences of complexly-interacted factors such as face poses, expressions and illumination variations to the proposed method are examined with a subset of session one of Multi PIE face database [4].Specifically, a total of 20,000 faces from the first 100

subjects with five poses(13_0, 14_0, 05_1, 05_0 and 04_1) individually(As is illustrated in Fig.3), each of which has two expressions that are seperately captured under twenty kinds of illumination conditions are extracted to form our training set. The gallery set contains only one frontal face image from the other 149 individuals, which has neural expression and a nearly ideal illuminaion(10). And the left 29,651 face images that belong to the same pose subsets and have the same combinations of expressions and illumination variations as the training ones but from the the other 149 registered subjects constitute the probe set. Just like in sec.4.1, we also resort to l_2-norm to regularize the objective, and the related coefficients are assigned to $\gamma = \eta = 10^{-6}$.

The HR and LR samples are obtained by smoothing and down-sampling the original face images to the resolutions of 48×48 and 12×12, respectively. The Rank-1 and cumulative recognition rates are illustrated in Fig.4(a) and (b). From Fig.4(a), we can see that our proposed method SDA achieves a pretty high recognition rate of 69.5% with 25-D features. Which greatly outperforms most of the borderlines (CLPM 52.1%, CM 14.6%, HR-PCA 17.5%, BL-PCA 2.6%, BL-LDA 3.9%) and is lower than but quite comparable to the result of HR-LDA algorithm, 77.2% in 24-D subspace.

By comparing Fig.2 and Fig.4, we can find out that the performances of the same algorithms on the FERET are superior to the ones obtained on Multi-PIE. The main reasons except the specific sizes of the two databases, we consider, lie in two aspects. The first one is that the identities in the standard training, gallery and probe sets of FERET on are not independent of each other which means that they are overlapped with a quite portion of sample IDs. Whereas, in our experiments on Multi-PIE, we have intended to avoid such circumstances and evaluated algorithms objectively. The second one is there are too many factors, for example, poses, expressions and illumination variations, and their interactions in Multi-PIE that could decline the performance of any algorithm. However, both experiments demonstrate that class-membership information maintain quite a degree of consistency to contribute to classification. And the results comparable to HR-LDA obtained by the proposed method justify our conjecture and motivation.

5. Conclusion

In this paper, we have formulated a method named simultaneous discriminant analysis (SDA) to deal with lowresolution face recognition problem. SDA learns two mappings for LR and HR face images respectively into the common space, where the discrimination is maximized for classification. Extensive experiments are conducted on FERET and Multi-PIE database, which proves its efficiency and superiority to previous methods.
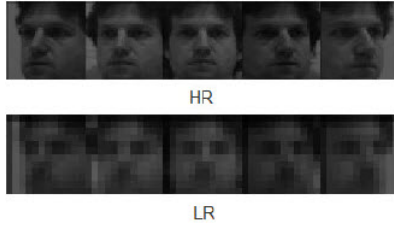
Figure 3. Face image samples from Multi-PIE database, left to right correspond to Pose 13 0, 14 0, 05 1, 05 0 and 04 1.
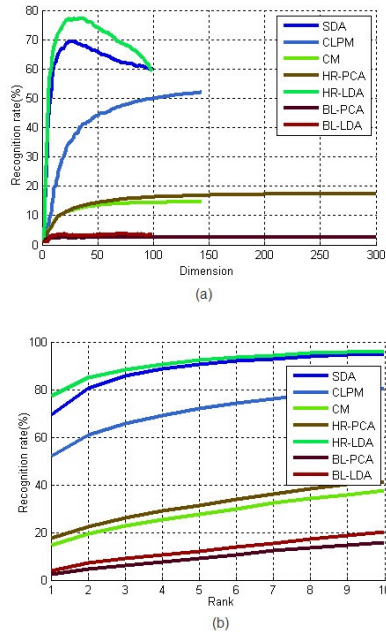

Figure 4. Rank-1 and cumulative recognition rates on Multi-PIE face database.

ACKNOWLEDGMENT

References

[1] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008. 1

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997. 2, 3

[3] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000. 1

[4] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Guide to the cmu multi-pie database. 2007. 5

[5] B. Gunturk, A. Batur, Y. Altunbasak, M. H. III, and R. Mersereau. Eigenface-domain super-resolution for face recognition. *Image Processing IEEE Transactions on*, 12(5):597–606, 2003. 1

[6] H. Huang and H. He. Super-resolution method for face recognition using nonlinear mappings on coherent features. *Neural Networks, IEEE Transactions on*, (99):1–10, 2010. 1, 2

[7] Z. Lei, T. Ahonen, M. Pietikainen, and S. Z. Li. Local frequency descriptor for low-resolution face recognition. *In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2011. 1

[8] Z. Lei and S. Z. Li. Coupled spectral regression for matching heterogeneous faces. *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*, 2009. 4

[9] Z. Lei and S. Z. Li. Contextual constraints based linear discriminant analysis. *Pattern Recognition Letters*, 32:626–632, 2011. 4

[10] B. Li, H. Chang, S. Shan, and X. Chen. Coupled metric learning for face recognition with degraded images. *Advances in Machine Learning*, pages 220–233, 2009. 1

[11] B. Li, H. Chang, S. Shan, and X. Chen. Low-resolution face recognition via coupled locality preserving mappings. *Signal Processing Letters, IEEE.*, 17(1):20–23, 2010. 1, 2, 4

[12] D. Lin and X. Tang. Inter-modality face recognition. *Computer Vision, ECCV 2006*, pages 13–26, 2006. 4

[13] C. Liu, H. Shum, and C. Zhang. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 192–198, 2001. 1

[14] V. Ojansivu and J. Heikkil. Blur insensitive texture classification using local phase quantization. *Image and Signal Processing*, pages 236–243, 2008. 1

[15] P. Philips, H. M. P. Pauss, and S. Rivzvi. The feret evaluation methodology for face-recognition algorithms. *IEEE Conf. Computer Vision and Pattern Recognition*, page 137C143, 1997. 4

[16] Y. Su, S. Shan, X. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 0:1–8, 2007. 1

[17] J. van Ouwerkerk. Image super-resolution survey. *Image and Vision Computing*, 24(10):1039–1052, 2006. 1

[18] P. H. Yeomans, S. Baker, and B. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. pages 1–8, 2008. 1

[19] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems 16, Proceedings of the 2003 Conference*, pages 595–602, 2004. 3