Coupled Metric Learning for Face Recognition with Degraded Images

Bo Li[1], Hong Chang[2,3], Shiguang Shan[2,3], and Xilin Chen[2,3]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[2] Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (CAS), Beijing, 100190, China
[3] Institute of Computing Technology, CAS, Beijing, 100190, China
{bli,hchang,sgshan,xlchen}@jdl.ac.cn

Abstract. Real-world face recognition systems are sometimes confronted with degraded face images, e.g., low-resolution, blurred, and noisy ones. Traditional two-step methods have limited performance, due to the disadvantageous issues of inconsistent targets between restoration and recognition, over-dependence on normal face images, and high computational complexity. To avoid these limitations, we propose a novel approach using coupled metric learning, without image restoration or any other preprocessing operations. Different from most previous work, our method takes into consideration both the recognition of the degraded test faces as well as the class-wise feature extraction of the normal faces in training set. We formulate the coupled metric learning as an optimization problem and solve it efficiently with a closed-form solution. This method can be generally applied to face recognition problems with various degrade images. Experimental results on various degraded face recognition problems show the effectiveness and efficiency of our proposed method.

1 Introduction

1.1 Degraded Face Recognition

During the past decades, face recognition has attracted a great deal of attention in the filed of pattern recognition and computer vision. Various face recognition algorithms have been proposed. An overview of popular methods is given in [1]. Unfortunately, the performance of face recognition systems will decrease in front of serious degraded test images, such as low-resolution images of only 12×12 pixels. Opposite to the *degraded images*, let us call the undegraded (e.g., clear, high-resolution and undamaged) ones as *normal images*.

Traditional methods for degraded face recognition usually take "two steps": image processing and recognition. Obviously, it is not a good choice to directly pass the degraded test images to the face recognition system enrolled with normal images, due to the sacrifice of important image details. A more commonly used way is to first restore the degraded images into normal ones. This kind of methods attempt to retrieve the loss

Z.-H. Zhou and T. Washio (Eds.): ACML 2009, LNAI 5828, pp. 220–233, 2009.

details of the degraded images. For this purpose, a lot of super-resolution and deblurring algorithms have been proposed, either for single image [2,3,4,5,6] or multi images [7,8,9]. Among these methods, learning-based super-resolution methods attract a lot of attention, such as [2,4,6,10]. More specific methods for face images have also been studied. Baker and Kanade [11] propose "hallucination" technique for face image super resolution. This learning-based method infers the high-resolution (HR) face images from an input low-resolution (LR) image using the face priors learned from training set. Liu et al. [12] propose a statistical modeling approach that integrates a global parametric model and a local nonparametric one.

However, there exist some limitations on the two-step methods. The main problem lies in that the preprocessing step only aims at minimizing the reconstruction error between restored images and the ground-truth, without any consideration of the subsequent recognition target. The restoration process with inconsistent target cannot always improve the performance of face recognition as much as we expected. Besides, the sophisticated super-resolution and deblurring algorithms used in this step are usually computational complex, which is not ideal for some real-time systems.

1.2 Recent Related Work

Recently, some algorithms without restoring the test image in the first step have been proposed, especially for low-resolution face recognition. Instead of reconstructing a HR image explicitly, the method proposed in [13] carries out the reconstruction in the Eigen-face space and only outputs the weights along the principal components for face recognition purpose. Jia and Gong [14] propose a tensor-based method to simultaneously super-resolve and recognize LR face images. Face images with different poses, expressions and illuminations are modeled in tensor space, where HR images in multi-modal can be estimated from the corresponding LR images in single-modal. This method is a tensor extension of the super-resolution problem for the single-modal LR face image. More recently, some researchers propose a new algorithm which could simultaneously carry out super-resolution and feature extraction for LR face recognition [15]. This method designs a new objective function which integrates the constraints for reconstruction and classification together. In face recognition tasks, the method requires to optimize the object function for each test image. Although the authors propose a speed up procedure, the computational complexity is still high, especially for real-world real-time applications.

Despite the recent progresses, most existing methods for degraded face recognition have limitations in the following aspects: (1) Inconsistent targets in restoration and recognition. Minimizing the reconstruction error in image restoration may not always guarantee good performance of the subsequent face recognition, even if the reconstruction is not explicitly performed [13,14]. (2) From empirical studies, normal images in the training (and gallery) set are not always good for recognition. On the contrary, high-frequency details and background noise in the normal images may decrease the recognition performance. (3) Computational complexity. Many super-resolution and deblurring methods, adopted in the preprocessing step, are time-consuming. Even the simultaneous method [15] is not efficient enough.

1.3 This Paper

To overcome the limitations stated above, we propose a new simple and elegant way to solve the problem of degraded face recognition. Different from previous methods, we address the problem through coupled metric learning which actually decides two transformations. One transformation maps degraded images to a new subspace, where higher recognition performance can be achieved; the other one maps normal images and class labels together to the same subspace for better class-wise feature representation. The coupled transformations are jointly determined by solving an optimization problem. The optimization procedure is fairly efficient with closed-form solutions.

Compared with some previous related methods, our work contributes in the following aspects. First, it is novel to learn coupled metrics, which takes into consideration both degraded face recognition and normal class-wise feature extraction. Second, it is very simple and efficient. Our method does not require any preprocessing (e.g., super-resolution or deblurring). The metric learning step is carried out in the off-line training phase. Third, it can be applied to face recognition with various degraded images as well as more general degraded object classification problems. Note that metric learning methods for classification has been studied by many researchers [16,17,18,19]. However, only a few methods try to learn coupled metrics and none of them has been successfully applied in the degraded face recognition problems mentioned above.

The rest of this paper is organized as follows. We propose the coupled metric learning method in Section 2, where problem formulation and optimization are presented in detail. Section 3 summarizes the algorithm for degraded face recognition and describes some implementation details. We then apply our method to several degraded face recognition problems in Section 4. Some discussions on nonlinear extension and other possible applications in given Section 5. Finally, Section 6 concludes the whole paper.

2 Coupled Metric Learning

2.1 Problem Formulation

Denote the feature vector of a degraded test image as $\mathbf{x}^p \in \mathcal{R}^m$. The training set consists of N normal facial images $\mathbf{y}_i \in \mathcal{R}^M$ and their degraded counter-parts $\mathbf{x}_i \in \mathcal{R}^m$, $i = 1, \ldots, N$. Suppose there are C classes. Let $\mathbf{t}_i$ denote the C-dimensional class indicator vector of $\mathbf{y}_i$ ($\mathbf{x}_i$), which can be seen as one column vector from a $C \times C$ identity matrix. The features of degraded and normal face images, $\mathbf{x}^p$, $\mathbf{x}_i$ and $\mathbf{y}_i (i = 1, \ldots, N)$, are usually constructed by concatenation of pixel values. m and M denote the dimensionalities of these feature vectors.

Let us represent the training set in the form of $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$, where $\mathbf{X}$ and $\mathbf{Y}$ are matrices with the sizes of $m \times N$ and $M \times N$, respectively. The labels of the training set can be represented using the class indicator matrix $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]$. Actually, the degraded and normal face images form subspaces $\mathcal{X} \subset \mathcal{R}^m$ and $\mathcal{Y} \subset \mathcal{R}^M$. The label space $\mathcal{T}$ is spanned by $\{\mathbf{e}_1, \ldots, \mathbf{e}_C\}$, where C is the class number and $\mathbf{e}_i$ is a unit vector with only one non-zero value at the i-th entry. For example, $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$ is a vector with size of $C \times 1$. Consequently, $\mathbf{t}_i = \mathbf{e}_j$ means that the i-th sample belongs to the j-th class.

Different from traditional methods, which recognize $\mathbf{x}^p$ by first reconstructing its normal counter-part, we address this challenging problem in the metric learning context.

2.2 Metric Learning in Coupled Spaces

The basic idea of our method is to learn coupled metrics which in fact map $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{T}$ to a joint new subspace $\mathcal{Z} \subset \mathcal{R}^d$, where the new distance measure is more ideal for face recognition.

On one hand, we try to learn a metric from degraded training images, $\mathbf{x}_i, i = 1, \ldots, N$, to get better class-separability. To this end, we define the linear transformation in matrix form as follows:

$$\widetilde{\mathbf{X}} = \mathbf{A}\mathbf{X} = [\mathbf{A}\mathbf{x}_1, \ldots, \mathbf{A}\mathbf{x}_N], \tag{1}$$

where $\widetilde{\mathbf{X}}$ represent the new features in the d-dimensional transformed space $\mathcal{Z}$. In this way, the $d \times m$ matrix $\mathbf{A}$ maps the degraded features to a new space where face recognition is actually performed. We call the transformation matrix $\mathbf{A}$ *recognition matrix*.

On the other hand, we expect to extract better features from normal training images, $\mathbf{y}_i, i = 1, \ldots, N$, and preserve the class label information at the same time. More specifically, we define the linear transformation on $\mathbf{Y}$ and $\mathbf{T}$ as

$$\widetilde{\mathbf{Y}} = \mathbf{F}\hat{\mathbf{Y}} = \mathbf{F}[\mathbf{Y}^T \alpha \mathbf{T}^T]^T, \tag{2}$$

where $\mathbf{F}$ is the $d \times (M + C)$ transformation matrix. The transformation decided by $\mathbf{F}$ results in a new class-wise feature representation in the feature space $\mathcal{Z}$, we thus name it as *feature matrix*. For mathematical conciseness, we use $\hat{\mathbf{Y}} = [\mathbf{Y}^T \alpha \mathbf{T}^T]^T$ to denote the expanded normal training matrix in the following context. α is a scaling parameter. To estimate $\mathbf{A}$ and $\mathbf{F}$ defined above, we formulate the coupled metric learning as an optimization problem. The objective function to be minimized is defined as

$$J(\mathbf{A}, \mathbf{F}) = \|\widetilde{\mathbf{X}} - \widetilde{\mathbf{Y}}\| = \|\mathbf{A}\mathbf{X} - \mathbf{F}\hat{\mathbf{Y}}\|, \tag{3}$$

with the constraints of $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T = \widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^T = \mathbf{I}$. $\mathbf{I}$ denotes the identity matrix and $\|\cdot\|$ the Frobenius norm. Note that $\mathbf{X}$ and $\hat{\mathbf{Y}}$ are first centered by extracting the means from the vectors involved.

The optimization procedure is similar with that of Canonical correlation analysis (CCA) [20,21]. Let us consider the simplest case when we map $\mathbf{X}$ and $\hat{\mathbf{Y}}$ to 1-dimensional target space $\mathcal{Z}$. The object function becomes

$$J(\mathbf{a}, \mathbf{f}) = \|\mathbf{a}\mathbf{X} - \mathbf{f}\hat{\mathbf{Y}}\|, \tag{4}$$

subject to $\|\mathbf{a}\mathbf{X}\| = \|\mathbf{f}\hat{\mathbf{Y}}\| = 1$. Here, the recognition matrix $\mathbf{A}$ and the feature matrix $\mathbf{F}$ are reduced to row vectors $\mathbf{a}$ and $\mathbf{f}$ with sizes of m and $M + C$, respectively. Note that, in our procedure, d should be not larger than the minimum of m and $M + C$. Eqn. (4) is equal to the following maximization problem

$$\max_{\mathbf{a},\mathbf{f}} \mathbf{a}\mathbf{X}\hat{\mathbf{Y}}^T\mathbf{f}^T, \text{ s. t. } \mathbf{a}\mathbf{X}\mathbf{X}^T\mathbf{a}^T = \mathbf{f}\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T\mathbf{f}^T = 1. \tag{5}$$

The corresponding Lagrangian is

$$L(\mathbf{a}, \mathbf{f}, \lambda_1, \lambda_2) = \mathbf{a}\mathbf{X}\hat{\mathbf{Y}}^T\mathbf{f}^T - \lambda_1(\mathbf{a}\mathbf{X}\mathbf{X}^T\mathbf{a}^T - 1) - \lambda_2(\mathbf{f}\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T\mathbf{f}^T - 1). \quad (6)$$

Setting the derivatives of L w.r.t. $\mathbf{a}$ and $\mathbf{f}$ to 0's, we obtain

$$\mathbf{X}\hat{\mathbf{Y}}^T\mathbf{f}^T - \lambda_1\mathbf{X}\mathbf{X}^T\mathbf{a}^T = 0, \quad (7)$$

$$\hat{\mathbf{Y}}\mathbf{X}^T\mathbf{a}^T - \lambda_2\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T\mathbf{f}^T = 0. \quad (8)$$

Subtracting $\mathbf{a}$ times (7) from $\mathbf{f}$ times (8) and considering the constrains, we can finally have $\lambda_1 = \lambda_2 = \lambda$,

$$\mathbf{X}\hat{\mathbf{Y}}^T(\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T)^{-1}\hat{\mathbf{Y}}\mathbf{X}^T\mathbf{a}^T = \lambda^2\mathbf{X}\mathbf{X}^T\mathbf{a}^T, \quad (9)$$

and

$$\mathbf{f} = \frac{1}{\lambda}\mathbf{a}\mathbf{X}\hat{\mathbf{Y}}^T(\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T)^{-1}. \quad (10)$$

Therefore, we can obtain the sequence of $\mathbf{a}$'s by solving a generalized eigendecomposition problem (Equation (9)) and then get $\mathbf{f}$'s from Equation (10). $\mathbf{A}$ is simply constructed by piling the first d largest eigenvectors of Equation (9) up. According to $\mathbf{A}$, $\mathbf{F}$ could be constructed. Note that, $\mathbf{X}\mathbf{X}^T$ and $\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T$ is usually non-invertible. In this case, we carry out the regularization operations on them: $R(\mathbf{X}\mathbf{X}^T) = (\mathbf{X}\mathbf{X}^T + \kappa\mathbf{I})^{-1}$ and $R(\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T) = (\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T + \kappa\mathbf{I})^{-1}$, where κ is set to a small positive value (e.g., $\kappa = 10^{-6}$).

3 Degraded Face Recognition

Once we get the recognition matrix $\mathbf{A}$ and the feature matrix $\mathbf{F}$ through coupled metric learning, we may perform face recognition in the transformed new feature space. In this section, we present the overall algorithm and give some implementation details.

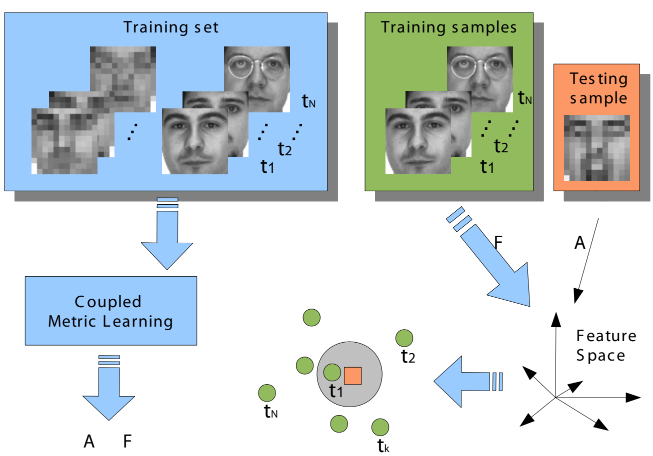


Fig. 1. An overview diagram for the algorithm

The overall algorithm is summarized in Figure 2. Note that the algorithm actually consists of two phases: an offline coupled metric learning phase and an online recognition phase. In the first phase, we learn the transformation matrices $\mathbf{A}$ and $\mathbf{F}$. In the second phase, degraded test image and the normal training images are jointly projected to a new feature space, using the learned transformations, respectively. The label of the test image is obtained using any distance-based classification method, such as k-nearest neighbor (kNN) classifier, support vector machines (SVMs), and so on. An overview diagram is shown in Figure 1 for better understanding the training and recognition procedure in our algorithm.

Input: Training degraded and normal face images and labels: $\mathbf{x}_i, \mathbf{y}_i, \mathbf{t}_i, i = 1, \ldots, N$
a degraded test face image: $\mathbf{x}^p$
Offline Coupled metric learning:
Learn coupled recognition matrix $\mathbf{A}$ and feature matrix $\mathbf{F}$
Online Recognition:
Transform the test image $\mathbf{x}^p$ to $\widetilde{\mathbf{x}}^p$: $\widetilde{\mathbf{x}}^p = \mathbf{A}\mathbf{x}^p$
Transform training images $\mathbf{y}_i$ to $\widetilde{\mathbf{y}}_i$ $(i = 1, \ldots, N)$: $\widetilde{\mathbf{y}}_i = \mathbf{F}\mathbf{y}_i$
Recognize $\widetilde{\mathbf{x}}^p$ from $\widetilde{\mathbf{y}}_i$
Output: Label $\mathbf{t}^p$ $(i = 1, \ldots, N)$

Fig. 2. Algorithm of degraded face recognition based on coupled metric learning

Note that our algorithm is ready to adopt the "probe-gallery" setting as well. Given a gallery set $(\mathbf{y}_i^g, \mathbf{t}_i^g), i = 1, \ldots, N_g$, with N_g being the number of normal face images in the set, we just need to transform these features to $\widetilde{\mathbf{y}}_i^g = \mathbf{F}\mathbf{y}_i^g$. The recognition is then performed on $\widetilde{\mathbf{x}}^p$ from $\widetilde{\mathbf{y}}_i^g$ $(i = 1, \ldots, N_g)$.

There are only two parameters involved in our proposed method. It is easy to set them according to their physical meanings. In our experiments, the dimensionality of the transformed target space d and the scaling parameter α are simply determined using several rounds of cross-validations. The kNN classifiers are used in the final classification step for face recognition.The kNN classifiers are used in the final classification step for face recognition.

4 Experiments

In this section, we access the efficacy of our proposed method for some real-world problems. The experiments are generally performed on face images with various degradations.

4.1 Face Database and Degraded Images

We make use of the AR face database [22], which consists of over $3,200$ color frontal view faces of 135 individuals. For each individual, there are 26 images, recorded in two sessions separated by two weeks and with different illumination, expressions, and facial disguises. In our experiment, we select 14 images for each individual with only illumination and expression changes. The image set from the first session (the former 7 images) is considered as training set and the other (the latter 7 images) as testing set.

The original images are cropped into 72×72 pixels and aligned according to the positions of two eyes. The corresponding serious degraded images are obtained as following procedures. The blurred images are generated by convoluting the normal images with point spread function (PSF). The larger the diameter of PSF, the more blurring the processed images. In the experiments, we use a "disk" kernel with diameter of 13 pixels as PSF. To get low resolution images of size 12×12, we take the operations of blurring and then down-sampling (with down-sampling factor equal to 6). We synthesize partially occluded images by setting the pixel values inside a square occlusion area to 0. Different sizes of occlusion areas are used in our experiments. Figure 3 shows an example face image from AR database and its counter-parts in low-resolution, blurring, and partial occlusion (from left to right in the figure).

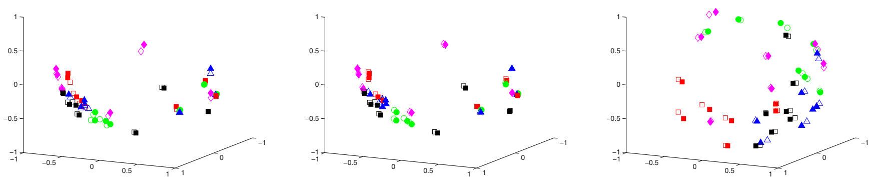
Fig. 3. A normal face image and the corresponding blurred one, low-resolution one, and occluded one

In our experiments, we normalize the columns of $\mathbf{X}$ and $\mathbf{Y}$ into the unit vectors. According to the direct sum kernels [23], we set a empirical value of $\alpha = 1$. Sometimes, this selection will affect the final recongnition performance. The dimensionality settings of the features for different kinds of dagraded face recognition could be found in Table 1.

We compare our proposed method with some traditional baseline methods. More specifically, approaches of the following four classes are included in our comparative study: (1) directly use the degraded test images for recognition; (2) directly do recognition with the ground-truth test images without degradation; (3) restore the degraded test images using different classical algorithms and then do recognition (two-step or restoration-based method); (4) learn coupled metric for recognition (CML-based method). In these comparison, the enrollments are normal images. Besides, we center the features before hand in all experiments.

4.2 An Illustrative Example

First of all, we give a preliminary comparison between these methods for blurred face recognition. In this experiment, we use the wiener filter algorithm [24] to estimate the deblurred images. Figure 4 shows some face samples of the first five persons embedded in three-dimensional space, where markers with the same type belong to the same person. The solid markers denote the normal samples while the hollow denote the degraded. From Figure 4(a), we can see that the blurred samples from different persons are mixed with each other. Figure 4(b) shows the results using deblurred samples, where



(a) (b) (c)

Fig. 4. Face samples of the first five persons embedded in 3-D space. (a) The blurred samples and the normal ones. (b) The deblurred samples and the normal ones. (c) The blurred samples and the normal ones after coupled metric learning. Here, the solid markers denote the normal samples while the hollow markers of the same type denote the degraded ones of the same person.

samples from the same person are closer, but samples from different persons are still difficult to discriminate. Finally, Figure 4(c) shows the results of coupled metric learning. We can see that the separability among different persons is increased. Intuitively, the situation in this subfigure will benefit classification. We also give the quantitative results of face recognition, as shown in Figure 5. Obviously, the performance of our method is significantly better than those of direct using degraded images and the restoration-based method. We can thus conclude that the restoration method could benefit recognition, while our proposed method is more outstanding for this purpose. Note that for visualization clarity, we select 35 samples from 5 persons in Figure 4(a)-(c), while the recognition rates in Figure 5 are computed using all samples.

Table 1. Dimensionality settings for the experiments of degraded face recognition on AR

Blurred Faces	Low-resolution Faces	Partially Occluded Faces
PCA: 945	PCA: 150	PCA: 250
LDA: 134	LDA: 134	LDA: 134
CML: 945	CML: 144	CML: 250

4.3 Face Recognition with Degraded Images

We further perform experiments on face recognition with degraded images, in order to demonstrate the effectiveness of the proposed algorithm and verify the discussions in previous sections. We compare different methods on face recognition tasks with degradation types including low resolution, blurring and partial occlusion, respectively.

On Blurred Faces. Firstly, let us test the coupled metric learning on the problem of face recognition with blurred test images. The blurred test images are generated from the normal test ones by convoluting with a "disk" PSF with diameter of 13 pixels. We apply two restoration-based methods by adopting Wiener filter [24] and Lucy-Richardson algorithm [25,26] in the deblurring step and compare these methods with ours.

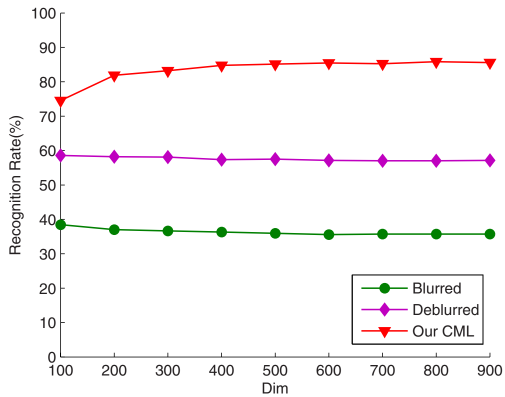


Fig. 5. The recognition rate curves for the illustrative experiments shown in Figure 4

As for the two traditional methods, we adopt two conventional face features, i.e., Eigen-faces (with PCA) [27] and Fisher-faces (with LDA) [28]. As for our method, we simply use pixel values as features. In the implementation, 945 Eigen-faces and 134 Fisher-faces are used for different methods. Notice that Fisher-faces are no more than 134 since the data set we used has 135 subjects. For our method, we use 945 Eigen-faces as the test features and original image intensities as the enrolled features.

The curves of cumulative recognition rates for different methods are shown in Figure 6. We compared our method with the different restoration and feature extraction methods which include: Normal-PCA (normal test images), Normal-LDA , Blurred-PCA (blurred test images), Blurred-LDA, Wiener-PCA (blurred test images restored by Wiener filter [24]), Wiener-LDA, Lucy-PCA (blurred test images restored by Lucy-Richardson algorithm [25,26]), Lucy-LDA, and Our CML (coupled metric learning method).

Our coupled metric learning method gets the best performance than all two-steps restoration-based methods. It archives the rank-1 recognition rate of 87.1%, even better than 85.6% of high-resolution LDA. From the blurred example shown in Figure 3, we can see that the blurring is very serious. Even though, our method still gets more satisfactory recognition results than normal situation. The reason may lie in that our metric learning method can find a feature space where the normal images, class information and the blurred ones are perfectly fused. Actually, multi-scale feature fusion is an effective scheme to improve the performance of face recognition systems [15].

On Low-Resolution Faces. Then, we test our method on the problem of face recognition with low-resolution test images. In this experiment, 72×72 high-resolution face images and the corresponding 12×12 low-resolution ones are used as normal and degraded images, respectively. We compare our method with two restoration-based methods using "spline" interpolation and learning-based super-resolution [4].

As the previous experiment, we also adopt 150 Eigen-faces and 134 Fisher-faces in the algorithms involved in comparison. As for our method, we simply use pixel values as features for both the LR test images and the normal enrolled ones.

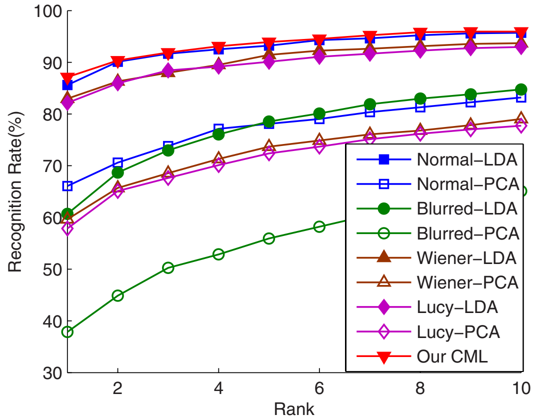


Fig. 6. Cumulative recognition results on blurred face recognition with different restoration and feature extraction methods

We compute the accumulative recognition rates for different methods that as shown in Figure 7. In the figure, the methods involved in our comparative study include Normal-LDA (HR test images), Normal-PCA, Spline-LDA (LR test images interpolatd by spline), Spline-PCA, LSR-LDA (LR test images super-resolved by learning-based method [4]), LSR-PCA, and Our CML. As we can see from Figure 7, our CML-based method outperforms most other methods and only a little worse than LDA with real high-resolution images. It achieves the best recognition rate of 86.3% for rank-1. The best rank-1 performance runner-up is 85.6%, given by high-resolution LDA. The performance of two other methods, learning-based method LDA and spline interpolation LDA, reach 82.0% and 74.4%, respectively.

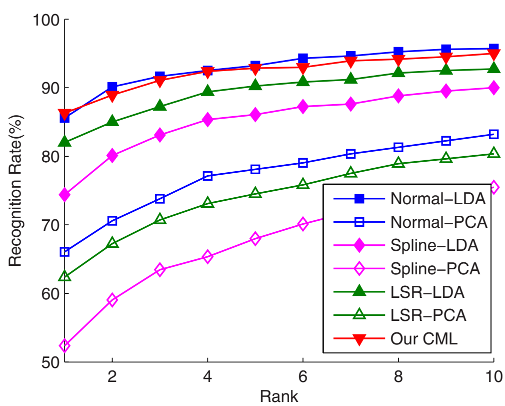


Fig. 7. Cumulative recognition results on LR face recognition with different restoration and feature extraction methods

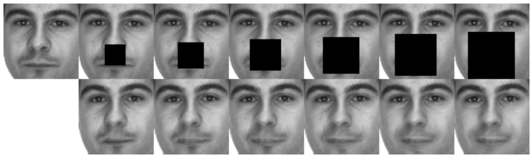


Fig. 8. Examples of synthetic partially occluded face images (the first row) and the corresponding restoration by HNE (the second row)

On Partially Occluded Faces. Finally, we present the experiments on partially occluded face recognition. Sometimes, the test face images are occluded by other object such as dark glasses. Similarly, our method can solve this problem through coupled metric learning.

We test our method on a synthetic set of partially occluded face images. We generate the occluded face images by setting the pixel intensity in a square area to zero. The sizes of the occlusion areas range from 20×20 to 45×45, with each expansion leading to 5-pixel larger in the area edges. Figure 8 shows some examples in the synthetic set and corresponding restored ones by holistic neighbor embedding (HNE) [29]. We compare our method with a two-step method which first restores the occluded images by HNE. The number of neighbors involved in the reconstruction is 7. As for our method, the partially occluded samples are represented by the first 250-dimensional features from PCA.

From the results show in Figure 9, the two-step method get very good performance, since the face images have strong priors and the un-occluded parts are accurate (see Figure 8). On the other hand, our method gets even better result than HNE-based methods and is comparable with the Normal-LDA method using the ground-truth images without occlusion.

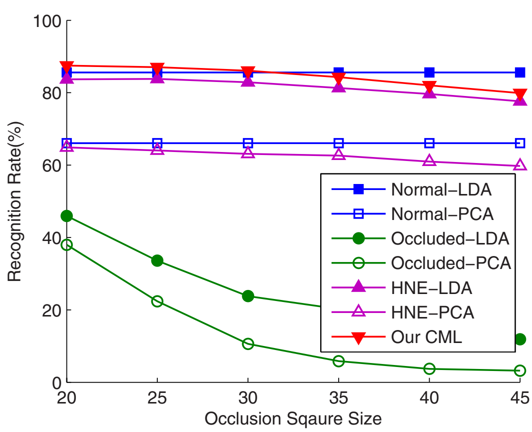


Fig. 9. Recognition rates on occluded face recognition with different restoration and feature extraction methods. The horizontal axis represents different sizes of occlusion squares.

5 Discussions

5.1 Real Dagraded Face Image

In Experiments, the synthesized dagraded face images are a little different from the real-world dagraded face image. For example, the low-resolution face images are noisy ones usually. Obviously, the performance of the original proposed method based on coupled metric learning will reduce. In this case, we can use the features which are nonsensitive for noise, such as Gabor wavelets, instead of the original intensity feature.

5.2 Nonlinear Extension with Kernels

The coupled metric learning method proposed in Section 2.2 can be extended from global linear forms to kernel version, since better feature representation and class separability is usually achieved through nonlinear mappings.

5.3 Other Possible Applications

It is worth noting that besides face recognition with degraded images, our proposed method can also be used to restore the degraded faces, as some super-resolution and deblurring methods do. As for super-resolution problems, the LR face to be reconstructed is mapped to a new space (learned by training HR and LR images), where we can do the reconstruction similar to the methods in [4] or [10]. Making good use of the relationship between LR and labeled HR images through coupled metric learning, our approach is more promising than some related super-resolution methods. Due to the space limitation, the experiments on face restoration with coupled metric learning are beyond the scope of this paper.

6 Concluding Remarks

This paper proposes a novel approach using coupled metric learning to address the problem of degraded face recognition. Different from some previous work, this method considers both the recognition of the degraded test faces and class-wise feature extraction of normal faces. The coupled metric learning is formulated as an optimization problem and can be solved efficiently with a closed-form solution. Experiments on various degraded face recognition tasks show that our method can get very satisfactory results.

In our future work, we will continue to study degraded face recognition problems along the metric learning direction. A possible extension of the current work is to study an new objective function which can put different weights on the coupled transformations. This is reasonable as one set of features may contain richer information than the other. However, it is not very straightforward if we still expect clean and closed-form solutions. Other possible research topics include investigating the relationship between reconstruction and recognition, as well as studying the break points and performance bounds for specific degraded face recognition.

Acknowledgement

This work is partially supported by National Natural Science Foundation of China under contract Nos.60803084, 60833013, and 60832004; 100 Talents Program of CAS, and National Basic Research Program of China (973 Program) under contract 2009CB320902.

References

1. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.: Face recognition: a literature survey. ACM Computing Surveys, 399–458 (2003)
2. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9), 1167–1183 (2002)
3. Biemond, J., Lagendijk, R., Mersereau, R.: Iterative methods for image deblurring. Proceedings of the IEEE 78(5), 856–883 (1990)
4. Chang, H., Yeung, D., Xiong, Y.: Super-resolution through neighbor embedding. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 275–282 (2004)
5. Fergus, R., Singh, B., Hertzmann, A., Roweis, S., Freeman, W.: Removing camera shake from a single photograph. ACM Transactions on Graphics 25(3), 784–794 (2006)
6. Freeman, B., Pasztor, E., Carmichael, O.: Learning low-level vision. International Journal of Computer Vision 40(1), 25–47 (2000)
7. Capel, D., Zisserman, A.: Super-resolution from multiple views using learnt image models. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 627–634 (2001)
8. Chen, J., Yuan, L., Tang, C., Quan, L.: Robust dual motion deblurring. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
9. Elad, M., Feuer, A.: Super-resolution reconstruction of image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(9), 817–834 (1999)
10. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw patches. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2008)
11. Baker, S., Kanade, T.: Hallucinating faces. Automatic Face and Gesture Recognition (2000)
12. Liu, C., Shum, H., Zhang, C.: A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 192–198 (2001)
13. Gunturk, B., Batur, A., Altunbasak, Y., Hayes, M., Mersereau, R.: Eigenface-domain super-resolution for face recognition. IEEE Transactions on Image Processing 12(5), 597–606 (2003)
14. Jia, K., Gong, S.: Multi-modal tensor face for simultaneous super-resolution. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, pp. 1683–1690 (2005)
15. Hennings-Yeomans, P., Baker, S., Kumar, B.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
16. Domeniconi, C., Peng, J., Gunopulos, D.: Locally adaptive metric nearest-neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9), 1281–1285 (2002)

17. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globally-consistent local distance functions for shape based image retrieval and classification. In: Proceedings of the Eleventh IEEE International Conference on Computer Vision (2007)
18. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood component analysis. In: Advances in Neural Information Processing Systems 17, pp. 513–520 (2005)
19. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. In: Weiss, Y., Schölkopf, B., Platt, J. (eds.) Advances in Neural Information Processing Systems 18. MIT Press, Cambridge (2006)
20. Hotelling, H.: Relations between two sets of variates. Biometrika 28, 312–377 (1936)
21. Hardoon, D., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. Neural Computation 16, 2639–2664 (2004)
22. Martinez, A., Benavente, R.: The AR face database. Technical report, CVC Tech. Report No. 24 (1998)
23. Zhang, Z.: Learning metrics via discriminant kernels and multidimensional scaling: toward expected euclidean representation. In: Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC, USA (2003)
24. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Englewood Cliffs (2002)
25. Lucy, L.B.: An iterative technique for the rectification of observed distributions. Astronomical Journal 79(6), 745–754 (1974)
26. Richardson, W.: Bayesian-based iterative method of image restoration. Journal of the Optical Society of America 62(1), 55–59 (1972)
27. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
28. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
29. Park, J., Oh, Y., Ahn, S., Lee, S.: Glasses removal from facial image using recursive error compensation. IEEE transactions on pattern analysis and machine intelligence 27(5), 805–811 (2005)