Occlusion Robust Face Recognition with Dynamic Similarity Features

Qingshan Liu, Wang Yan, Hanqing Lu, Songde Ma National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing, P. R. China, 100080 {qsliu, wyan, luhq, masd}@nlpr.ia.ac.cn

Abstract

In this paper, we present a new scheme for face recognition. The main idea is to represent the images with the similarity features against the reference set and to provide the relative match for two images. For any image, we first compute the similarities between it and all the reference images, and then we take these similarities as its feature. Based on the similarity features, a linear discriminating classifier is constructed to recognize the querying image. Inspired by research in cognitive psychology, the perceptual distance based dynamic similarity function is proposed to compute the similarity features. The proposed method can be regarded as a generalization of kernel discriminant analysis, and it can well deal with the nonlinear variations, especially occlusion. Extensive experiments are conducted to show its performance and robustness to occlusion.

1. Introduction

Face recognition is a hot topic in the field of computer vision and pattern recognition, due to its potential applications in biometrics, surveillance, human computer interface, etc. Many face recognition methods have been proposed in the past two decades [1]. Among them, Eigenface and Fisherface are two popular methods [2, 3]. Both of them first take the gray values of pixels as features, and then use Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to map the features into a low dimensional subspace for classification respectively. The idea of PCA is to generate a set of orthonormal projections aiming at maximizing the covariance over all the samples. LDA seeks a linear transformation which maximizes the between class scatter and minimizes the within class scatter. Thus, PCA is optimal for reconstruction, while LDA aims at better discrimination.

Both PCA and LDA are linear methods, so they are inadequate to represent nonlinear variations in real images. Since Support Vector Machine (SVM) got a great success in the field of pattern recognition [11], the kernel methods have attracted more attention due to its good nonlinear properties. The kernel methods first map the input data into an implicit feature space with a nonlinear mapping, and the data is analyzed in the feature space to get a nonlinear representation. Kernel PCA (KPCA) [4] and Kernel Discriminant Analysis (KDA) [6] are two nonlinear subspace methods and are successfully applied for face recognition [5, 7, 8], in which PCA and LDA are performed in the implicit feature space to produce nonlinear principal component subspace and discriminating subspace respectively. In implementation, in order to avoid the computation of implicit features, a kernel function is defined to calculate the dot product between two implicit feature vectors. The kernel must satisfy the Mercer conditions [11].

In this paper, we present a new method for face recognition, in which the images are represented by the similarity features against the reference set and the relative match strategy is provided for two images. For any image, we first compute the similarities between it and all the reference images, and then we regard these similarities as its feature. LDA based classifier is constructed on the similarity features, for it aims to separate one class from the others. Inspired by research in cognitive psychology, the perceptual distance based dynamic similarity function is proposed to compute the similarity features. The proposed method can be regarded as a generalization of KDA, and KDA is actually a special case, i.e., the similarity function satisfies the Mercer conditions and the reference set is same as the labeled training set. We test the proposed method and compare it with KDA on the FERET face database, which is widely used to evaluate the face recognition algorithms. The experimental results show its promising performance.

2. Our work

In this section, we first give the definition of dynamic similarity feature, and then construct LDA based classifier. Finally we discuss the link between the proposed method and KDA.



2.1. Dynamic similarity features

Given the reference images $R = \{r_1, r_2, ..., r_n\}$ and a similarity function *s*. Each image is represented by a M dimension vector. We define the feature vector of an image *x* as:

$$F_{x} = (s(x, r_{1}), s(x, r_{2}), ..., s(x, r_{n}))^{T}, \qquad (1)$$

where $s(z, x_i)$ represents the similarity between x and r_i . We call it the similarity feature.

The Minkowski distance is widely used for measuring similarity between two images,

$$d(x, y) = \left(\sum_{i=1}^{M} |x(i) - y(i)|^{p}\right)^{\frac{1}{p}},$$
(2)

where *p* is the Minkowski factor, and x(i) means the *i*-th element of *x*. When *p* is equal to 2, it is the well-known Euclidean distance (L_2 distance). Manhattan distance (L_1 distance) is the case of p = 1. We can see that the Minkowski distance has an assumption, i.e., similar images should be similar in all dimensions. But the psychology studies show that this assumption runs counter to human vision perception [12].

In this paper, we proposed a perceptual distance based similarity function. Actually the perceptual distance is a modified version of Minkowski distance, which is demonstrated to be consistent to the cognitive psychology and is also given a good evaluation in image retrieval [9]. The perceptual distance is defined as follows.

Let $d_i = |x(i) - y(i)|$, and assume d_i to be ordered as $d_1 \le d_2 \le ... \le d_M$. The definition of perceptual distance is:

$$d(x, y) = d_i(m, M) = \left(\sum_{d_i \in \Delta_m} d_i^{p}\right)^{1/p},$$
 (3)

where Δ_m represents the set of $\{d_1, d_2, ..., d_m\}$, m < M. Different from the Minikoski distance, the perceptual distance only considers the first *m* most similar respects and dynamically adjusts similar respects for matching two images. This dynamic property we call it dynamic distance. With the perceptual distance, we define the similarity function in this paper as:

$$s(x, y) = e^{-\lambda d(x, y)}, \qquad (4)$$

where λ is a positive constant and d(x, y) is the perceptual distance between x and y. We set p = 2 in the paper that is similar to L_2 distance. As for λ and m, we set them according to the experiments. Because of the dynamic property of the perceptual distance, the proposed similarity function is robust to occlusion to some extent, which will be investigated in the experiments.

2.2. Linear discriminating classifier

Feature extraction and classifier construction are two basic problems for all the pattern recognition applications. According to the above description, we represent an image using the similarity feature against the reference set. Based on the extracted features, we employ LDA based subspace analysis to construct the classifier. There are two advantages: 1). LDA aims to maximize the between class scatter and minimize the within class scatter, so it is good for separate one class from the others. 2). LDA is a subspace method, so it can well handle the problem that the dimension of the similarity feature increases with the number of the reference samples.

Let T be the labeled training set $X = (x_1, x_2, ..., x_N)$ for learning classifier. The feature vector of i-th training sample is $F_i = (s(x_i, r_1), s(x_i, r_2), ..., s(x_i, r_n))^T$. Based on all the F_i , i = 1, 2, ..., N, we first compute the between class scatter S_B and the within class scatter S_W . Then mathematically LDA seeks a linear projection W by maximizing Fisher function [3]:

$$J(W) = \arg\max_{W} \frac{W^{T} S_{B} W}{W^{T} S_{w} W}.$$
 (5)

Its solution is equivalent to finding the first leading eigenvectors of the matrix $S_W^{-1}S_B$. There are often no enough training samples to make S_W nonsingular in practical applications. In this paper, we adopt the strategy of simultaneous diagonalization to deal with this numerical computation problem, for this strategy is given a good evaluation in [10].

For a new querying image x, we first extract its similarity feature F_x as (1), and then project it into the LDA based subspace for classification, $\beta = W^T F_x$.

2.3. Discussion

In this subsection, we discuss the link between the proposed method and KDA. In order to better understand their relationship, we first give a brief review of KDA.

KDA first maps the labeled training data $X = (x_1, x_2, ..., x_N)$ into an implicit feature space Φ by a nonlinear mapping, $x \rightarrow \phi(x)$, and then performs LDA in Φ to find nonlinear discriminating subspace. In implementation, a kernel $k(x, y) = \phi(x) \cdot \phi(y)$ that satisfies the Mercer condition is introduced to avoid the computation of implicit features. The idea of performing LDA in Φ is to make the between class scatter S_B^{ϕ} maximum and the within class scatter minimum according to the Fisher function. Though S_B^{ϕ} and S_W^{ϕ} are unknown, they need not to be computed explicitly due to the effect of the dot product kernel. Define the Gram matrix $K = [K_1, K_2, ..., K_N]$, $K_i = (k(x_j, x_1), k(x_j, x_2), ..., k(x_j, x_n))^T$. KDA is mathematically equivalent to maximizing the following function [6].



 Table 1. The recognition rate (%) of different m

m / M	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
Recognition rate	93.56	93.97	95.15	96.23	96.99	97.41	97.15	96.99	95.98	92.22

$$J(\alpha) = \arg \max_{\alpha} \frac{|\alpha^{T} G_{B} \alpha|}{|\alpha^{T} G_{W} \alpha|}, \qquad (6)$$

where G_B and G_W are actually the between class scatter and within class scatter based on all the K_i , i = 1, 2, ..., N respectively. For a given new sample x, its projection in KDA subspace is $\beta = \alpha^T K_x$, $K_x = (k(x, x_1), k(x, x_2), ..., k(x, x_n))^T$.

Now we analyze the definitions of K_x and the similarity feature F_x . We can see that their expression is similar. However, $k(x, x_i)$ in K_x needs to satisfy the Mercer condition, while $s(x,r_i)$ in F_x has no such constrain. The proposed similarity function (4) seems not to satisfy the Mercer condition. If $s(x,r_i)$ satisfies the Mercer condition and the reference set R and the labeled training set X are same, the proposed method is actually equivalent to KDA. Thus, KDA can be regarded as a special case of the proposed method, and the proposed method is a generation of KDA and gives a reasonable interpretation for some proposed non-Mercer kernels [13, 14]. The proposed method also has nonlinear property if the similarity function is a nonlinear function. In addition, it gives a concept of relative measure for image matching against the reference set.

3. Experiments

We test the proposed method in the FERET database, which is widely used to evaluate the face recognition algorithms. The experimental data include 1002 front view face images selected from training CD, and FA and FB sets for testing. The FA has 1196 subjects and the FB set has 1195 subjects, and there is only one image per subject in both the FA and FB sets respectively. In order to justly compare the proposed method with KDA and LDA, we take 1002 images from the training CD as the reference set and the labeled training set for the proposed method. The FA images are used for gallery images, and the FB images are taken as probe images. The images are normalized to 48×54 by eye locations. Histogram equalization is performed as preprocessing.

3.1. Parameter setting

As for KDA, there are two popular kernels, i.e., polynomial kernel, $k(x, y) = (a(x \cdot y) + b)^d$, and Gaussian kernel, $k(x, y) = \exp(-\gamma \left\|\frac{x-y}{\sigma}\right\|^2)$. We compare KDA with both of them, and for simplicity,

we note them as KDA_P and KDA_G respectively. We set a = 0.001, b = 0 and d = 2 as in [7] for the polynomial kernel. We test the KDA_G with $\gamma = [0.1, 0.2, ..., 1]$ and the performance of $\gamma = 0.3$ is better, we report the results of $\gamma = 0.3$ in the following.

The proposed similarity function has three parameters, i.e., p, λ , and m. Here, we set p = 2 similar to the well-known L_2 distance. We set $\lambda = 0.1$ for it gives a better performance in $\lambda = [0.1, 0.2, ..., 1]$. About the dynamic parameter m, we investigate m/M from 0.55 to 1. The results are reported in Table 1. We can see that m/M = 0.8 gives a better recognition rate of 97.41 %. It shows that 20% of M points should belong to dissimilar points and is not good for recognition. In the following experiments, we set m/M = 0.8.

3.2. Comparing with KDA and LDA

We compare the proposed method with KDA_P, KDA_G and LDA. We can see that the proposed method outperforms the other three methods. Both the best recognition rates of KDA_G and KDA_P are not up to 96.7%, (KDA_G: 96.65 %, KDA_P: 95.9%), while the proposed method gets the recognition rate of 97.41 %. It is implied that the perceptual distance based dynamic similarity feature is more consistent to human vision perceptual. LDA gives the worst performance, for it is a linear method.



Figure 1. The comparison of four methods

3.3. Investigate the robustness to occlusion

The proposed method has another advantage, i.e., it is robust to occlusion to some extent. In order to



evaluate this property, we randomly crop a patch for the probe image. In our experiments, we crop the patches with 5×5 , 7×7 11×11 and 17×17 . Figure 2 shows some examples. Figure 3 reports the experimental results, where 0 means no occlusion. We can see that our method is more robust to occlusion than KDA, for the proposed method dynamically focuses on the first similar *m* pixels, while KDA_G and KDA_P consider all the pixels. Especially the polynomial kernel is very sensitive to the occlusion.



Figure 3. Performance with occlusions

4. Conclusion

In this paper, we present a new occlusion robust face recognition method. Based on the reference set, any image can get a set of similarities between it and all the reference images, and we take these similarities as the feature vector of the image. The LDA based classifier is constructed by the similarity features. Inspired by research in cognitive psychology, the perceptual distance based similarity function is proposed to compute the similarity features. The proposed method actually provides a relative measure concept for recognition, and it also has nonlinear property similar to kernel discriminate analysis (KDA) that is embodied in similarity function. Moreover, the proposed method can be regarded as a generalization of KDA. Extensive experiments are conducted to show its performance and robustness to occlusion.

Acknowledgement

This work is partially supported by the National Key Basic Research and Development Program (973) under Grant No. 2004CB318107, and the Natural Sciences Foundation of China under Grant No. 60405005, 60121302 and NSFC—RGC 60318003.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, Face Recognition: A Literature Survey, *CS-Tech Report-4167*, University of Maryland, 2000.

[2] M. Turk and A. Pentland, Eigenfaces for recognition, *J. cognitive Neuroscience*, 1991, 3(1): 72-86.

[3] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. PAMI*, 1997, 19(7): 711-720.

[4] B. Scholkopf, A. Smola, K. R. Muller, Nonlinear component analysis as a kernel Eigenvalue problem, *Neural Computation*, 1998, 10(5): 1299-1319.

[5] M. H. Yang, N. Ahuja, D. Kriegman, Face recognition using kernel Eigenfaces, *Proc. Int. Conf. Image Processing*, 2000.

[6] Smoke, G. Ratsch, J. Weston, Fisher Discriminant Analysis with Kernels, *Proc. Neural Networks for Signal Processing Workshop*, 1999.

[7] Q. S. Liu, R. Huang, H. Q. Lu, and S. D. Ma, Face Recognition Using Kernel Based Fisher Discriminant Analysis, *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[8] M. H. Yang, Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods, *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[9] B. Li, E. Y. Chang, Discovery of A Perceptual Distance Function for Measuring Image Similarity, *ACM Multimedia Journal Special Issue on Content-Based Image Retrieval*, 2003.

[10] C. Liu and H. Wechsler. Enhanced Fisher Linear Discriminant Models for Face Recognition, *Proc. Int. Conf. Pattern Recognition*, 1998.

[11] E. Osuna, R. Freund, and F. Girosi, Support Vector Machines: Training and Applications, *Tech Report*, AI Lab, MIT, 1997.

[12] D. L. Medin, R. L. Goldstone and D. Gentner, Respects for similarity, *Psychological Review*, 100(2): 254-278, 1993.

[13] Nuno Vasconcelos, Purdy Ho, and Pedro Moreno, The Kullback-Leibler Kernel as a Framework for Discriminant and Localized Representations for Visual Recognition, *Proc. European Conf. Computer Vision*, 2004.

[14] S. Lyu, Mercer Kernels for Object Recognition with Local Features, *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2005.

