A SVM-Based Feature Extraction for Face Recognition

Peng Cui(✉) and Tian-tian Yan

Harbin University of Science and Technology, Harbin 150080, China
cuipeng83@163.com

Abstract. Social computing, a cross science of computational science and social science, is affecting people's learning, work and life recently. Face recognition is going deep into every field of social life, and the feature extraction is particularly important. Linear Discriminant Analysis (LDA) is an effective feature extraction method. However, the traditional LDA cannot solve the nonlinear problem and small sample problem existing in high dimensional space. In this paper, the method of the Support Vector-based Direct Discriminant Analysis (SVDDA) is proposed. It incorporates SVM algorithm into LDA, extends SVM to nonlinear eigenspace, and optimizes eigenvalue to improve performance. Moreover, this paper combines SVDDA with the social computing theory. The experiments were tested on different face datasets. Compared with other existing methods, SVDDA has higher robustness and optimal performance.

Keywords: Discriminant analysis · Face recognition · Support vector machine · Feature extraction

1 Introduction

In recent years, on the social computing research has become an increasingly active area, related research and practice has been growing rapidly, and its appearance is closely related to the use of the Internet, Web2.0 technology and social software. It has a profound impact on all aspects of people's lives [1]. Social computing is a cross discipline combining computing technology with social science. It uses computational techniques to help us understand the social rules, mutual communication and cooperation, and uses the principle and method of swarm intelligence to solve the problem. However, there is no clear and universally accepted definition. On social computing environment, vast social softwares have been produced and used widely. This phenomenon promotes social computing to penetrate into every field of society, and achieves the digital dynamic blend in the field of science, technology, humanities in the fields of science, technology, humanities, etc.

The face recognition products have been widely used in the field of financial, legal, military, public security, frontier defense, government, aerospace, power, factories, education, health care, and many enterprises and institutions, etc. With the further development of technology and the improvement of social recognition, face

W. Che et al. (Eds.): ICYCSEE 2016, Part I, CCIS 623, pp. 120–126, 2016.
DOI: 10.1007/978-981-10-2053-7_12

recognition technology will be applied in more fields. It is closely related to people's life. Moreover, feature extraction is a key step in face recognition.

In the past decades, subspace analysis methods are the most efficient method for feature extraction on face recognition. Among these methods, LDA and the improved algorithm based LDA are popular, and obtain good recognition performance [2, 3]. Yu and Yang proposed Direct Linear Discriminant Analysis (D-LDA) to find most important eigenvector projection in the eigenspace (nonzero eigenvalues subspace) between-class scatter or (zero eigenvalues subspace) within-class scatter matrix [4]. For improving performance, LDA is extended to nonlinear subspace by kernel trick, including Kernel Fisher Discriminant (KFDA) [5], Generative Discriminant Analysis (GDA) [6], Kernel Direct Discriminant Analysis (KDDA) [7] and Kernel Discriminant Analysis (KDA) [8]. Kim proposed Support Vector Machine Discriminant Analysis SVM-DA [9], which incorporated SVM into LDA to solve dimension disaster and small sample problem. However, the algorithm is same as traditional LDA, which ignored eigenvector set of within-class scatter matrix zero eigenvalues space. Inspired by SVM-DA, KDDA and DLDA, we propose a Support Vector–based Direct Discriminant Analysis (SVDDA), which extends SVM to nonlinear eigenspace, and optimizes eigenvalue to improve performance.

2 Support Vector-Based Direct Discriminant Analysis

The null space of within class scatter (S_W) is given up firstly in LDA or SVM-DA, which can include discriminant information. Secondly, the null space of between-class scatter S_B is given up. In the feature selection process, SVM-DA gives up some small eigenvalues. Though it is effective to eliminate noise, the null space of S_W is given up. We believe that the null space which can contain some important class information that should be retained. Firstly, diagonalize S_B, and solve eigenvalues; in terms of eigenvalues, give up useless null space. Then diagonalize S_W, but retain those small eigenvalues which can contain important information. If S_W is singular, let S_W equal to $S_W + S_B$ for solving small samples problem.We extend LDA to SVM nonlinear space, and propose SVDDA.

Let $\tilde{\Omega}_{ij}$ represents parameter of SVM hyperplane, which can separate, class i and class j into generative space H. The expression is

$$\tilde{\Omega}_{ij} = \sum_{\forall x_k \in S_{ij}} \alpha_{ij}^k y_{ij}^k \phi(x_k) = \Phi \alpha_{ij} \tag{1}$$

Where S_{ij} is set of support vectors, α_{ij}^k and y_{ij}^k are respectively weight and label corresponding to support vector x_k, $\Phi = [\varphi(x_1)\ldots\varphi(x_N)]$ is training set which is mapped to kernel feature space. $\alpha_{ij}^k y_{ij}^k$ denotes corresponding to support vectors in Φ. SVM-based between-class scatter matrix [9] in kernel feature space can be represented as

$$\tilde{S}_B = \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} p_i p_j \frac{\tilde{\Omega}_{ij}\tilde{\Omega}_{ij}^T}{\|\tilde{\Omega}_{ij}\|^2} = \frac{1}{N^2} \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} N_i N_j \frac{\tilde{\Omega}_{ij}\tilde{\Omega}_{ij}^T}{\|\tilde{\Omega}_{ij}\|^2} \tag{2}$$

In terms of kernel function, Eq. (2) can be converted into form of dot product:

$$\tilde{S}_B = \tilde{\Omega} A \tilde{\Omega}^T = \Phi \alpha A \alpha^T \Phi^T \tag{3}$$

Where $\tilde{\Omega} = [\Omega_{1,2}, \ldots, \Omega_{1,c}, \ldots, \Omega_{(c-1),c}]$, $A = diag[(N_1 \times N_2/N^2), \ldots, (N_{c\text{-}1} \times N_c/N^2)]$, and $\alpha = [\alpha_{1,2}, \ldots, \alpha_{(c-1),c}]$. The within-class scatter matrix is

$$S_W = \sum_{i=1}^{c} \sum_{\forall x_i \in C_i} (\phi(x_j) - \overline{\phi_i})(\phi(x_j) - \overline{\phi_i})^T = \Phi(I - B - B^T + BB^T)\Phi^T \tag{4}$$

Where $\overline{\phi_i}$ denotes mean vector of samples in class i, $\overline{\phi_i} = \frac{1}{N_i} \sum_{\forall x_i \in C_i}^{c} \phi(x_j)$, B is $N \times N$ block diagonal matrix given by $(B_i)_{i=1,\ldots,c}$, B_i is $N_i \times N$ matrix of all terms which equal to $1/N_i$, $B_i = 1/N_i$, i = 1,2,…,c. In terms of Eqs. (3) and (4), the new Fisher criterion is

$$V = \arg\max_{v} \frac{\left|V^T \Phi \alpha A \alpha^T \Phi^T V\right|}{\left|V^T \Phi (I - 2B + BB^T) \Phi^T V\right|} = \arg\max_{v} \frac{\left|V^T \tilde{S}_B V\right|}{\left|V^T S_W V\right|} \tag{5}$$

As basis vector V corresponding to nonzero eigenvalues should locate in generative space of training vector, it exists efficient E satisfy

$$V = \sum_{i=1}^{N} e_i \phi(x_k) = \Phi E \tag{6}$$

Where $E = [e_1, \ldots, e_n]^T$ N-dimensional vector. Equation (6) is substituted into Eq. (7),

$$E = mag \max_{E} \frac{E^T \Phi^T \Phi \alpha A \alpha^T \Phi^T \Phi E}{E^T \Phi^T \Phi (I - 2B + BB^T) \Phi^T \Phi E} \tag{7}$$

As $\Phi^T \Phi$ only need to compute dot product in generative space H, it can be replaced with a kernel function. The function should be same as the kernel used in SVM hyperplane. Let K be $N \times N$ matrix,which be made up of dot product of training samples in H.

Let K replace $\Phi^T \Phi$,then Eq. (7) is represented as

$$E = m \arg\max_{E} \frac{E^T K \alpha A \alpha^T K E}{E^T K (I - 2B + BB^T) K E} = \frac{E^T \tilde{S}_B^K E}{E^T S_W^K E} \tag{8}$$

$\tilde{S}_B^K$ and S_W^K are new defined scatter, and E which can be solved by eigenvalues problem.

3 Feature Extraction of SVDDA

As the dimension of the feature space F is expressed as possible arbitrarily large, or may be infinitely, it is difficult to directly calculate the matrix $\tilde{S}_B^K$ and S_W^K. Eigenvalue problem of $\tilde{S}_B^K$ can be obtained indirectly by $\tilde{S}_B$.

(1) Feature analysis of $\tilde{S}_B$

First diagonalize $\tilde{S}_B$ and find $\tilde{\Lambda}_B$, $\tilde{\Lambda}_B$ is a diagonal matrix in descending order. If $\tilde{S}_B$ is singular, it is necessary to give up those eigenvalues and eigenvectors. $V_m = [v_1, \ldots, v_m] = \Phi E_m$, where $E_m = [e_1, \ldots, e_m]$ is corresponding eigenvector with eigenvalues greater than 0. $\tilde{\tilde{\Lambda}}_B \tilde{\tilde{\Lambda}}_B$ is main sub-array of $\tilde{\Lambda}_B$,which is $m \times m$ diagonal matrix with eigenvalues greater than 0.

(2) Feature analysis of S_W^K

Let $Z = V_m \tilde{\tilde{\Lambda}}_B^{-1/2}$, then $Z^T \tilde{S}_B Z = I$. We project S_W to S_W^K of generative space by Z

$$Z^T S_W Z = (E_m \Lambda_b^{-1/2})^T (\Phi^T S_W \Phi)(E_m \Lambda_b^{-1/2}) \tag{9}$$

Where $\Phi_b^T S_W \Phi_b$ equals to S_W^K. By diagonalization of $Z^T S_W Z$, we obtain Λ_W

$$G Z^T S_W Z G = \Lambda_W \tag{10}$$

Where G is eigenvector of $Z^T S_W Z$. The eigenvectors with maximum eigenvalues are given up, a M selected eigenvectors are represented as $G_W = [g_1, \ldots, g_M]$. Let $P = Z G_M$, then $P^T S_W P = \tilde{\Lambda}_W$, where $\tilde{\Lambda}_W = diag[\lambda'_1, \ldots, \lambda'_M]$ is $M \times M$ main sub-matrix of Λ_W,

$$(E_m \tilde{\tilde{\Lambda}}_B^{-1/2} G_M)^T (\Phi^T S_W \Phi)(E_m \tilde{\tilde{\Lambda}}_B^{-1/2} G_M) = \tilde{\Lambda}_W \tag{11}$$

we obtain

$$\begin{aligned} \tilde{\Lambda}_W^{-1/2} P^T S_W P \tilde{\Lambda}_W^{-1/2} = I_W &= (E_m \tilde{\tilde{\Lambda}}_B^{-1/2} G_M \tilde{\Lambda}_W^{-1/2})^T (\Phi^T S_W \Phi)(E_m \tilde{\tilde{\Lambda}}_B^{-1/2} G_M \tilde{\Lambda}_W^{-1/2}) \\ &= E_{opt}^T S_W^K E_{opt} \end{aligned} \tag{12}$$

Where E_{opt} is optimal eigenvector in generative space, and $E_{opt} = E_m \tilde{\tilde{\Lambda}}_B^{-1/2} G_M \tilde{\Lambda}_W^{-1/2}$. To make the method robust, modified criterion is used. When $Z^T S_W Z$ is singular, then

$$V = mag \max_V \frac{|(V S_B V)|}{|(V^T \tilde{S}_B V) + (V S_W V)|} \tag{13}$$

The modified criterion is equivalent to standard Fisher criterion. Hence $Z^T S_W Z$ in Eq. (10) is replaced with $Z^T (\tilde{S}_B + S_W) Z$.

(3) Dimension reduction and feature extraction.

Any input pattern z, eigenvector projection can be represented as

$$y = E_{opt}^T \Phi^T \phi(z) = (E_m \tilde{\tilde{\Lambda}}^{-1/2} G_M \tilde{\Lambda}_W^{-1/2})(\Phi^T \phi(z)) = E_{opt}^T . \gamma(z) \quad (14)$$

where $\gamma(z) = [k(x_1, z), \ldots, k(x_N, z)]^T$ is $N \times 1$ kernel vector.

4 Experiment and Result Analysis

4.1 Dataset and Experimental Method

To assess the proposed method, experiments are conducted on FERET, AR, PIE datasets. FERET dataset consists of 1400 images from 200 subjects, and image resolution is 80 × 80. AR dataset consists of 1200 images from 120 subjects, and image resolution is 40 × 50. PIE dataset consists of 1360 images from 68 subjects, and image resolution is 32 × 32. Figure 1 shows sample images from three datasets. The images in first row are randomly selected 10 images from FERET dataset. The images in second row are randomly selected 10 images from AR dataset. The images in third row are are randomly selected 10 images from PIE dataset.

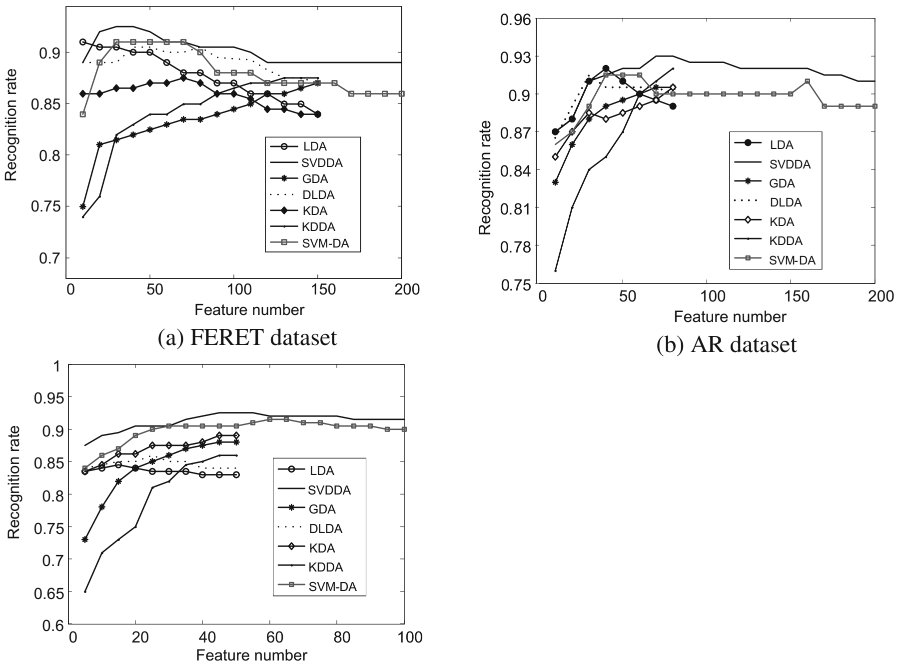
Fig. 1. Some sample images from different datasets

To evaluate the proposed method, we compare SVDDA with LDA, DLDA, KDA, GDA, KDDA and SVM-DA. 4, 5, 5 images per person from FERET, AR and PIE datasets are randomly selected for training and the rest for testing. After 7 runs, results are averaged.

4.2 Experimental Results

Figure 2 shows the recognition rate from 7 different methods on FERET dataset. It can be seen that recognition rate of SVDDA is obviously better than other methods. The recognition rates of LDA, DLDA and SVM-DA are close. However, the PCA reduction dimension is all used by LDA and DLDA, so their dimensions are limited. KDDA and GDA achieve low recognition rate. Recognition rate of KDA is better than KDDA and GDA.

As seen in the three experiments, SVDDA receives the highest average recognition rate; DLDA is relatively close to the SVM-DA, they are better than other nonlinear



(a) FERET dataset

(b) AR dataset

(c) PIE dataset

Fig. 2. Recognition rate of different methods on three datasets

method. Though KDDA receives lowest average recognition rate, the recognition rate is better in high dimensions. This shows that the non-linear methods don't obtain superior performance in all case. As the dimensionality of Fisher criterion-based subspace increases, there are more and more useless information, so discriminant information of low dimensions is particularly important. In comparison with other methods, computation time of SVDDA is slightly higher than others, but considering its high recognition rate, and its robustness that can be applied to all feature dimensions, thus its performance is optimal.

5 Conclusion

The emergence of social computing theory provides modern methods and means for effectively dealing with the complex and dynamic changing emerging social and the engineering problems. It promotes the formation and development of computational sociology. Based on the theory of social computing, we propose SVDDA. It is used to extracting nonlinear feature. This method uses SVM projection to replace the traditional mean shift, and modifies discriminant criterion to solve small sample problem existing in high dimensional space, and changes computation order of eigenvalues of scatter to retain important information by giving up null space of within-class scatter and retaining null space of within-class scatter. However there are still some problems, such as the choice of the kernel parameter and computation time, so how to choose effective kernel function and reduce the computational complexity still require further study.

Acknowledgement. This research was funded by science technology research of Heilongjiang provincial education department under Grant No. 11551086.

References

1. Wu, Y-.h., Liu, X., Ma, X.-l.: Overview of research on application and innovation of e-learning in education based on social computing. Open Educ. Res. **18**(4), 99–105 (2012)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 711–720 (1997)
3. Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D.L., Weng, J.: Discriminant analysis of principal components for face recognition. In: Wechsler, H., Phillips, P.J., Bruce, V., Soulié, F.F., Huang, T.S. (eds.) Face Recognition: From Theory to Applications. NATO ASI Series, pp. 73–85. Springer, Heidelberg (1998)
4. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data with application to face recognition. Pattern Recogn. **34**, 2067–2070 (2001)
5. Müller, K., Mika, S., Rtsch, G.: An introduction to kernel-based learning algorithms. IEEE Trans. Neural Networks **21**, 181–201 (2001)
6. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Comput. **12**, 2385–2404 (2000)
7. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Netw. **14**, 117–126 (2003)
8. Devi, H.S., Laishram, R., Thounaojam, D.M.: Face recognition using R-KDA with non-linear SVM for multi-view database. Procedia Comput. Sci. **54**, 532–541 (2015)
9. Kim, S.K., Park, Y.J., Toh, K.A.: SVM-based feature extraction for face recognition. Pattern Recogn. **43**, 2871–2881 (2010)
10. Li, F.-g., Liang, Y., Gao, X.-z.: Research on text categorization based on LDA-wSVM model. Appl. Res. Comput. **32**(1), 21–25 (2015)
11. Yang, L.-y., Qin, A.: Face recognition algorithms based on uncorrelated multilinear PCA. Radio Commun. Technol. **42**(1), 73–75 (2016)
12. Dong, Y., Xu, Y.-b., Zhuo, L., et al.: Research on spam identification based on social computing and machine learning. J. Shandong Univ. (Nat. Sci.) **48**(7), 72–78 (2013)
13. Meng, X.-f., Li, Y., Zhu, J.J.H.: Social computing in the era of big data: opportunities and challenges. J. Comput. Res. Develop. **50**(12), 2483–2491 (2013)