http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (6): 1246-1251, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1246.1251

A New Direct Linear Identify Algorithm for the Small Samples

Xu Xiao-Yuan and Zheng Yu-Jie
JiangSu Radio and TV University, YingTian Street No. 832, Nan Jing City, 210017, China

Abstract: In solving the small sample problems, Direct Linear Identity Algorithm (DLDA) is an effective method by null space information. But the traditional DLDA method needs to handle a high-dimensional scatter matrix at the first step, reducing the stability of the algorithm. In this study, a new Direct Linear Identity Algorithm (DLDA/QR) is proposed by the introduction of the QR decomposition algorithm, improving the stability of the algorithm. And it gives the experimental data and analysis.

Key words: Algorithm, DLDA/QR, decomposition, face recognition, small sample

INTRODUCTION

During the feature extraction process, in order to ensure that the scatter matrix within classes is nonsingular, the number of samples should be greater than its dimension in collecting samples. However, in the face recognition and other applications, due to the number of samples used for training is usually limited by the actual sample database while the dimension of the image data is very high, for example, in the ORL face database, the actual number of face images in the database (40×10 = 400), is much smaller than the original image dimension (112×92 = 10304), so this kind of problems as we know is called small sample problems. In this kind of problems, it is quite difficult to find enough training samples to ensure the scatter matrix within classes reversible and affected by the number of training samples, the scatter matrix within classes is often singular. Therefore, in the case of small samples, how to get the optimal discriminant feature extraction set is a recognized problem. Ways to solve the problem can be summarized into the following two categories:

One method is from the algorithm's point of view, directly aimed at the small sample problems, a solving algorithm is put forward at the corresponding matrix analysis. Another method is from the point of view of sample models, in order to eliminate the within class scatter matrix singularity, sample vector dimension can be reduced firstly. Under the guidance of this thought, we can use the following two ways: one is directly operating in the image space, by reducing the dimensions of the image (image resolution) to reduce the dimension (Bai *et al*., 2012; Cao and Chen, 2011) and eliminate the singularity; the other is firstly to preprocess the original image space and do feature extraction in this space. The common method is to reduce the original image dimension, using the principal component analysis (PCA, also known as the K-L transform). After the original image space is descending into an appropriate space, then Fisher criteria is used to distinguish feature extraction and the extraction method is also called two-step feature extraction method based on PCA+LDA, the most typical example of which is fisherfaces (Delgado-Gomez *et al*., 2009).

Currently, to solve the small sample size problem, two step method based on PCA+LDA has become a practical and feasible solution but there are still many defects in practical applications: For example, after descending dimension by PCA, how to determine the dimension of samples so as to avoid the information distortion and if assuming that, in the space of descending dimension, there is singularity in within class scatter matrix, how to effectively extract distinguishing feature from the corresponding zero space. In order to solve these problems, many scholars put forward their solutions striving to get effective discriminant feature information extraction. Direct linear discriminant analysis method (Gan and He, 2011; Lee *et al*., 2012; Yong *et al*., 2011) (Direct LDA, DLDA) using the null space information, effectively solved the small sample problem but in the processing steps that need to be addressed in a high dimensional scatter matrix, algorithm stability is thus reduced. For the defect of the algorithm, the QR decomposition algorithm (Ye and Li, 2005; Salmela *et al*., 2008) is introduced and a new DLDA/QR algorithm is presented which effectively improves the stability and achieves the experimental data and analysis.

Corresponding Author: Xu Xiao-Yuan, JiangSu Radio and TV University, YingTian Street No. 832, Nan Jing City, 210017, China

DIRECT LDA

In direct LDA, the basic idea of which is in the null space of the within class scatter matrix looking for a set of eigenvectors which can make within class non zero scatter distance maximum and should be orthogonal used as a projection axis. Here, we assume that n_b and N_w represent respectively the null space of S_b and S_w. Thus, $N'_b = R^n\text{-}N_b$ and $N'_w = R^n\text{-}N_w$ represent, respectively the supplement space of N_b and N_w. From the intersection of $N'_b \cap N_w$, the optimal discriminant vector set can be obtained. The algorithm process can be described as below:

Step 1: Calculate within class scatter matrix S_b and scatter matrix between classes S_w

Step 2: Present rank(S_w) = r, if r = n get a set of feature vectors of the maximum eigenvalue $(S_w+S_t)^{-1} S_b$, use this as the differential vector, otherwise, go to the next step

Step 3: On the singular value decomposition of S_w as $S_w = U\sum V^T$, in which, S_w is a real symmetric matrix, so U = V

Step 4: Assume $V = (v_1, ..., v_r, v_{r+1},..., v_n)$ and $Q = (v_{r+1},..., v_n)$, through the proof, the zero space of S_w can be obtained $N_w = Q = \text{span}(v_{r+1},..., v_n)$

Step 5: Calculate $\tilde{s}_b = QQ^TS_b(QQ^T)^T$

Step 6: Calculate a set of feature vectors according to the largest eigenvalue of $\tilde{s}_b$ and use them as the differential vectors

From the point of calculation amount, the Chen's method is not suitable for processing high dimensional sample vectors. Because the third and sixth step of the algorithm are in the same space of the original image vector dimension, if the dimension of the original image space is very high, occurring frequently in image recognition problems, this algorithm is becoming less effective for the large amount of calculation.

A NEW DIRECT LINEAR DISCRIMINANT ANALYSIS ALGORITHM (DLDA/QR)

Analyzing the defects of DLDA algorithm and introducing QR decomposition, we put forward a new direct linear discriminant analysis algorithm: DLDA/QR algorithm. The algorithm retains thinking of the DLDA algorithm, reduces the time complexity of the algorithm and improves the stability and effectiveness.

Firstly the optimization is realized as below:

$$G = \underset{G^TG=I}{\arg\max}\, \text{trace}\left(G^TS_bG\right) \tag{1}$$

From the Eq. 1, at the first step of this new algorithm, we can focus on solving the maximum of the within class scatter matrix. This process can be referenced by the principal component analysis method and also can be realized by QR decomposition method. In the new algorithm, the matrix is defined as follows:

$$H_b = [\sqrt{N}(\mu_1 - \mu)\cdots\sqrt{N}(\mu_C - \mu)] \tag{2}$$

And meet:

$$S_b = H_bH_b^T \tag{3}$$

Thus, we can get the QR decomposition:

$$H_b = (Q_1, Q_2)\begin{pmatrix} R \\ 0 \end{pmatrix} \tag{4}$$

Among which, $Q_1 \in X^{m\times t}$, $Q_2 \in X^{m\times(m\text{-}t)}$, $R \in X^{t\times L}$ and $t = \text{rank}(S_b)$.

Then, we come to the full rank decomposition. Therefore, for any arbitrary orthogonal matrix W ($W \in R^{t\times t}$), $G = Q_1W$ can solve the optimization problems of Eq. 1. Here special attention should be paid to the rank (t) of matrix H_b, the ceiling value of which is C-1. In the process of actual application, data C is usually linearly independent. In this case, after decomposition, the dimension of QR becomes t (t = C-1 = rank(S_b)).Then introducing QR decomposition, we have achieved vector dimensionality reduction similar to the PCA method.

At the first step of DLDA/QR algorithm, we are concerned with the maximum of the scatter matrix between classes. At the second step, were concerned with the within class scatter matrix. In this process the feature extraction can refer to the DLDA method but in order to improve the stability and effectiveness of the algorithm, the process of feature extraction is limited in a small space. Q_1 is obtained by calculating the matrix and we can calculate the corresponding scatter matrix $\tilde{s}_t = Q_1S_tQ_1$ and $\tilde{s}_b = Q_1S_tQ_1$. The two scatter matrix are in the new reduced dimensional space that we can also verify that matrix $\tilde{s}_t$ and matrix $\tilde{s}_b$ are t×t and in which $\tilde{s}_b$ is non singular.

In this process, we calculate the matrix which can simultaneously diagonalize $\tilde{s}_t$ and $\tilde{s}_b$:

$$V^T \tilde{s}_t V = \wedge,\ V^T \tilde{s}_b V = I \quad (5)$$

Among them $\wedge$ is a diagonal matrix, whose diagonal elements is sorted in ascending order and I is a unit matrix.

At the same time realizing the diagonalization of the process, the first step is to carry out the diagonalization of symmetric matrix $\tilde{s}_b$. Because it is a t×t matrix and typically t<<n, therefore, we can easily incorporate this matrix diagonalization.

We assume the existence of a matrix U which satisfies:

$$U^T \tilde{s}_b U = \wedge_b \quad (6)$$

Among them $U^T U = I$ and $\wedge_b$ is a diagonal matrix whose diagonal elements are sorted in descending order.

Now that:

$$Z = U\wedge_b^{-1/2} \quad (7)$$

So, it can be:

$$\left(U\wedge_b^{-1/2}\right)^T \tilde{s}_b U\wedge_b^{-1/2} = I \Rightarrow Z^T \tilde{s}_b Z = I \quad (8)$$

So the diagonalization is achieved in a new between-class scatter matrix. Assuming that:

$$\tilde{s}_t = Z^T \tilde{s}_t Z \quad (9)$$

And using the similar method, we can realize the total scatter matrix diagonalization.

Similarly, assuming the existence of a matrix satisfies:

$$Y^T \tilde{s}_t Y = \wedge_t \quad (10)$$

Among them, $\wedge_t$ is a diagonal matrix and the diagonal elements are sorted in ascending order (Eq. 10).

Usually, a set of vectors are selected to constitute a matrix ($P = (y_1,..., y_i,..., y_s)$). The number of the vectors is s ($s \le t$) and y_i represents the column i of matrix Y. Assuming:

$$P^T \tilde{s}_t P = \wedge_s \quad (11)$$

$$V_s = ZP \quad (12)$$

Thus, we can get:

$$V_s^T \tilde{s}_t V = P^T Z^T \tilde{s}_t ZP = P^t \tilde{s}_t P = \wedge_s \quad (13)$$

and

$$V_s^T \tilde{s}_b = P^T Z^T \tilde{s}_b ZP = P^T IP = I \quad (14)$$

In conclusion, we may safely draw that $V_s = ZP$ can simultaneously diagonalize the matrix $\tilde{s}_b$ and $\tilde{s}_t$.

Therefore, we get the final transformation matrix as follows:

$$E = Q_1 ZP\wedge_s^{-1/2} \quad (15)$$

Then, for any test sample, in a low dimension space the projection data is get by:

$$\Omega_{test} = E^T x_{test} \quad (16)$$

In addition, the simultaneous diagonalization process can be described by a theorem to explain.

Theorem 1: We can simultaneously diagonalize two symmetric matrices $\tilde{s}_t$ and $\tilde{s}_b$, $V^T \tilde{s}_t V = \wedge$, $V^T \tilde{s}_b V = I$, among which V and $\wedge$ are eigenvectors and eigenvalues of $\tilde{s}_b^{-1} \tilde{s}_t$ and satisfy $\tilde{s}_t^{-1} \tilde{s}_t V = \wedge V$.

Assuming that the diagonal elements of the diagonal matrix $\wedge$ are sorted in ascending order, accordingly, the transformation matrix $E = Q_1 V_s \wedge_s^{-1/2}$, among which V_s is a matrix composed of the anterior column of the matrix V and $V_s = t \times s$. $\wedge_s$ is a matrix composed of the matrix V and $\wedge_s = s \times s$.

Based on the above discussion, we derive the steps of DLDA/QR algorithm as follows:

Step 1: In the original sample space H_b, S_b and S_t are respectively calculated. QR decomposition is used to obtain Q_1 and then we get two new scatter matrices:

$$\tilde{s}_t = Q_1 S_t Q_1,\ \tilde{s}_b = Q_1 S_b Q_1$$

Step 2: $(\tilde{s}_b)^{-1} \tilde{s}_t$ is obtained by calculating the eigenvalue and eigenvector of the matrix respectively. $\wedge$ represents the eigenvalue of the matrix and V means the eigenvector. Calculating $\wedge_s$ and V_s, we get the final transforming matrix $E = Q_1 V_s \wedge_s^{-1/2}$

Step 3: According to Eq. 16, testing samples are projected to a lower dimensional space and then classified

THE EXPERIMENTAL DATA AND RESULTS ANALYSIS

This study uses the ORL face database (Lee *et al.*, 2012; Yong *et al.*, 2011) as experimental data base. ORL face database is composed of 40 human facial images, the original dimension of which is 92×112. The facial expression and facial details are different. Some of the database are shown in Fig. 1.

The experiments are based on the face database of ORL. In the process of experiments, the training samples and the testing samples both are randomly selected. We selected 9 face samples of each class from the ORL as training samples and so the number of the face samples is 40×9. And the testing samples are selected from the remaining data. After feature extraction, nearest neighbor classifier is used finally to classify. In the experiment, we used the Euclidean distance and the cosine distance as the distance measure. Table 1 and 2, respectively show the mean and variance in 10 different sets of experimental data.

From the experiments of Table 1 and 2, we can see that the recognition rate of DLDA/QR is slightly better than the traditional DLDA/QR algorithm and the performance of the both is better than Fisherface algorithm. Figure 2 shows that when the number of

Table 1: Recognition rate by Euclidean distance measure in ORL

	No. of Training sample		
Class (ϑ)	4	5	6
Fisherface	89.83±1.85	92.92±1.34	93.15±1.01
DLDA	92.56±1.32	95.01±0.76	95.12±0.76
DLDA/QR	92.70±1.24	96.05±0.45	96.54±1.10

Values are as Mean±SD

Table 2: Recognition rate by cosine distance measure in ORL

	No. of Training sample		
Class (ϑ)	4	5	6
Fisherface	91.60±0.65	93.20±1.54	93.30±1.55
DLDA	92.61±1.03	94.14±1.55	94.76±1.53
DLDA/QR	93.04±0.82	95.55±1.24	96.30±0.61

Values are as Mean±SD

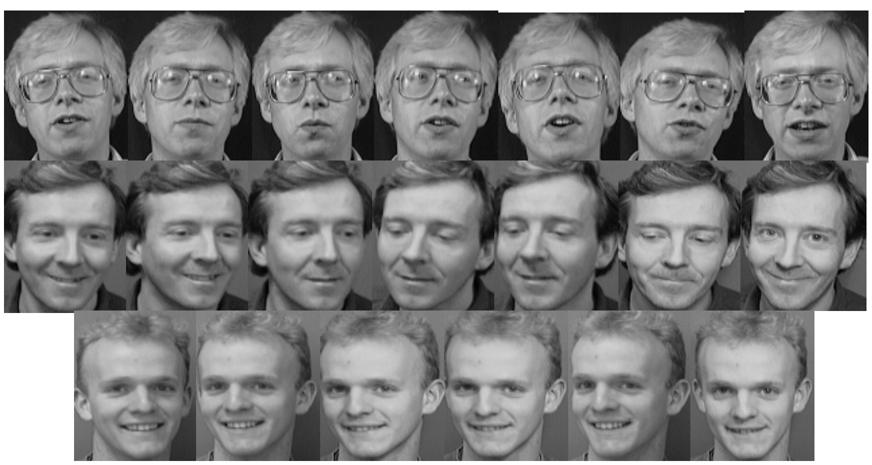

Fig. 1: Some facial image of ORL face database

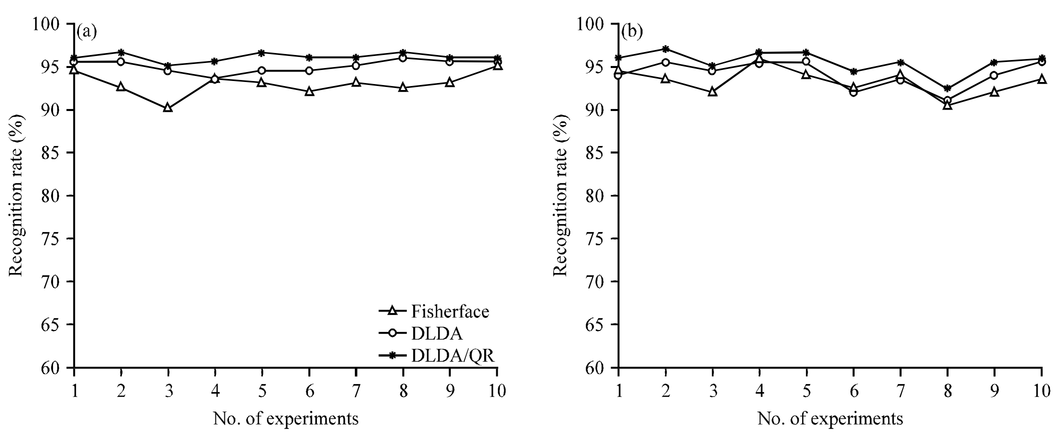

Fig. 2(a-b): Recognition rate comparison in different numbers of experiments (a) Euclidean distance measure and (b) Cosine distance measure

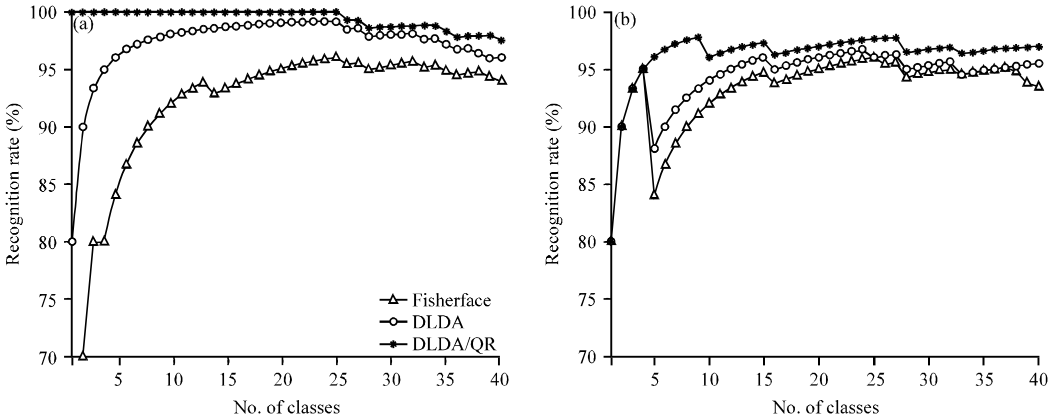

Fig. 3(a-b): Recognition rate comparison in different numbers of classes (a) Euclidean distance measure and (b) Cosine distance measure

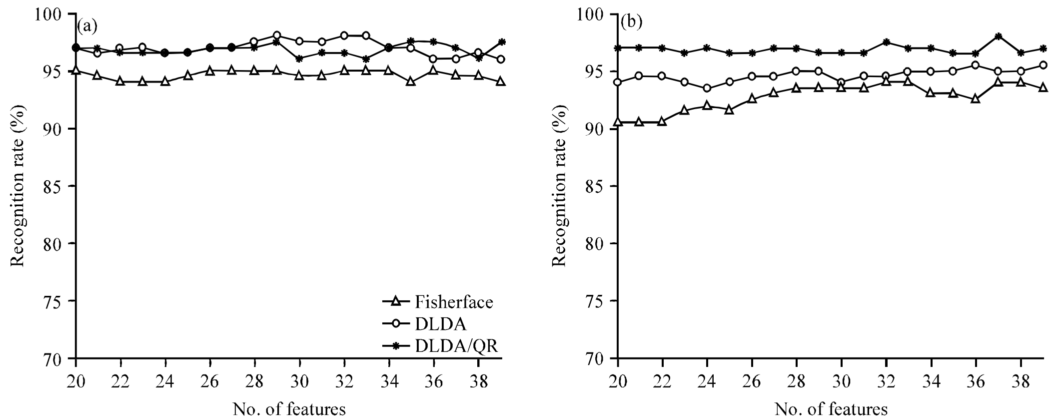

Fig. 4(a-b): Recognition rate comparison in different numbers of features (a) Euclidean distance measure and (b) Cosine distance measure

classes is 40, the feature dimension is 39, the training samples are 5 and in 10 different training sets, the DLDA/QR algorithm improves the stability of the original algorithm. Figure 3 shows when the final feature dimension is 39 and the number of extracted training samples is 5, the recognition rate comparison in different classes. Figure 4 shows when the fixed number of sample classes is 40 and the number of extracted training samples is 5, the recognition rate comparison in different feature numbers. The results show that the performance of DLDA/QR is better than the old methods and improves the stability and effectiveness.

CONCLUSION

For the defects of instability of the process in DLDA algorithm, QR decomposition algorithm is introduced into the computation. And based on this, a new direct linear analysis method named DLDA/QR is put forward. Through experiments on the ORL face database, the stability and validity of the new method is verified. The new method not only improved the effectiveness but also maintained better stability.

REFERENCES

Bai, R., M. Cao, Z. Su, W. Ostachowicz, and H. Xu, 2012. Fractal dimension analysis of higher-order mode shapes for damage identification of beam structures. Math. Prob. Eng.

Cao, L. and D.Y. Chen, 2011. Face recognition algorithm based on the wavelet approximation coefficients and non-parametric discriminant analysis. J. Beijing Inform. Sci. Technol. Univ., 6: 20-23.

Delgado-Gomez, D., J. Fagertun, B. Ersboll, F.M. Sukno and A.F. Frangi, 2009. Similarity-based fisherfaces. Pattern Recognit. Lett., 1: 1110-1116.

Gan, J.Y. and S.B. He, 2011. Nonlinear radon transform and its application to face recognition. Pattern Recognit. Artif. Intell., 4: 405-409.

Lee, F., K. Kotani, Q. Chen and T. Ohmi, 2012. Face recognition using adjacent pixel intensity difference quantization histogram combined with markov stationary features. Int. J. Applied Cryptogr., 4: 327-335.

Salmela, P., A. Burian, H. Sorokin and J. Takala, 2008. Complex-valued QR decomposition implementation for MIMO receivers. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing Proceeding, March 31-April 4, 2008, Las Vegas, NV., USA., pp: 1433-1436.

Ye, J.P. and Q. Li, 2005. A two-stage linear discriminant analysis via QR-decomposition. IEEE Trans. Pattern Anal. Mach. Intell., 27: 929-941.

Yong, X., W. Lucai and A. Qionglong, 2011. QR code recognition method based on correlation match. Chin. J. Sci. Instrum., 3: 571-576.