Regularisation of eigenfeatures by extrapolation of scatter-matrix in face-recognition problem

A. Sharma and K.K. Paliwal

The proposed face recognition algorithm regularises the within-class scatter matrix by extrapolating its eigenvalues. This way both the range space and the null space information of the within-class scatter matrix is utilised. The results are promising when the algorithm is experimented on face recognition datasets and compared with several other methods.

Introduction: In face recognition problems, the dimensionality of the feature space is very large compared to the number of training data samples available. It is therefore necessary to reduce the dimensionality of the space for improving the robustness (or generalisation capability) and computational complexity of the face recognition classifier. Many dimensionality reduction techniques have been proposed in the literature [1, 2]. Among them, the linear discriminant analysis (LDA) technique is perhaps the most widely studied technique for dimensionality reduction. The LDA technique finds an orientation $\mathbf{W}$ that transforms high dimensional feature vectors belonging to different classes to a lower dimensional feature space such that the projected feature vectors of a class on this lower dimensional space are well separated from the feature vectors of other classes. In the LDA technique, $\mathbf{W}$ can be obtained by the eigenvalue decomposition of the product of the inverse of within-class scatter matrix $\mathbf{S}_W$ and between-class scatter matrix $\mathbf{S}_B$, i.e. $\mathbf{S}_W^{-1}\mathbf{S}_B$. Since in face recognition problems, the dimensionality is large compared to the number of training samples available, $\mathbf{S}_W$ becomes singular and the evaluation of eigenvalues and eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ becomes impossible. There are, however, several techniques that overcome this drawback (see [3] for details).

Both the range space and the null space of $\mathbf{S}_W$ are meaningful for classification purpose. The range space based methods discard the null space of $\mathbf{S}_W$. On the other hand, the null space based methods [4] discard the range space and utilise only the null space of $\mathbf{S}_W$.

In this Letter we propose a method that utilises both the null space and the range space of $\mathbf{S}_W$ by extrapolating the eigenvalues of the range space into the null space of $\mathbf{S}_W$. The extrapolation is done by using exponential functions. It has an advantage over the regularised LDA method [5] that it does not compute extrapolation by using any heuristic approach.

Extrapolation of within-class scatter matrix: To describe the proposed approach, we first define the basic mathematical notions. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of n training samples in a d-dimensional feature space where $d > n$ and $\Omega = \{\omega_1, \omega_2, \ldots \omega_n\}$ be the class labels of samples of set $\mathcal{X}$. There are c class labels, i.e. $\omega_j \in \{1, 2, \ldots C\}$. Let the between-class scatter matrix $\mathbf{S}_B$, within-class scatter matrix $\mathbf{S}_W$ and total scatter matrix $\mathbf{S}_T$ be defined as in [6]. Further, we assume that the n samples in $\mathcal{X}$ are linearly independent. Therefore, the ranks of matrices $\mathbf{S}_T$, $\mathbf{S}_B$, and $\mathbf{S}_W$ can be given as $r_t = n - 1$, $r_b = c - 1$ and $r_w = n - c$, respectively. To reduce the complexity, a pre-processing step has been carried out by applying PCA to remove the null space of $\mathbf{S}_T$ which will transform the samples to the range space of $\mathbf{S}_T$. This will give us transformed within-class scatter matrix $\hat{\mathbf{S}}_W$ and transformed between-class scatter matrix $\hat{\mathbf{S}}_B$. The matrix $\hat{\mathbf{S}}_W$ can be decomposed as $\hat{\mathbf{S}}_W = \mathbf{U}\mathbf{D}_W^2\mathbf{U}^{\mathrm{T}}$, where $\mathbf{D}_W$ is a diagonal matrix the elements of which (arranged in descending order) are the square-root of the eigenvalues of $\hat{\mathbf{S}}_W$ and $\mathbf{U}$ is a matrix consisting of the corresponding eigenvectors as columns. Since in the face recognition application $\hat{\mathbf{S}}_W$ is singular, the eigenvalues $\mathbf{D}_W$ can be represented as:

$$\mathbf{D}_W = \begin{bmatrix} \Sigma_W & 0 \\ 0 & 0 \end{bmatrix}, \text{ where } \mathbf{D}_W \in \Re^{r_t \times r_t} \text{ and } \Sigma_W \in \Re^{r_w \times r_w}$$

Now we would like to estimate $\hat{\mathbf{D}}_W$ by extrapolating Σ_W as follows:

$$\hat{\mathbf{D}}_W = \begin{bmatrix} \Sigma_W & 0 \\ 0 & \hat{\Sigma}_W \end{bmatrix}, \text{ where } \hat{\Sigma}_W \text{ is obtained from the extrapolation of } \Sigma_W$$

The estimation $\hat{\Sigma}_W$ has been obtained by exponential fitting of the diagonal elements of Σ_W. The exponential function $g(z) = a\,\mathbf{exp}(bz) + c\,\mathbf{exp}(dz)$ has been utilised to estimate $\hat{\Sigma}_W$, where a, b, c and d are constants and z is any variable (here eigenvalues, i.e. $diag(\Sigma_W)$). Since the lower eigenvalues contain more noise which will make the estimation unreliable, not all the eigenvalues are used to estimate $\hat{\Sigma}_W$. Therefore, only the first few predominant eigenvalues have been utilised in the estimation. This way a more meaningful extrapolated curve could be obtained which will provide a reliable estimate of $\hat{\Sigma}_W$. Using the arguments given in [7], we can select the predominant eigenvalues either as those eigenvalues which are greater than the average of all the eigenvalues in $\mathbf{D}_W$ or by those which preserve 50% of the total sum of all the eigenvalues in $\mathbf{D}_W$. We found the later procedure to be better in the present classification task and, hence, it is used in this Letter. An example of extrapolation is shown in Fig. 1 using a face recognition dataset. The eigenvalues Σ_W are shown up to r_w and the estimated eigenvalues $\hat{\Sigma}_W$ are shown up to r_t.

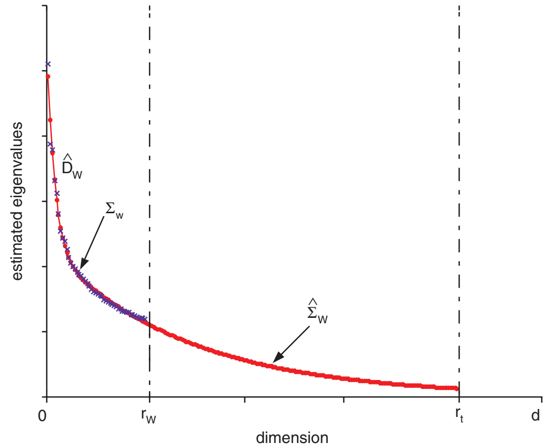

Fig. 1 *Eigenvalue extrapolation using exponential fitting*

Results: Two face datasets used in the experiments are ORL database [8] and AR database [9]. The ORL database contains 400 images of 40 people having 10 images per person. The dimension of feature sample is 10304. We have used five images per person for training and the remaining five images per person for testing. The AR database contains 100 classes. We have used seven images per subject for training and the remaining seven images per subject for testing. The dimension size is 4980. The proposed method was compared with the following methods: the range space method, the null space based method (OLDA) [4], the PCA plus LDA method [10], the regularised LDA method [5] and the maximum uncertainty LDA (MLDA) method [7]. Table 1 shows the recognition accuracy on both the datasets for all the methods. It can be seen that the proposed method outperforms the other methods.

Table 1: Recognition accuracy on ORL database and AR database

Database	Range space method (%)	Null space based method (OLDA) (%)	PCA plus LDA (%)	Regularised LDA (%)	MLDA (%)	Extrapolation method (%)
ORL database	87.0	91.0	81.5	90.5	91.0	91.5
AR database	70.1	81.6	82.0	82.6	77.3	86.6
Average	78.6	86.3	81.8	86.6	84.2	89.1

Conclusion: The regularisation of the within-class scatter matrix was done by extrapolating its eigenvalues. The exponential fitting was used for this estimation. The proposed method has shown promising results when compared with other methods on AR- and ORL-face recognition databases.

20 March 2010
doi: 10.1049/el.2010.0755
One or more of the Figures in this Letter are available in colour online.

A. Sharma and K.K. Paliwal (*Signal Processing Lab, Griffith University, Brisbane QLD-4111, Australia*)

E-mail: sharma_al@usp.ac.fj

References

1 Bishop, C.M.: 'Pattern recognition and machine learning' (Springer, 2006)

2 Sharma, A., and Paliwal, K.K.: 'A gradient linear discriminant analysis for small sample sized problem', *Neural Process. Lett.*, 2008, **27**, (1), pp. 17–24

3 Sharma, A., and Paliwal, K.K.: 'Rotational linear discriminant analysis for dimensionality reduction', *IEEE Trans. Knowl. Data Eng.*, 2008, **20**, (10), pp. 1336–1347

4 Ye, J.: 'Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems', *J. Mach. Learn. Res.*, 2005, **6**, pp. 483–502

5 Guo, Y., Hastie, T., and Tinshirani, R.: 'Regularized discriminant analysis and its application in microarrays', *Biostatistics*, 2007, **8**, (1), pp. 86–100

6 Duda, R.O., Hart, P.E., and Stork, D.G.: 'Pattern classification' (Wiley, New York, 2000)

7 Thomaz, C.E., Kitani, E.C., and Gillies, D.F.: 'A maximum uncertainty LDA-based approach for limited sample size problems – with application to face recognition'. Proc. 18th Brazilian Symp. on Computer Graphics and Image Processing, 2005 (IEEE CS Press), pp. 89–96

8 Samaria, F., and Harter, A.: 'Parameterization of a stochastic model for human face identification'. Proc. Second IEEE Workshop Applications of Computer Vision, 1994, pp. 138–142

9 Martinez, A.M.: 'Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (6), pp. 748–763

10 Swets, D.L., and Weng, J.: 'Using discriminative eigenfeatures for image retrieval', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (8), pp. 831–836