

IMPROVED PSEUDOINVERSE LINEAR DISCRIMINANT ANALYSIS METHOD FOR DIMENSIONALITY REDUCTION

KULDIP K. PALIWAL* and ALOK SHARMA*^{†,‡,§,¶}

**Signal Processing Laboratory, School of Engineering
Griffith University, QLD-4111, Brisbane, Australia*

†University of the South Pacific, Fiji

*‡Laboratory of DNA Information Analysis
Human Genome Center, Institute of Medical Science
University of Tokyo, 4-6-1 Shirokanedai
Minato-ku, Tokyo 108-8639, Japan*

§aloks@ims.u-tokyo.ac.jp

¶sharma_al@usp.ac.fj

Received 4 November 2010

Accepted 22 September 2011

Published 11 May 2012

Pseudoinverse linear discriminant analysis (PLDA) is a classical method for solving small sample size problem. However, its performance is limited. In this paper, we propose an improved PLDA method which is faster and produces better classification accuracy when experimented on several datasets.

Keywords: Pseudoinverse; linear discriminant analysis; dimensionality reduction; computational complexity.

1. Introduction

Dimensionality reduction is an important aspect of pattern classification. It helps in improving the robustness (or generalization capability) of the pattern classifier and in reducing its computational complexity. The linear discriminant analysis (LDA) method⁵ is a well-known dimensionality reduction technique studied in the literature. The LDA technique finds an orientation matrix \mathbf{W} that transforms high-dimensional feature vectors belonging to different classes to lower dimensional feature vectors such that the projected feature vectors of a class are well separated from the feature vectors of other classes. The orientation \mathbf{W} is obtained by maximizing the Fisher's criterion function $J_1(\mathbf{W}) = |\mathbf{W}^T \mathbf{S}_B \mathbf{W}| / |\mathbf{W}^T \mathbf{S}_W \mathbf{W}|$, where \mathbf{S}_B is between-class scatter matrix and \mathbf{S}_W is within-class scatter matrix. It has been shown in the literature that modified version of Fisher's criterion $J_2(\mathbf{W}) = |\mathbf{W}^T \mathbf{S}_B \mathbf{W}| / |\mathbf{W}^T \mathbf{S}_T \mathbf{W}|$ produces similar results, where \mathbf{S}_T is total scatter matrix.⁶

In the conventional LDA technique, the matrix \mathbf{S}_W or matrix (depending upon the criterion taken) needs to be nonsingular. However, in many pattern classification applications these matrices become singular. This problem is known as small sample size (SSS) problem.⁶ In order to overcome this problem, several methods have been proposed in the literature.^{a,2,4,11,13,16–19,21,22} Among these methods, the pseudoinverse LDA (PLDA) method¹⁸ stands as a forerunner and a classical method for solving SSS problem. The PLDA method has been widely studied.^{10,11,15,18} It finds the orientation matrix \mathbf{W} by computing eigenvalue decomposition (EVD) of $\mathbf{S}_W^+ \mathbf{S}_B$, where \mathbf{S}_W^+ is the pseudoinverse of \mathbf{S}_W . However, this has a problem that its computational complexity is $O(d^3)$, which is prohibitively high (when the dimensionality d is very large). Due to this reason, the PLDA method has been cited in the literature in several papers but hardly compared with other techniques for pattern classification. In order to reduce this computational complexity, Liu *et al.*¹¹ introduced a fast PLDA method. In their method the orientation \mathbf{W} is computed by finding the range space of \mathbf{S}_W followed by the range space of \mathbf{S}_B . The null space of \mathbf{S}_W is discarded in this process. Their fast PLDA method has been shown equivalent to the PLDA method.¹¹ Though the fast PLDA method is computationally faster than the PLDA method, it has a drawback. It discards null space of \mathbf{S}_W in computing the orientation matrix \mathbf{W} , which has been shown to contain useful discriminant information for classification.⁴ Considering this drawback we propose the use of modified Fisher's criterion $J_2(\mathbf{W}) = |\mathbf{W}^T \mathbf{S}_B \mathbf{W}| / |\mathbf{W}^T \mathbf{S}_T \mathbf{W}|$ for the pseudoinverse method. Therefore, in the proposed method the orientation \mathbf{W} is computed by finding the range space of \mathbf{S}_T followed by the range space of \mathbf{S}_B . In this method the null space of \mathbf{S}_T has been discarded. It is known that discarding the null space of \mathbf{S}_T does not cause any loss of discriminant information.⁹ Thus, this has an advantage over the fast PLDA method that it improves the classification performance. In addition, it is shown to be computationally faster than fast PLDA method.

2. Improved PLDA Method

In order to describe improved PLDA method, we first define some notations. Let \mathfrak{X} be a set of n training vectors in a d -dimensional feature space, and $\Omega = \{\omega_i : i = 1, 2, \dots, c\}$ be the finite set of c class labels, where ω_i denotes the i th class label. The set \mathfrak{X} can be subdivided into c subsets $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_c$ (where subset \mathfrak{X}_i belongs to ω_i); i.e. $\mathfrak{X}_i \subset \mathfrak{X}$ and $\mathfrak{X}_1 \cup \mathfrak{X}_2 \cup \dots \cup \mathfrak{X}_c = \mathfrak{X}$. Let n_i be the number of samples in class ω_i such that:

$$n = \sum_{i=1}^c n_i.$$

^aAll these methods except the method by Zhang *et al.*²² try to maximize Fisher's criterion or modified Fisher's criterion either in one stage or in two stages. In the Zhang's method, the difference between $\mathbf{S}_W^{-1} \mathbf{S}_B$ and $\mathbf{W} \mathbf{D} \mathbf{W}^T$ is minimized, where \mathbf{W} is an orthogonal matrix and \mathbf{D} is a diagonal matrix. This method also deals with the case when $N > d$ (where N is the number of samples and d is the dimensionality).

The samples or vectors of set \mathcal{X} can be written as:

$$\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}, \quad \text{where } \mathbf{x}_j \in \mathbf{R}^d.$$

Let $\boldsymbol{\mu}_j$ be the centroid of \mathcal{X}_j and $\boldsymbol{\mu}$ be the centroid of \mathcal{X} , then the between-class scatter matrix \mathbf{S}_B is given by

$$\mathbf{S}_B = \sum_{j=1}^c n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^\top.$$

The within-class scatter matrix \mathbf{S}_W is defined as

$$\mathbf{S}_W = \sum_{j=1}^c \mathbf{S}_j,$$

where $\mathbf{S}_j = \sum_{\mathbf{x} \in \mathcal{X}_j} (\mathbf{x} - \boldsymbol{\mu}_j)(\mathbf{x} - \boldsymbol{\mu}_j)^\top$.

The total-class scatter matrix \mathbf{S}_T is defined as

$$\mathbf{S}_T = \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^\top.$$

The matrix \mathbf{S}_T can also be formed as $\mathbf{S}_T = \mathbf{A}\mathbf{A}^\top$, where $\mathbf{A} \in \mathfrak{R}^{d \times n}$ is defined as

$$\mathbf{A} = [(\mathbf{x}_1 - \boldsymbol{\mu}), (\mathbf{x}_2 - \boldsymbol{\mu}), \dots, (\mathbf{x}_n - \boldsymbol{\mu})]. \quad (1)$$

In a similar way, \mathbf{S}_B can be formed as $\mathbf{S}_B = \mathbf{B}\mathbf{B}^\top$, where rectangular matrix $\mathbf{B} \in \mathfrak{R}^{d \times c}$ can be defined as

$$\mathbf{B} = [\sqrt{n_1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}), \sqrt{n_2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}), \dots, \sqrt{n_c}(\boldsymbol{\mu}_c - \boldsymbol{\mu})].$$

Let the ranks of matrices $\mathbf{S}_T, \mathbf{S}_B$, and \mathbf{S}_W be t, b and w , respectively.

The orientation matrix \mathbf{W} can be obtained by first finding the range space of \mathbf{S}_T followed by the range space of \mathbf{S}_B , i.e. if EVD of \mathbf{S}_T is

$$\mathbf{S}_T = \mathbf{U}_1 \boldsymbol{\Lambda} \mathbf{U}_1^\top,$$

where $\mathbf{U}_1 \in \mathfrak{R}^{d \times t}$ corresponds to the range space of \mathbf{S}_T and $\boldsymbol{\Lambda} \in \mathfrak{R}^{t \times t}$ is a diagonal matrix, then $\hat{\mathbf{S}}_T = \mathbf{U}_1^\top \mathbf{S}_T \mathbf{U}_1$ and $\hat{\mathbf{S}}_B = \mathbf{U}_1^\top \mathbf{S}_B \mathbf{U}_1$, the orientation matrix \mathbf{W} can be found by finding EVD of $\hat{\mathbf{S}}_T^+ \hat{\mathbf{S}}_B$.

In order to find the range of \mathbf{S}_T , we can compute EVD of $\mathbf{A}^\top \mathbf{A} \in \mathfrak{R}^{n \times n}$ instead of $\mathbf{S}_T = \mathbf{A}\mathbf{A}^\top \in \mathfrak{R}^{d \times d}$, this will significantly reduce the computational complexity.⁶

If the eigenvectors and eigenvalues of $\mathbf{A}^\top \mathbf{A} \in \mathfrak{R}^{n \times n}$ are $\mathbf{E} \in \mathfrak{R}^{n \times n}$ and $\mathbf{D} \in \mathfrak{R}^{n \times n}$, respectively, then

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= \mathbf{E}\mathbf{D}\mathbf{E}^\top \\ &= [\mathbf{E}_1, \mathbf{E}_2] \begin{bmatrix} \mathbf{D}_1 & \\ & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_1^\top \\ \mathbf{E}_2^\top \end{bmatrix}, \\ &\quad \text{where } \mathbf{E}_1 \in \mathfrak{R}^{n \times t}, \mathbf{E}_2 \in \mathfrak{R}^{n \times (n-t)} \text{ and } \mathbf{D}_1 \in \mathfrak{R}^{t \times t} \\ &= \mathbf{E}_1 \mathbf{D}_1 \mathbf{E}_1^\top \end{aligned} \quad (2)$$

and orthonormal eigenvectors \mathbf{U}_1 defining the range space of \mathbf{S}_T can be given as

$$\mathbf{U}_1 = \mathbf{A}\mathbf{E}_1\mathbf{D}_1^{-1/2}.$$

Since discarding the null space of \mathbf{S}_T does not cause any loss of discriminant information,⁹ we can use $\mathbf{U}_1 \in \mathbb{R}^{d \times t}$ to transform the original d -dimensional space to a lower t -dimensional space. The matrices \mathbf{A} and \mathbf{B} can be written in the lower dimensional space as follows:

$$\begin{aligned} \hat{\mathbf{A}} &= \mathbf{U}_1^T \mathbf{A} \in \mathbb{R}^{t \times n} \\ &= \mathbf{D}_1^{-1/2} \mathbf{E}_1^T \mathbf{A}^T \mathbf{A} \\ &= \mathbf{D}_1^{-1/2} \mathbf{E}_1^T \mathbf{E}_1 \mathbf{D}_1 \mathbf{E}_1^T \quad (\text{from Eq. (2)}) \\ &= \mathbf{D}_1^{1/2} \mathbf{E}_1^T. \end{aligned} \tag{3}$$

The matrix $\hat{\mathbf{B}}$ can be economically constructed from $\hat{\mathbf{A}}$. In order to do this, we first write the transformed matrix $\hat{\mathbf{A}}$ as $\hat{\mathbf{A}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ and then compute $\hat{\mathbf{B}}$ as

$$\hat{\mathbf{B}} = \left[\frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} \mathbf{v}_j, \frac{1}{\sqrt{n_2}} \sum_{j=n_1+1}^{n_1+n_2} \mathbf{v}_j, \dots, \frac{1}{\sqrt{n_c}} \sum_{j=n_1+n_2+\dots+n_{c-1}+1}^n \mathbf{v}_j \right]. \tag{4}$$

This will give transformed between-class scatter $\hat{\mathbf{S}}_B = \hat{\mathbf{B}}\hat{\mathbf{B}}^T$. From Eq. (3), the transformed total-scatter matrix $\hat{\mathbf{S}}_T = \hat{\mathbf{A}}\hat{\mathbf{A}}^T = \mathbf{D}_1^{1/2} \mathbf{E}_1^T \mathbf{E}_1 \mathbf{D}_1^{1/2} = \mathbf{D}_1^{\hat{\mathbf{A}}}$, this will give $\hat{\mathbf{S}}_T^+ \hat{\mathbf{S}}_B = \mathbf{D}_1^{-1} \hat{\mathbf{B}}\hat{\mathbf{B}}^T$. The EVD of $\hat{\mathbf{S}}_T^+ \hat{\mathbf{S}}_B$ will give eigenvectors $\hat{\mathbf{W}} \in \mathbb{R}^{t \times h}$ (where h is less than or equal to the rank of $\hat{\mathbf{S}}_T^+ \hat{\mathbf{S}}_B$, in other words, $1 \leq h \leq c - 1$) corresponding to its leading eigenvalues. The orientation matrix \mathbf{W} can be obtained as follows:

$$\mathbf{W} = \mathbf{U}_1 \hat{\mathbf{W}} = \mathbf{A}\mathbf{E}_1\mathbf{D}_1^{-1/2} \hat{\mathbf{W}}.$$

The implementation of the improved PLDA method is summarized in Table 1.

3. Computational Complexity and Storage Requirements

In this section, the computational complexity and storage requirements of the proposed improved PLDA method are discussed. We also compare its computational complexity and storage requirements with PLDA and fast PLDA methods.¹¹ The

Table 1. Improved PLDA method.

1. Construct matrix \mathbf{A} from Eq. (1).
2. Compute eigenvalues $\mathbf{E}_1 \in \mathbb{R}^{n \times t}$ and eigenvectors $\mathbf{D}_1 \in \mathbb{R}^{t \times t}$ of $\mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$.
3. Compute transformed matrix $\hat{\mathbf{B}}$ (from Eq. (4)).
4. Compute the EVD of $\mathbf{D}_1^{-1} \hat{\mathbf{B}}\hat{\mathbf{B}}^T$ to get $\hat{\mathbf{W}} \in \mathbb{R}^{t \times h}$ (where $1 \leq h \leq c - 1$).
5. Compute $\mathbf{W} = \mathbf{D}_1^{-1/2} \hat{\mathbf{W}}$, then $\mathbf{W} \leftarrow \mathbf{E}_1 \mathbf{W}$ and then $\mathbf{W} \leftarrow \mathbf{A}\mathbf{W}$.

Table 2. Computational complexity of the improved PLDA method.

Steps	Complexities
Step 1. Formation of matrix \mathbf{A}	$2dn$
Step 2. Multiplication of $\mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$ and computation of $\mathbf{E}_1 \in \mathbb{R}^{n \times t}$ and $\mathbf{D}_1 \in \mathbb{R}^{t \times t}$ using eigenvalue decomposition of $\mathbf{A}^T \mathbf{A}$	$dn^2 + 17n^3$
Step 3. Computation of transformed matrix $\hat{\mathbf{B}}$ (from Eq. (4))	n^2
Step 4. Multiplication of $\mathbf{D}_1^{-1} \hat{\mathbf{B}} \hat{\mathbf{B}}^T$ and its EVD	$(2t^2c + t^2) + 17t^3$
Step 5. Multiplication of $\mathbf{W} = \mathbf{D}_1^{-1/2} \hat{\mathbf{W}}$, $\mathbf{W} \leftarrow \mathbf{E}_1 \mathbf{W}$ and $\mathbf{W} \leftarrow \mathbf{A} \mathbf{W}$	$t(c-1) + 2nt(c-1) + 2dn(c-1)$
Total estimated	$dn^2 + 2dnc + 2dn + (34n^3 + 4n^2c + 2n^2 + nc)$ (since $t \approx n$ and $c - 1 \approx c$)

estimated computational complexity of the improved PLDA method is listed in Table 2.

Since the dimensionality d in a SSS problem is very large compared to the number of training samples n ($d \gg n$), the computational complexity of the improved PLDA method boils down to $dn^2 + 2dnc + 2dn$ flops.

In the PLDA method, the computation of EVD of $\mathbf{S}_W^+ \mathbf{S}_B$ is required. This has the computational complexity of $O(d^3)$. The fast PLDA method requires approximately $3dn^2 + 2dnc + 3dn$ flops. When the dimensionality is very large $d \gg n$ then the proposed method is approximately three times faster than the fast PLDA method. The storage requirements of all the methods are same. In all the methods, the orientation matrix $\mathbf{W} \in \mathbb{R}^{d \times h}$ computed during training session is required to be stored for the testing session which requires approximately dh storage.

4. Datasets and Experimentation

The following types of datasets are utilized for the experimentation: DNA microarray gene expression data, face recognition data and text classification data. We have also used randomly generated data to investigate the effect of dimensionality d on the computation time. Five DNA microarray gene expression datasets are utilized. We use the splitting of the data into the training and test samples as provided by the distributors.^b For face recognition, AR database¹² is utilized for the experimentation. A subset of AR database is used here with 1400 face images from 100 persons (14 images per person). Training set contains seven images per person and

^bMost of the DNA microarray gene expression datasets can be downloaded from <http://sdmc.lit.org.sg/GEDatasets/Datasets.html> or <http://cs1.shu.edu.cn/gzli/data/mirror-kentridge.html> or <http://leo.ugr.es/elvira/DBCRepository>.

Table 3. Datasets used in the experimentation.

Datasets	Class	Dimension	Number of training samples	Number of testing samples
Acute Leukemia ⁷	2	7129	38	34
ALL Subtype ²⁰	7	12558	215	112
MLL ¹	3	12582	57	15
GCM ¹⁴	14	16063	144	54
Lung Cancer ⁸	2	12533	32	149
Face AR ¹²	100	19800	700	700
Dexter ³	2	20000	300	300

Table 4. Classification accuracy (in percentage) and CPU time on datasets.

Datasets	Fast PLDA method		Improved PLDA method		Fisherface LDA method		Null space-based method	
	Classn. accuracy	CPU time	Classn. accuracy	CPU time	Classn. accuracy	CPU time	Classn. accuracy	CPU time
Acute Leukemia	88.2	0.05	97.1	0.02	100.0	0.10	97.1	0.15
ALL Subtype	59.8	1.00	85.7	0.53	80.4	2.67	86.6	3.34
MLL	80	0.15	100.0	0.06	100.0	0.36	100.0	0.55
GCM	46.3	0.90	70.4	0.29	59.3	1.98	70.4	2.81
Lung Cancer	94.6	0.09	98.0	0.03	98.0	0.14	98.0	0.23
Face AR14	83.1	19.17	88.9	10.47	83.0	29.53	85.0	45.05
Dexter	67.3	2.13	93.7	1.19	93.3	6.44	93.7	7.75

the remaining seven images per person are used for testing. The dimensionality d is 19,800. We use a subset of Dexter dataset³ for text classification in a bag-of-word representation. This dataset has sparse continuous input variables. The description of all the datasets is given in Table 3.

The fast PLDA method and the improved PLDA method have been experimented on all the above datasets. In addition, Fisherface LDA method¹⁷ and null space-based method¹⁹ have been used for comparison purpose. The nearest neighbor classifier has been used to classify test feature vector. The classification accuracy and CPU time of these methods are depicted in Table 4. It can be observed from Table 4 that improved PLDA method is outperforming fast PLDA method in terms of classification accuracy and CPU time. Furthermore, the improved PLDA method is computationally efficient than Fisherface LDA and null space-based methods. It can also be observed (in terms of classification accuracy) that improved PLDA method is outperforming Fisherface LDA method and proving as good as results with the null space-based method.

We have also generated random data and increased its dimensionality from 10,000 to 100,000, to measure the CPU time of the improved PLDA and fast PLDA methods. The CPU time curve as a function of data dimensionality is shown in Fig. 1. It can be seen from the figure that as the dimensionality of data increases the improved PLDA method performs faster than the fast PLDA method.

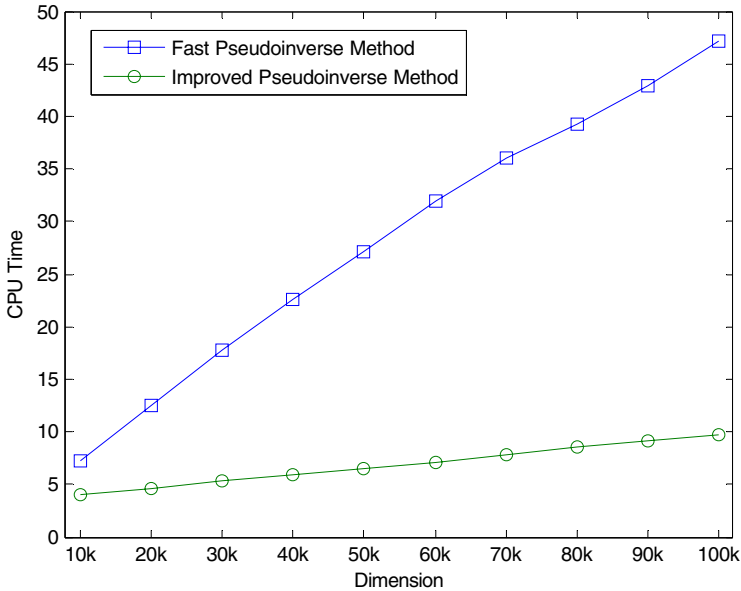


Fig. 1. CPU time as a function of data dimensionality.

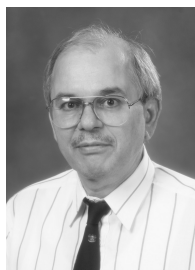
5. Conclusion

An improved PLDA method has been proposed in this paper. It is outperforming other pseudoinverse methods in terms of computation complexity and classification accuracy when experimented on several datasets.

References

1. S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub and S. J. Korsmeyer, MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia, *Nat. Genet.* **30** (2002) 41–47.
2. P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces versus Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Patt. Anal. Mach. Intell.* **19**(7) (1997) 711–720.
3. C. L. Blake and C. J. Merz, UCI repository of machine learning databases, Irvine, CA, University of Calif., Dept. of Information and Comp. Sci. (1998), <http://www.ics.uci.edu/~mllearn>.
4. L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin and G.-J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, *Patt. Recogn.* **33** (2000) 1713–1726.
5. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
6. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press Inc., Hartcourt Brace Jovanovich, Publishers, 1990).

7. T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander, Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science* **286** (1999) 531–537.
8. G. J. Gordon, R. V. Jensen, L.-L. Hsiao, S. R. Gullans, J. E. Blumenstock, S. Ramaswamy, W. G. Richards, D. J. Sugarbaker and R. Bueno, Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma, *Cancer Res.* **62** (2002) 4963–4967.
9. R. Huang, Q. Liu, H. Lu and S. Ma, Solving the small sample size problem of LDA, *Proc. ICPR 2002*, Vol. 3 (2002), pp. 29–32.
10. W. J. Krzanowski, P. Jonathan, W. V. McCarthy and M. R. Thomas, Discriminant analysis with singular covariance matrices: Methods and applications to spectroscopic data, *Appl. Stat.* **44** (1995) 101–115.
11. J. Liu, S. C. Chen and X. Y. Tan, Efficient pseudo-inverse linear discriminant analysis and its nonlinear form for face recognition, *Int. J. Patt. Recogn. Artif. Intell.* **21**(8) (2007) 1265–1278.
12. A. M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(6) (2002) 748–763.
13. K. K. Paliwal and A. Sharma, Improved direct LDA and its application to DNA gene microarray data, *Patt. Recogn. Lett.* **31**(16) (2010) 2489–2492.
14. S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander and T. R. Golub, Multiclass cancer diagnosis using tumor gene expression signatures, *Proc. Natl. Acad. Sci. U.S.A.* **98**(26) (2001) 15149–15154.
15. S. Raudys and R. P. W. Duin, On expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix, *Patt. Recogn. Lett.* **19** (1998) 385–392.
16. A. Sharma and K. K. Paliwal, Regularisation of eigenfeatures by extrapolation of scattermatrix in face-recognition problem, *Electron. Lett.* **46**(10) (2010) 450–475.
17. D. L. Swets and J. Weng, Using discriminative eigenfeatures for image retrieval, *IEEE Trans. Patt. Anal. Mach. Intell.* **18**(8) (1996) 831–836.
18. Q. Tian, M. Barbero, Z. H. Gu and S. H. Lee, Image classification by the Foley-Sammon transform, *Opt. Eng.* **25**(7) (1986) 834–840.
19. J. Ye, Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems, *J. Mach. Learn. Res.* **6** (2005) 483–502.
20. E. J. Yeoh, M. E. Ross, S. A. Shurtleff, W. K. Williams, D. Patel, R. Mahfouz, F. G. Behm, S. C. Raimondi, M. V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C. H. Pui, W. E. Evans, C. Naeve, L. Wong and J. R. Downing, Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling, *Cancer* **1**(2) (2002) 133–143.
21. H. Yu and J. Yang, A direct LDA algorithm for high-dimensional data-with application to face recognition, *Patt. Recogn.* **34** (2001) 2067–2070.
22. T. Zhang, B. Fang, Y. Y. Tang, Z. Shang and G. He, A least-squares model to orthogonal linear discriminant analysis, *Int. J. Patt. Recogn. Artif. Intell.* **24**(4) (2010) 635–650.



Kuldip K. Paliwal received the B.S. degree from Agra University, Agra, India, in 1969, the M.S. degree from Aligarh Muslim University, Aligarh, India, in 1971 and the Ph.D. degree from Bombay University, Bombay, India, in 1978.

He has been carrying out research in the area of speech processing since 1972. He has worked at a number of organizations including Tata Institute of Fundamental Research, Bombay, India, Norwegian Institute of Technology, Trondheim, Norway, University of Keele, U. K., AT&T Bell Laboratories, Murray Hill, New Jersey, U.S.A., AT&T Shannon Laboratories, Florham Park, New Jersey, U.S.A., and Advanced Telecommunication Research Laboratories, Kyoto, Japan. Since July 1993, he has been a professor at Griffith University, Brisbane, Australia, in the School of Microelectronic Engineering. His current research interests include speech recognition, speech coding, speaker recognition, speech enhancement, face recognition, image coding, pattern recognition and artificial neural networks. He has published more than 300 papers in these research areas.

Dr. Paliwal is a Fellow of Acoustical Society of India. He has served the IEEE Signal Processing Society's Neural Networks Technical Committee as a founding member from 1991 to 1995 and the Speech Processing Technical Committee from 1999 to 2003. He was an Associate Editor of the IEEE Transactions on Speech and Audio Processing during the periods 1994–1997 and 2003–2004. He also served as Associate Editor of the IEEE Signal Processing Letters from 1997 to 2000. He is also in the Editorial Board of the IEEE Signal Processing Magazine. He was the General Co-Chair of the Tenth IEEE Workshop on Neural Networks for Signal Processing (NNSP2000). He has co-edited two books: *Speech Coding and Synthesis* (published by Elsevier), and *Speech and Speaker Recognition: Advanced Topics* (published by Kluwer). He has received IEEE Signal Processing Society's best (senior) paper award in 1995 for his paper on LPC quantization. He is currently serving the Speech Communication journal (published by Elsevier) as its Editor-in-Chief.



Alok Sharma received the B.Tech degree from the University of the South Pacific (USP), Suva, Fiji, in 2000 and the MEng degree, with an academic excellence award, and the Ph.D. degree in the area of pattern recognition from Griffith University, Brisbane, Australia, in 2001 and 2006, respectively.

He is currently a research fellow at the University of Tokyo. He is also with the Signal Processing Laboratory, Griffith University and the University of the South Pacific. He participated in various projects carried out in conjunction with Motorola (Sydney), Auslog Pty. Ltd. (Brisbane), CRC Micro Technology (Brisbane), and the French Embassy (Suva). His research interests include pattern recognition, computer security, and human cancer classification. He reviewed several articles from journals like IEEE Trans. NN, IEEE Trans. SMC, Part A: SH, IEEE Journal on STSP, IEEE Trans. KDE, IEEE Tans. EC, Computers and Security, Pattern Recognition, etc. He is a member of IEEE.