	MITSUBISHI ELECTRIC ITE-VIL	Title:	Face Recognition by Fisher and Scatter Linear Discriminant Analysis		
Internal Reference:	**VIL03-D326**	Author:	M.Z. Bober, K. Kucharski and W. Skarbek	Rev.	**A**
Publication Date:	November 2003	Published in:	Lecture Notes in Computer Science, Publisher: Springer-Verlag Heidelberg, ISSN: 0302-9743, Volume: Volume 2756 / 2003, November 2003, Pages: 638 – 645, ISBN: 3-540-40730-8		

Face Recognition by Fisher and Scatter Linear Discriminant Analysis

Miroslaw Bober
Krzysztof Kucharski
Wladyslaw Skarbek

Abstract

Fisher linear discriminant analysis (FLDA) based on variance ratio is compared with scatter linear discriminant (SLDA) analysis based on determinant ratio. It is shown that each optimal FLDA data model is optimal SLDA data model but not opposite. The novel algorithm 2SS4LDA (*two singular subspaces for LDA*) is presented using two singular value decompositions applied directly to normalized multiclass input data matrix and normalized class means data matrix. It is controlled by two singular subspace dimension parameters q and r, respectively. It appears in face recognition experiments on the union of MPEG-7, Altkom, and Feret facial databases that 2SS4LDA reaches about 94% person identification rate and about 0.21 average normalized mean retrieval rank. The best face recognition performance measures are achieved for those combinations of q,r values for which the variance ratio is close to its maximum, too. None such correlation is observed for SLDA separation measure.

Face Recognition by Fisher and Scatter Linear Discriminant Analysis

Mirosław Bober[1], Krzysztof Kucharski[2], and Władysław Skarbek[2]

[1] Visual Information Laboratory, Mitsubishi Electric, Guilford, UK
[2] Faculty of Electronics and Information Technology
Warsaw University of Technology, Poland
W.Skarbek@ire.pw.edu.pl

Abstract. Fisher linear discriminant analysis (FLDA) based on variance ratio is compared with scatter linear discriminant (SLDA) analysis based on determinant ratio. It is shown that each optimal FLDA data model is optimal SLDA data model but not opposite. The novel algorithm 2SS4LDA (*two singular subspaces for LDA*) is presented using two singular value decompositions applied directly to normalized multiclass input data matrix and normalized class means data matrix. It is controlled by two singular subspace dimension parameters q and r, respectively. It appears in face recognition experiments on the union of MPEG-7, Altkom, and Feret facial databases that 2SS4LDA reaches about 94% for the person identification rate and about 0.21 for the average normalized mean retrieval rank. The best face recognition performance measures are achieved for those combinations of q, r values for which the variance ratio is close to its maximum, too. None such correlation is observed for SLDA separation measure.

1 Introduction

Linear Discriminant Analysis, shortly LDA, deals with the training sequence $X = [x_1, \ldots, x_L]$ of multidimensional data vectors ($x_i \in \mathbb{R}^N,\ i = 1, \ldots, L$).

In general the data vectors are obtained using a measurement process for objects from certain classes $\mathcal{C}_1, \ldots, \mathcal{C}_J$, i.e. i-th element of X belongs to class $\mathcal{C}_{j(i)}$. Let the number of elements x_i which represent class j be L_j, i.e. $L = L_1 + \cdots + L_J$. We can identify elements in X extracted from j-th class by the index set I_j: $I_j \doteq \{i : x_i \text{ represents class } \mathcal{C}_j\}$.

LDA is based on statistical concepts of data variances and covariances. The *unbiased vector within-class variance* $\mathrm{var}_w(X)$ and the *unbiased vector between-class variance* $\mathrm{var}_b(X)$ have the form : $\mathrm{var}_w(X) \doteq \frac{1}{L-J} \sum_{j=1}^{J} \sum_{i \in I_j} \|x_i - \overline{x}^j\|^2$, $\mathrm{var}_b(X) \doteq \frac{1}{J-1} \sum_{j=1}^{J} L_j \|\overline{x}^j - \overline{x}\|^2$, where the class vector mean $\overline{x}^j$ and grand vector mean $\overline{x}$ are: $\overline{x}^j \doteq \sum_{i \in I_j} x_i / L_j$, $\overline{x} \doteq \sum_{i=1}^{L} x_i / L$.

The data covariances within classes are represented by the *within-class scatter matrix* $S_w(X) \doteq \sum_{j=1}^{J} \sum_{i \in I_j} (x_i - \overline{x}^j)(x_i - \overline{x}^j)^t / (L - J)$, while the between class covariances are defined by the matrix $S_b(X) \doteq \sum_{j=1}^{J} L_j (\overline{x}^j - \overline{x})(\overline{x}^j - \overline{x})^t / (J - 1)$ called the *between-class scatter matrix.*

For the given dimension r of the feature vector $y = W^t x$, LDA attempts to find a linear transformation matrix $W \in \mathbb{R}^{N \times r}$, $W = [w_1, \ldots, w_r]$, $w_i \in \mathbb{R}^N$ for the training sequence X which gives the best separation for the classes. Then the scatter matrices and data variances are transformed accordingly: $S_w(Y) = W^t S_w(X) W$, $\mathrm{var}_w(Y) = \mathrm{tr}(W^t S_w(X) W)$, $S_b(Y) = W^t S_b(X) W$, $\mathrm{var}_b(Y) = \mathrm{tr}(W^t S_b(X) W)$, where $Y = [W^t x_1, \ldots, W^t x_L]$ is the sequence of feature vectors for X.

In this paper two measures $f(W)$ and $g(W)$ of separation concept are considered. The first one, originally proposed by Fisher ([5]), takes into account the ratio of between-class variance to within-class variance. While the second one, very commonly cited in face recognition applications (e.g. [4, 6, 7]) replaces variances by determinants ($|\cdot|$) of corresponding scatter matrices ([2]) as measures of data scattering. Both measures lead to two modeling techniques with two different families of LDA models $\mathcal{F}_r$ and $\mathcal{S}_r$ defined for $1 \leq r \leq N$:

1. Fisher linear discriminant analysis (FLDA) with models in $\mathcal{F}_r(X)$:

$$f(W) \doteq \frac{\mathrm{tr}(W^t S_b W)}{\mathrm{tr}(W^t S_w W)}$$

$$\mathcal{F}_r(X) \doteq \{W \in \mathbb{R}^{N \times r} : W = \arg\max f(W),\ W^t S_w W = I,\ W \perp \ker(S_w)\}$$

2. Scatter linear discriminant (SLDA) with models in $\mathcal{S}_r(X)$:

$$g(W) \doteq \frac{|W^t S_b W|}{|W^t S_w W|}$$

$$\mathcal{S}_r(X) \doteq \{W \in \mathbb{R}^{N \times r} : W = \arg\max g(W),\ |W^t S_w W| \neq 0\}$$

The paper is organized as follows. In section 2 algorithmic characterization of FLDA models is given and efficient algorithm 2SS4LDA is described. In section 3 mutual relations between FLDA and SLDA models are presented. Experiments on facial databases are discussed in section 4.

2 Algorithm for FLDA models

The requirement $W^t S_w W = I$ for FLDA model ensures that class mean shifted, data vector components become decorrelated and of unit variance in LDA coordinates. The column vectors of $W = [w_1, \ldots, w_r]$ of FLDA model are sought within the hyper-ellipsoid $\mathcal{B} \doteq \{x : x^t S_w x = 1, x \perp \ker(S_w)\}$.

Let us consider the reduced eigenvalue decomposition (REVD) for $S_w = U_{q_0} \Lambda_{q_0} U_{q_0}^t$, where the first $q_0 = \mathrm{rank}(S_w)$, columns in U and Λ are chosen, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \geq \cdots \geq \lambda_N$. Then the search space has the form: $\mathcal{B} = \{x : x^t S_w x = 1, x \perp \ker(S_w)\} = \{A\alpha : A \doteq U_{q_0} \Lambda_{q_0}^{-1/2},\ \alpha \in \mathbb{R}^{q_0}, \|\alpha\| = 1\}$. Now, the behavior of the objective function $w^t S_b w$ can be analyzed using REVD for $A^t S_b A \doteq V_{r_0} \Sigma_{r_0} V_{r_0}^t$, where $r_0 = \mathrm{rank}(A^t S_b A) : w^t S_b w = \alpha^t A^t S_b A \alpha = \alpha^t V_{r_0} \Sigma_{r_0} V_{r_0}^t \alpha$. Maximization of $w^t S_b w$ with the constraint $1 - \alpha^t \alpha = 0$ by

Lagrangian multipliers leads to the stationary points $\alpha_k = v_k$ with value σ_k, $k = 1, \ldots, r_0$. Therefore the optimal point for the Fisher goal function $f(W) = f(w_1, \ldots, w_r)$ can be combined from locally optimal points $w_k = Av_k$ of the quadratic form $w^t S_w w$ for $r \leq r_0$:

$$f(W) = \mathrm{tr}(W^t S_b W)/r = \sum_{k=1}^{r} w_k^t S_b w_k / r \tag{1}$$

$$rf(W) \leq \sum_{k=1}^{r} v_k^t A^t S_b A v_k = \sum_{k=1}^{r} v_k^t V_{r_0} \Sigma_{r_0} V_{r_0}^t v_k = \sum_{k=1}^{r} \sigma_k \tag{2}$$

Hence the optimal $W = AV_r$, $r \leq r_0$, and FLDA models can be compactly characterized as follows: $\mathcal{F}_r(X) = \{W \in \mathbb{R}^{N \times r} : W = U_{q_0} \Lambda_{q_0}^{-1/2} V_r\}$, $S_w = U_{q_0} \Lambda_{q_0} U_{q_0}^t$, $A = U_{q_0} \Lambda_{q_0}^{-1/2}$, $A^t S_b A = V_{r_0} \Sigma_{r_0} V_{r_0}^t$, and for $r \leq r_0$ $W^t S_b W = V_r^t (A^t S_b A) V_r = V_r^t V_{r_0} \Sigma_{r_0} V_{r_0}^t V_r = \Sigma_r$.

By the above FLDA properties we propose the novel algorithm 2SS4LDA (*two singular subspaces for LDA*). It is based on two singular value approximations applied directly for the normalized multiclass input data matrix and the normalized class means data matrix. It is controlled by subspace dimension parameters q and r. The first singular subspace of dimension q is designed for original data and it used to compute new coordinates for class means. The second singular subspace is built in this new coordinates. In a sense it is nested SVD procedure. The feature vectors are computed using r left singular vectors spanning the second singular subspace.

Algorithm 1. 2SS4LDA - Two Singular Subspaces for LDA

Input *Data sequence* $X = [x_1, \ldots, x_L], x_i \in \mathbb{R}^N$, *class membership vector* I, *desired LDA feature vector dimension* r, *and desired first singular subspace dimension* q.

Output *Corrected values of singular subspace dimensions* q, r *and FLDA model* $W \in \mathbb{R}^{N \times r}$.

Method *Perform the following steps:*

1. *Compute the global centroid* c *and class centroids:* $C \leftarrow [c_1, \ldots, c_J]$
2. *Perform centroid shifting and normalization for data matrices* X, C :

$$\textit{if } i \in I_j \textit{ then } y_i \leftarrow (x_i - c_j)/\sqrt{L - J}, i = 1, \ldots, L,$$

$$d_j \leftarrow (c_j - c)\sqrt{L_j/(J-1)},\ j = 1, \ldots, J$$

3. *Find the singular subspace of* $Y = [y_1, \ldots, y_L]$ *by performing SVD for* Y *obtaining* $q_0 \doteq \mathit{rank}(Y)$ *left singular vectors* $U_{q_0} \leftarrow [u_1, \ldots, u_{q_0}]$ *corresponding to positive singular values* $\Lambda_{q_0}^{1/2} \leftarrow [\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_{q_0}}]$

4. If $q > q_0$ then $q \leftarrow q_0$
If $q < q_0$ then $U_q \leftarrow [u_1, \ldots, u_q]$ and $\Lambda_q^{1/2} \leftarrow [\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_q}]$
5. Compute whitening projection matrix: $A_q \leftarrow U_q \Lambda_q^{-1/2}$
6. Make whitening projection for normalized class means:

$$d_j \leftarrow A_w^t d_j, \; j = 1, \ldots, J$$

7. Find the singular subspace of $D \doteq [d_1, \ldots, d_J]$ by performing SVD for D obtaining $r_0 \doteq rank(D)$ left singular vectors $V_{r_0} \leftarrow [v_1, \ldots, v_{r_0}]$ corresponding to positive singular values
8. If $r > r_0$ then $r \leftarrow r_0$
If $r < r_0$ then $V_r \leftarrow [v_1, \ldots, v_r]$
9. Compute FLDA model, i.e. the projection matrix W: $W \leftarrow A_q V_r$

Note that for $q_0 = \text{rank}(S_w)$ the above algorithm produces the exact FLDA model and for $q < q_0$ its approximation is obtained. In face recognition problem the optimal value of the measure $f(W)$ is obtained for q much less than the rank of normalized data matrix Y. It means that the better mean class separation occurs for Y projected (with whitening) onto its singular subspace of much lower dimension than the dimension of subspace spanned by vectors in Y.

Note that q_0, the rank of Y (equal to the rank of S_w) is bounded by $\min(L-J, N)$, while r_0, the rank of D (equal to the rank of S_b) is bounded by $\min(J-1, q_0)$. Therefore, the constraint for the feature vector size is $r \leq \min(L-J, J-1, N)$.

In FLDA modeling for face recognition, J is the number of persons in the training database and in our experiments it is less than N and $L-J$. Hence if $q = q_0$ then the rank of data matrix D is equal to the number of training persons minus one: $r_0 = J - 1$.

3 On relation of FLDA and SLDA models

It is well known (by the analysis of stationary points for $g(W)$) that the optimal SLDA models are sought by solving the generalized eigenvalue problem (cf. [2]) of the following form $S_b W = S_w W \Lambda_r$, where $\Lambda_r = \text{diag}(\lambda_1, \ldots, \lambda_r)$, $\lambda_1 \geq \cdots \geq \lambda_r > 0$.

Let $\mathcal{E}_r(X)$ denotes the set of all solutions of the above generalized eigenvalue problem. It means that $\mathcal{S}_r(X) \subset \mathcal{E}_r(X)$.

It can be easily proved that local maxima for $g(W)$ are always global, i.e. for optimal $W \in \mathcal{S}_r(X)$ the scatter separation measure has always the same value:

$$g(W) = \frac{|W^t S_b W|}{|W^t S_w W|} = \frac{|W^t S_w W \Lambda_r|}{|W^t S_w W|} = \frac{|W^t S_w W||\Lambda_r|}{|W^t S_w W|} = \prod_{k=1}^{r} \lambda_i$$

It also implies that $\mathcal{E}_r(X) \subset \mathcal{S}_r(X)$ and in conclusion $\mathcal{S}_r(X) = \mathcal{E}_r(X)$.

By Lagrangian optimization of $w^t S_b w$ at $w^t S_w w = 1$ we get $S_b w = \mu S_w w$. Hence the optimal FLDA models have to satisfy the generalized eigenvalue equation and therefore $\mathcal{F}_r(X) \subset \mathcal{E}_r(X)$.

The above inclusion is strict. To show it let us observe that by equation (2) and by $\mathrm{tr}(I) = r$ if $W \in \mathcal{F}_r(X)$ we have $f(W) = \sum_{k=1} \sigma_k / r$.

On the other hand there is simple relation between any two models $W_1, W_2 \in \mathcal{E}_r(X)$: there exists a block diagonal matrix $\Gamma = \mathrm{diag}(B_1, \ldots, B_{r'})$ such that $W_1 = W_2 \Gamma$. Suppose that also $W_2 \in \mathcal{F}_r(X)$. If there is no multiple eigenvalues then $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_r)$ is diagonal what is the case for real life data matrices. Then for any matrix $Z \in \mathbb{R}^{r \times r}$: $\mathrm{tr}(\Gamma^t Z \Gamma) = \sum_{k=1}^{r} \gamma_k^2 Z_{kk}$. Hence if there is no multiple eigenvalues:

$$f(W_1) = f(W_2 \Gamma) = \frac{\mathrm{tr}(\Gamma^t W_2^t S_b W_2 \Gamma)}{\mathrm{tr}(\Gamma^t W_2^t S_w W_2 \Gamma)} = \sum_{k=1}^{r} \alpha_k \sigma_k \leq \frac{\sum_{k=1}^{r} \sigma_k}{r}$$

where $\alpha_k \doteq \gamma_k^2 / (\gamma_1^2 + \cdots + \gamma_r^2)$. The above inequality becomes equality only if $\gamma_1 = \cdots = \gamma_r$.

In general case when $\Gamma = \mathrm{diag}(B_1, \ldots, B_{r'})$, denoting by $Z_{kk}^{(b)}$ the submatrix of Z corresponding to B_k we have: $\mathrm{tr}(\Gamma^t Z \Gamma) = \sum_{k=1}^{r'} \mathrm{tr}(B_k^t Z_{kk}^{(b)} B_k)$. Hence

$$f(W_1) = f(W_2 \Gamma) = \frac{\mathrm{tr}(\Gamma^t W_2^t S_b W_2 \Gamma)}{\mathrm{tr}(\Gamma^t W_2^t S_w W_2 \Gamma)} = \frac{\mathrm{tr}(\Gamma^t \Sigma_r \Gamma)}{\mathrm{tr}(\Gamma^t \Gamma)} = \sum_{k=1}^{r'} \alpha_k \sigma_k \leq \frac{\sum_{k=1}^{r} \sigma_k}{r}$$

where $\alpha_k \doteq \mathrm{tr}(B_k^t B_k) / (\mathrm{tr}(B_1^t B_1) + \cdots + \mathrm{tr}(B_{r'}^t B_{r'}))$. The above inequality becomes equality only if $\mathrm{tr}(B_1^t B_1) = \cdots = \mathrm{tr}(B_{r'}^t B_{r'})$.

Therefore, independently of multiplicity of eigenvalues, there are always optimal models in SLDA which are not optimal in FLDA. In summary:

$$\mathcal{F}_r(X) \subsetneq \mathcal{S}_r(X) = \mathcal{E}_r(X)$$

We have shown that there are optimal SLDA models which are not optimal FLDA models. Can we find optimal SLDA which are not FLDA models, but they achieve maximum value of $f(W)$? The answer is positive. Namely, those SLDA models $W \in \mathbb{R}^{N \times r}$ for which all diagonal elements of the within-class scatter matrix $W^t S_w W$ are identical, give maximum value of LDA separation measure $f(W)$, i.e. for W satisfying the equation $S_b W = S_w W \Lambda$:

$$f(W) = \frac{\mathrm{tr}(W^t S_b W)}{\mathrm{tr}(W^t S_w W)} = \frac{\mathrm{tr}(W^t S_w W \Lambda)}{\mathrm{tr}(W^t S_w W)} = \frac{\sum_{k=1}^{r} \alpha_k \lambda_k}{\sum_{k=1}^{r} \alpha_k}$$

where $\alpha_k \doteq (W^t S_w W)_{kk} \geq 0$.

Hence, the maximum value of $f(W)$ (equal to $\sum_k \lambda_k / r$) is achieved if and only if $\alpha_1 = \cdots = \alpha_r$. Of course, the class of such matrices W includes $\mathcal{F}_r(X)$. Inclusion is strict as any matrix $W' = \sqrt{\alpha} W$ obtained by uniform scaling by $\sqrt{\alpha} \neq 1$ of FLDA optimal model W gives SLDA optimal model which is not FLDA (since

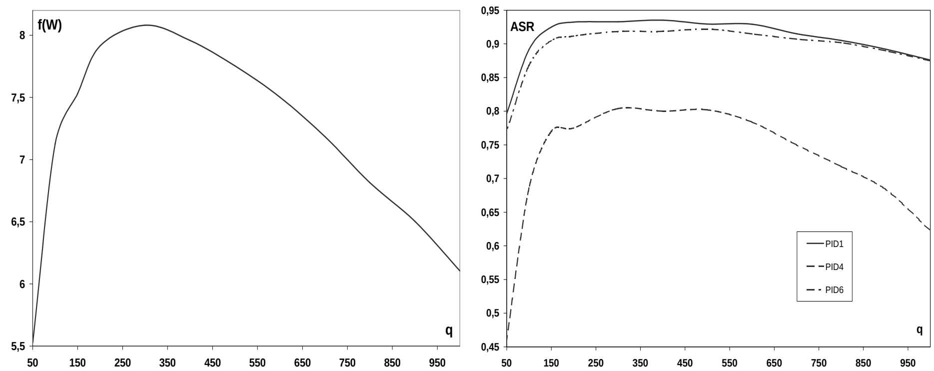

Fig. 1. Left: FLDA separation measure $f(W)$ as function of singular subspace dimension q for class mean shifted original data (model dimension $r = 48$). **Right:** Corresponding average success rate (ASR) using 2SS4LDA algorithm in three person identification experiments (descriptor size: 240 bits).

$W'^t S_w W' = \alpha I$). Let us denote the class of such models by $\mathcal{F}_r^{(\alpha)}(X)$. Then $\mathcal{F}_r(X) = \mathcal{F}_r^{(1)}(X)$.

Another observation: if multiplicities of all eigenvalues in Λ_r are equal to one then any SLDA model makes diagonalisation of $S_w : W^t S_w W = \Gamma^t \Gamma$, i.e. LDA projected variables are uncorrelated. This very commonly happens in practice.

If $\Gamma^t \Gamma$ is not a scalable form of unit matrix, i.e. $\Gamma^t \Gamma \neq \alpha I$, then the model W is optimal SLDA model which achieves less value of separation measure $f(W)$ than optimal FLDA models: $W \in \mathcal{S}_r(X) - \bigcup_{\alpha>0} \mathcal{F}_r^{(\alpha)}(X)$.

Finally, the relation between two concepts of optimal separation of classes can be summarised by the following inclusions:

$$\mathcal{F}_r(X) \subsetneq \bigcup_{\alpha>0} \mathcal{F}_r^{(\alpha)}(X) \subsetneq \mathcal{S}_r(X) = \mathcal{E}_r(X) \ .$$

4 Experiments

In order to perform comparative tests of our method in face recognition task we use images from three databases: Altkom (1200 images, 80 persons), MPEG-7 (3175 images, 635 persons) and FERET subset (4000 images, 875 persons).

Every image has size 46x56 and eyes manually located. Initial preprocessing includes automatic background cutting off.

The half of Altkom and MPEG databases constitutes the training set on which the model matrix W is calculated according to the proposed algorithm 2SS4LDA for FLDA models and generalized Schur algorithm (GSCHUR) for SLDA models (cf. [3]). The other half along with FERET images is used for extracting feature vectors and testing.

The four experiments are considered conforming MPEG-VCE face recognition visual core experiment (cf. [1]): image retrieval(FIR) and three various

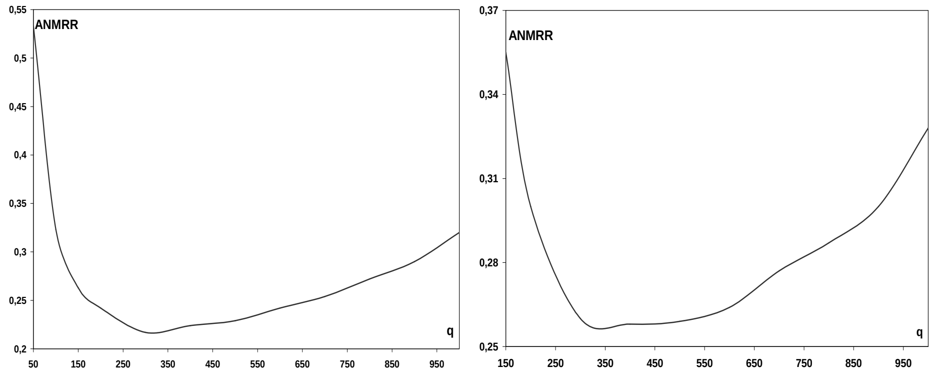

Fig. 2. Average normalized mean retrieval ratio (ANMRR) in function of subspace dimension q in FLDA experiments (**Left:** $r = 48$, 240 bits; **Right:** $r = 128$, 640 bits).

person identifications(PID). In FIR every single image from testing set becomes a query while in PID the specific disjoint subsets of testing set are chosen for query and test respectively.

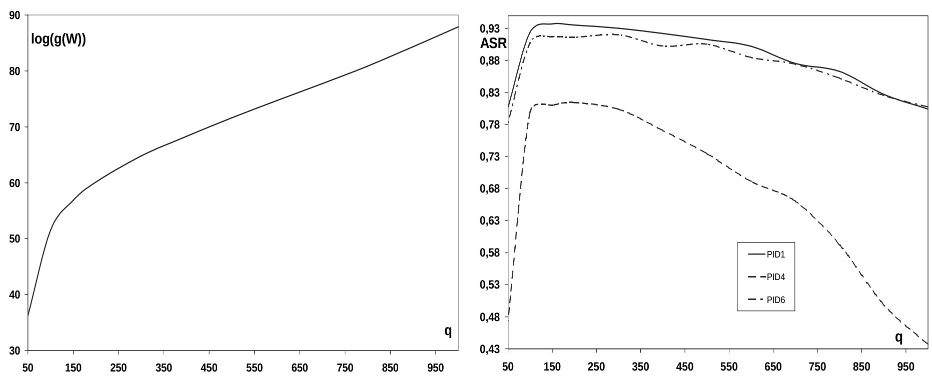

Fig. 3. Left: SLDA separation measure $g(W)$ as function of dimension q for principal subspace which is used for input data matrix singularity conditioning (model dimension $r = 48$). **Right:** Corresponding average success rate (ASR) using GSCHUR algorithm in three person identification experiments (descriptor size: 240 bits).

5 Conclusions

We have proved that each optimal model of data sequence X in Fisher linear discriminant analysis is also optimal in scatter linear data discriminant analysis. Moreover, there is infinity of models optimal in SLDA which are not optimal in FLDA. There is also infinity of SLDA models which maximize Fisher class separation measure $f(W)$. There exists mathematically closed formulas for those interesting subclasses of SLDA class of data models. They are based on the types of within-class scatter matrix diagonalisation.

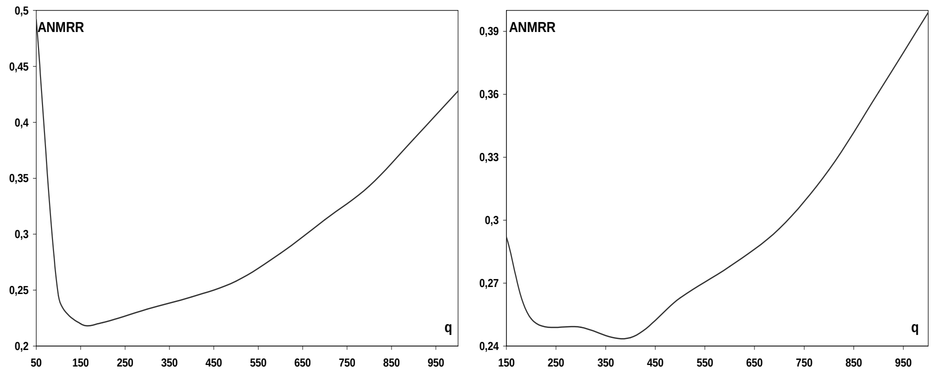

Fig. 4. Average normalized mean retrieval ratio (ANMRR) in function of subspace dimension q in SLDA experiments (**Left:** $r = 48$, 240 bits; **Right:** $r = 128$, 640 bits).

The proposed algorithm 2SS4LDA performs two singular value decompositions applied directly to normalized multiclass input data matrix and normalized class means data matrix. By tuning two singular subspace dimension parameters q and r, we can optimize ratio of between and within-class variance what leads to better performance in face recognition application.

We have observed that using GSCHUR algorithm with regularization by projection of the original data X onto singular subspace of dimension q gives the best results very close to the best results of 2SS4LDA but for quite different settings of q and r.

High correlation of the class separation measure in FLDA with face recognition performance was found in contrary to SLDA case.

On very demanding facial databases of MPEG-7 VCE, the LDA classifier built by proposed algorithm gives 94% for the person identification rate and about 0.21 for the average normalized mean retrieval rank.

References

1. Bober M., Description of MPEG-7 Visual Core Experiments, ISO/IEC JTC1/SC29/WG11, report N4925, July 2002, Klagenfurt.
2. Devijver P.A., Kittler J., Pattern Recognition: A Statistical Approach, Prentice Hall, Englewood Cliffs, N.J., 1982
3. Golub G., Van Loan C., Matrix Computations. Baltimore: Johns Hopkins University Press, 1996
4. Li Y., Kittler J., Matas J., Effective Implementation of Linear Discriminant Analysis for Face Recognition and Verification, CAIP'99, pages 234-242, 1999
5. Ripley B.D., Pattern Recognition and Neural Networks. Cambridge University Press, 1996
6. Swets D.L., Weng J., Using Discriminant Eigenfeatures for Image Retrieval, IEEE Trans. on PAMI, 18(8):831-837, August 1996
7. Zhao W., Krishnaswamy A., Chellappa R., Swets D.L., Weng J., Discriminant Analysis of Principal Components for Face Recognition, In 3rd International Conference on Automatic Face and Gesture Recognition, pages 336-341, 1998