A two-stage linear discriminant analysis for face-recognition

Alok Sharma \(^{a,b,c,*}\), Kuldip K. Paliwal \(^a\)

\(^a\) Signal Processing Lab, Griffith University, Australia
\(^b\) School of Engineering & Physics, University of the South Pacific, Fiji
\(^c\) Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan

A R T I C L E   I N F O

Article history:
Received 20 September 2011
Available online 10 February 2012
Communicated by K.A. Toh

Keywords:
Two-stage linear discriminant analysis
Small sample size problem
Classification accuracy

A B S T R A C T

A two-stage linear discriminant analysis technique is proposed that utilizes both the null space and range space information of scatter matrices. The technique regularizes both the between-class scatter and within-class scatter matrices to extract the discriminant information. The regularization is conducted in parallel to give two orientation matrices. These orientation matrices are concatenated to form the final orientation matrix. The proposed technique is shown to provide better classification performance on face recognition datasets than the other techniques.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Linear discriminant analysis (LDA) is a well known technique for dimensionality reduction and feature extraction (Duda et al., 2000; Sharma and Paliwal, 2006, 2008, 2010, 2012; Chen et al., 2000; Lu et al., 2003a,b, 2005; Yang et al., 2003; Yu and Yang, 2001; Swets and Weng, 1996; Belhumeur et al., 1997; Ye, 2005; Guo et al., 2007; Thomaz et al., 2005; Huang et al., 2002; Tian et al., 1986; Zhao et al., 2003; Jiang et al., 2008; Gao and Davis, 2006; Paliwal and Sharma, 2010, 2011; Mandal et al., 2010). Dimensionality reduction plays crucial role in the face recognition problem. It is generally applied for improving robustness (or generalization capability) and reducing computational complexity of the face recognition classifier. In the LDA technique, the orientation matrix \(W\) is computed from the eigenvalue decomposition (EVD) of \(S_W S_B\) (Duda et al., 2000), where \(S_W \in \mathbb{R}^{d \times d}\) is within-class scatter matrix, \(S_B \in \mathbb{R}^{d \times d}\) is between-class scatter matrix and \(d\) is the dimensionality of feature space. In the face recognition problem, the matrix \(S_W\) becomes singular and its inverse computation becomes impossible. Several techniques are reported in the literature that overcome this drawback of LDA (Chen et al., 2000; Lu et al., 2003a,b, 2005; Yang et al., 2003; Yu and Yang, 2001; Swets and Weng, 1996; Belhumeur et al., 1997; Ye, 2005; Guo et al., 2007; Thomaz et al., 2005; Sharma and Paliwal, 2010, 2012; Huang et al., 2002; Tian et al., 1986; Zhao et al., 2003; Jiang et al., 2008; Paliwal and Sharma, 2010, 2011; Mandal et al., 2010).

In LDA, there are four informative spaces namely, null space of \(S_W (S_W^\text{null})\), range space of \(S_W (S_W^\text{range})\), null space of \(S_B (S_B^\text{null})\) and range space of \(S_B (S_B^\text{range})\). All these four individual spaces have significant discriminant information (refer Appendix I for empirical demonstration). To approximate the inverse computation of \(S_W\), different combinations of these spaces are used in the literature for finding the orientation matrix \(W\). For an instance the pseudo-inverse technique (Tian et al., 1986) uses \(S_W^\text{null}\) and \(S_B^\text{null}\) to compute the orientation matrix. The regularized LDA technique (Zhao et al., 2003) uses \(S_W^\text{null}\), \(S_B^\text{null}\) and \(S_B^\text{range}\). However, due to the use of small value of regularization parameter (compared to the large eigenvalues of \(S_W\)), the \(S_W^\text{null}\) gets de-emphasize in the inverse operation of \(S_W\). Therefore, the influential spaces in the regularized LDA technique are \(S_W^\text{null}\) and \(S_B^\text{null}\). Similarly, the null LDA technique (Chen et al., 2000) uses \(S_W^\text{null}\) and \(S_B^\text{null}\). These techniques basically utilize two spaces in the orientation matrix computation and discard the other two spaces. Since the individual spaces contribute crucial discriminant information for classification, discarding some spaces would sacrifice the classification performance of the classifier. Theoretically, if all the four spaces can be inherited appropriately in the computation of orientation matrix \(W\) then the classification performance can be improved further.

In this paper, we exploit ways of utilizing all the four spaces. The inclusion of all the spaces of scatter matrices is done in two analyses. Fig. 1 illustrates the proposed strategy. The orientation matrix can be computed from the input data by carrying out two discriminant analyses in parallel. In the first analysis, the orientation matrix \(W_1\) is computed by retaining top eigenvalues and eigenvectors of \(S_W^{-1} S_B\), where non-singular matrix \(S\) is the approximation of singular matrix \(S\). This will retain \(S_W^\text{null}\) and \(S_B^\text{range}\). In the second analysis, the orientation matrix \(W_2\) is obtained by retaining top eigenvalues and eigenvectors of \(S_B^{-1} S_W\). This will retain \(S_W^\text{range}\) and \(S_B^\text{null}\). The orientation matrices obtained by these two analyses are...
concatenated to get the final orientation matrix \( W \), i.e., \( W = [W_1, W_2] \). For brevity we call the proposed technique the two-stage LDA technique. The non-singular approximation \( S \) of singular matrix \( S \) can be evaluated in two ways: (1) using regularized LDA technique (Zhao et al., 2003) where \( b \) is discarded.

\[ S_{rb} = U_B D_B^\dagger U_B^T \]

and \( \tilde{S}_W = U_W D_W^\dagger U_W^T \)

where \( D_W \in \mathbb{R}^{n \times n} \) and \( D_B \in \mathbb{R}^{(c-r) \times (c-r)} \) are diagonal matrices whose elements (arranged in descending order) are the square-root of the eigenvalues of \( S_W \) and \( S_B \), respectively; and \( U_W \in \mathbb{R}^{n \times n} \) and \( U_B \in \mathbb{R}^{(c-r) \times (c-r)} \) are orthogonal matrices consisting of the corresponding eigenvectors as columns. Since the rank of \( S_W \) is \( r_w \), the matrix \( U_W \) can be formed as \( U_W = [U_{Wrr}, U_{Wln}] \) where \( U_{Wrr} \in \mathbb{R}^{r_w \times r_w} \) corresponds to the range space of \( S_W \) and \( U_{Wln} \in \mathbb{R}^{(n-r_w) \times (n-r_w)} \) corresponds to the null space of \( S_W \). In a similar way, we can write \( U_B = [U_{Brr}, U_{Bln}] \) where \( U_{Brr} \in \mathbb{R}^{(c-r) \times (c-r)} \) corresponds to the range space of \( S_B \) and \( U_{Bln} \in \mathbb{R}^{(c-r) \times (c-r)} \) corresponds to the null space of \( S_B \).

### 3. Two-stage LDA technique

It is well known in the literature that the null space of \( \tilde{S}_W \) contains crucial information for classification (Chen et al., 2000; Ye, 2005). The null space based LDA techniques retain the null space information of \( S_W \), however, they discard the range space information of \( S_W \). It has been seen that the range space information of \( S_W \) is also important for classification (Swets and Weng, 1996; Belhumeur et al., 1997) and by discarding it could penalize classification performance. Some techniques (e.g., Guo et al., 2007; Zhao et al., 2003; Jiang et al., 2008; Sharma and Paliwal, 2010) estimates non-singular within-class scatter matrix \( S_B \) by adding a small positive constant (known as regularization parameter) to the eigenvalues of \( S_B \) (Guo et al., 2007; Zhao et al., 2003) or by extrapolating the eigenvalues of \( S_W \) in its null space (Jiang et al., 2008; Sharma and Paliwal, 2010). Therefore, obtaining the eigenvectors corresponding to the top eigenvalues of \( \tilde{S}_W \) in these techniques the null space information of \( S_W \) and the range space information of \( S_B \) are effectively retained. Although, the range space information of \( S_W \) is utilized in these techniques, it has very less influence as it is de-emphasized in the inverse operation of \( S_W \) (see Fig. 2). Nonetheless, theoretically the latter implementation would contain more information than the former techniques. To see the qualitative contribution of \( S_B \) in obtaining the orientation matrix, we decompose \( S_W \) into its eigenvalues and eigenvectors as

\[ S_W = U_W D_W U_W^T \]

where diagonal matrix \( D_W \) is \( \Sigma_W = \begin{bmatrix} \Sigma_{rr} & 0 \\ 0 & \Sigma_{ln} \end{bmatrix} \) with \( \Sigma_W \in \mathbb{R}^{n \times n} \) and \( \Sigma_W \in \mathbb{R}^{(c-r) \times (c-r)} \) is the estimation or regularization of eigenvalues \( \Sigma_W \).

From Eq. (5), \( \tilde{S}_W \) can be formed as

\[ \tilde{S}_W = [U_{Wrr}, U_{Wln}] \begin{bmatrix} \Sigma_{rr} & 0 \\ 0 & \Sigma_{ln} \end{bmatrix} \begin{bmatrix} U_{Wrr}^T \\ U_{Wln}^T \end{bmatrix} = U_{Wrr} \Sigma_{rr} U_{Wrr}^T \]

The EVD of Eq. (8) can be computed and the range space information of \( S_B \) can be used in the formation of orientation matrix. Three things can be observed here:

(1) The null space of \( S_B \) is discarded.

(2) The range space information of within-class scatter matrix in the inverse operation is de-emphasized.

(3) The null space of the product \( S_W S_B \) is discarded.
Fig. 2. This figure uses regularization method to get non-singular estimate $S_W$ from the singular matrix $S_W$ and illustrates the de-emphasis of the range space information of $S_W$ in its inverse operation. The terms $r_w$ and $r_r$ are the ranks of $S_W$ and $S_R$, respectively. The region between 1 and $r_w$ is the range space of $S_W$ and the region between $r_w$ and $r_r$ is the null space of $S_W$. The eigenvalues of $S_W$ are added by the regularization parameter $\alpha$ which gives the eigenvalues of $S_W$ (i.e., $S_W = S_W + \alpha I$). The regularization parameter is usually a small quantity obtained by performing cross-validation procedure on the training feature vectors. This parameter addition helps in defining the eigenvalues of $S_W$ in the null space region. Thereby enabling the inverse operation of $S_W$. The small eigenvalues of $S_W$ (in the null space) get emphasized in the inverse operation. These eigenvalues are used as weighting coefficients for their corresponding eigenvectors and therefore the eigenvectors of $S_W$ in the range space are de-emphasized.

It is known that though the null space of $S_b$ is less effective, it contains some useful information for classification (Gao and Davis, 2006; Paliwal and Sharma, 2010, Appendix I). Therefore, theoretically if the null space of $S_b$ is included in computing the orientation matrix then the classification performance can be improved. Furthermore, if the range space of $S_b$ can be utilized effectively then it can help in retaining more information. Next, if the eigenvectors of $S_W^{-1}S_b$ are represented by $E = [E_b, E_r]$ (where $E_b \in \mathbb{R}^{C-1 \times r_b}$ and $E_b \in \mathbb{R}^{r_r \times (r_r - r_b)}$) then it is possible that some eigenvalues (which are not in the range space of $S_W^{-1}S_b$) are complex valued which would give complex eigenvectors as columns of $E_b \in \mathbb{R}^{C-1 \times (r_r - r_b)}$ and cannot be included in the formation of orientation matrix. Some of the eigenvalues of a singular matrix can become complex due to limited size of the hardware (Golub and Loan, 1996). Since the matrix $S_W^{-1}S_b$ is positive semi-definite and singular (with rank $r_b$), then theoretically it should produce $r_b$ positive eigenvalues and the remaining eigenvalues should be zero. However, due to the hardware limitations, it may produce some very small non-zero eigenvalues (positive or negative). The small negative eigenvalues will lead to complex eigenvectors. For example, if the size of $S_W^{-1}S_b$ is $10 \times 10$ and its rank is 3 then it will give 3 eigenvectors corresponding to the positive eigenvalues which are defined as its range space $E_b$. The remaining 7 eigenvectors define the null space $E_r$ some of its eigenvectors corresponding to very small negative eigenvalues will be complex valued. In our implementation we use only the $E_b$ and discard $E_r$.

In order to retain more information for the purpose of improving the classification performance further, we investigate a strategy to: (1) include the null space of $S_b$, (2) include the range space of $S_W$, and (3) extract the null space information of $S_W^{-1}S_b$.

One strategy would be to estimate eigenvalues for the null space of $S_b$ (as done e.g., for $S_W$ in regularized LDA technique) and perform eigenvalue decomposition of $S_W^{-1}S_b$. The term $S_b$ is the regularized or estimated matrix of $S_b$ and can be defined as

$$S_b = U_b \tilde{D}_b U_b^T \tag{9}$$

where $\tilde{D}_b = \begin{bmatrix} \Sigma_b & 0 \\ 0 & \Sigma_b \end{bmatrix}$ and $\Sigma_b = \sigma_{b}^{\alpha} S_W^{-\frac{1}{2}} E_r E_b^{-T} S_W^{-\frac{1}{2}}$ is the estimation of eigenvalues in the null space of $S_b$. This strategy may satisfy above points 1 and 3. However, it could have either no effect on classification performance or can deteriorate the classification performance. See Appendix II for details.

In order to satisfy the above three points we can do as follows. Consider a matrix $C$ which has leading and lagging eigenvectors represented by $L$ and $G$, respectively. Then the leading and lagging eigenvectors of $C^-$ can be given by $G$ and $L$, respectively. Therefore, $S_b$ can be estimated to be non-singular matrix $S_b$ to retain its null space1 which can be used to approximate the null space of $S_W^{-1}S_b$ by obtaining the range space of $S_W^{-1}S_b$. The eigenvectors of $S_W^{-1}S_b$ can be denoted by $E = [E_b, E_r]$ (where $E_b \in \mathbb{R}^{C-1 \times r_b}$ and $E_b \in \mathbb{R}^{r_r \times (r_r - r_b)}$).

Since the rank of $S_b$ is $r_b < r_w$, only leading $r_b$ eigenvectors (i.e., eigenvectors corresponding to largest eigenvalues) of $E_b$ can be considered to form an orientation matrix. The remaining $r_w - r_b$ eigenvalues could be noisy which would give erroneous corresponding weighted eigenvectors. If the leading eigenvectors of $E_b$ is denoted by $E_b \in \mathbb{R}^{C-1 \times r_b}$ then it can be considered as approximated null space of $S_W^{-1}S_b$. Since the range space of $S_W^{-1}S_b$ is $E_b$ and its approximated null space is $E_{b, b}$, the orientation matrix in $r_b$-dimensional space would be $W = [E_b, E_r] \in \mathbb{R}^{C-1 \times 2r_b}$ or in $d$-dimensional space would be $W = U_{d \times d} \in \mathbb{R}^{d \times 2r_b}$. Theoretically, this strategy would include all the four spaces and retrieve the null space information of $S_W^{-1}S_b$. The summary of the algorithm is depicted in Table 1.

4. Computational considerations

The computational complexity of the two-stage LDA technique is higher than other techniques like null space based technique

---

1 Regularization of $S_b$ can be done in a similar manner as we have done regularization of $S_W$ matrix. An example of this can be viewed from Fig. 2 by replacing the matrix $S_W$ by matrix $S_b$ and by replacing the rank $r_w$ by rank $r_b$. 
(OLDA) (Ye, 2005), PCA plus LDA (Swets and Weng, 1996; Belhumeur et al., 1997) and DLDA (Yu and Yang, 2001) as eigenvector computation is required both in the null space as well as in the range space of scatter matrices. However, by applying PCA as a pre-processing step the computational complexity would be reduced by removing the null space of total scatter matrix and then transforming feature vectors on the \( r \)-dimensional space. The computational complexity of the pre-processing step (Step 1) would be \( O(dn^2) \), where \( d \) is the dimensionality of feature vectors and \( n \) is the number of training feature vectors. The computational complexity of eigenvalue decomposition of the scatter matrices in Step 2 would be \( O(n^3) \). Some additional computational complexity will be required to estimate eigenvalues in the null space of scatter matrices in the two-stage LDA technique. Now we use the regularization technique (as above). The results are demonstrated in Table 4. It can be observed from this table that the null space of \( S_b \) provides additional and complementary information over the other three spaces (range space of \( S_B \), null space of \( S_W \) and range space of \( S_b \)). We report here the results for the two-stage LDA technique with and without the null space of \( S_b \). For the two-stage LDA technique with the null space of \( S_b \), the procedure is same as described in Table 1. For using the two-stage LDA technique without the null space of \( S_b \), we modified the procedure given in Table 1 as follows. Instead of using the range space of \( S_b^{-1}S_W \), we use the range space of \( S_W \) in Step 4. Multiple runs of N-fold cross-validation are carried out on all the four datasets and the resulting average recognition accuracies with the null space of \( S_b \) and without the null space of \( S_b \) are depicted in Table 3. It can be observed from this table that the null space of \( S_b \) does provide complementary information over the other three spaces and plays a useful role in the proposed two-stage LDA technique.

So far we have provided results where the two-stage LDA technique is used to reduce the dimensionality to \( 2r_B \). Now we show its performance as a function of dimensionality. To demonstrate this, we varied the dimensions from 5 to \( 2r_B \) (where, \( 2r_B = 2(c − 1) \) and \( c \) is the number of classes) and computed the average recognition accuracy by doing multiple runs of N-fold cross-validation technique (as above). The results are demonstrated in Table 4. It can be seen from this table that the recognition performance improves by increasing the dimensionality.

In the experiments described above, we have used the extrapolation procedure for obtaining the non-singular estimates of scatter matrices in the two-stage LDA technique. Now we use the regularization method to obtain the non-singular estimate of these matrices. In order to do this, we vary the regularization parameter \( \alpha \) in the following manner. For estimating within-class scatter matrix in full space we define regularization parameter \( \alpha = \delta * /S_W \) where \( \lambda_{S_W} \) is the maximum eigenvalue of within-class scatter matrix and \( \delta \) is a small positive number. Similarly, for estimating between-class scatter matrix in full space we define \( \alpha = \delta * /S_b \), where \( \lambda_{S_b} \) is the maximum eigenvalue of between-class scatter matrix. The average recognition accuracy is then obtained by conducting multiple runs of N-fold cross-validation on the four face recognition datasets. The results are shown in Table 5. We can see from this table that the recognition performance can be improved by choosing the regularization parameter appropriately. However, it must be

### Table 1
The algorithm.

| Step 1 | Pre-processing stage: apply PCA to find range space \( U_R \in \mathbb{R}^{d×n} \) of total scatter matrix \( S_T \) and apply it to find transformed within-class scatter matrix \( S_W \in \mathbb{R}^{d×n} \) and between-class scatter matrix \( S_B \in \mathbb{R}^{r_B×n} \) (where \( r_B \) is the rank of \( S_b \)) |
| Step 2 | Estimate non-singular matrices \( S_W \) and \( S_B \) from singular matrices \( S_W \) and \( S_B \), respectively, by using either regularization technique or extrapolation technique |
| Step 3 | Decompose \( S_W^{-1}S_B \) into its eigenvalues and eigenvectors, and find the leading \( r_B \) number of eigenvectors \( W_i \in \mathbb{R}^{d×r_B} \) (i.e., eigenvectors corresponding to largest eigenvalues), where \( r_B \) is the rank of between-class scatter matrix |
| Step 4 | Similarly (as Step 3) find leading \( r_B \) number of eigenvectors \( W_i \in \mathbb{R}^{d×r_B} \) from the eigenvalue decompose of \( S_B^{-1} S_W \) |
| Step 5 | Form \( W = [W_1, W_2] \) and compute orientation matrix \( W = U_B W \in \mathbb{R}^{d×2r_B} \) |

### Table 2
Performance of the techniques in terms of average recognition accuracy over multiple runs of N-fold cross-validation on ORL, AR, Yale and FERET databases.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>ORL (%)</th>
<th>AR (%)</th>
<th>Yale (%)</th>
<th>FERET (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLDA (Yu and Yang, 2001)</td>
<td>89.5</td>
<td>80.8</td>
<td>93.5</td>
<td>92.9</td>
</tr>
<tr>
<td>Null space based technique (OLDA) (Ye, 2005)</td>
<td>91.5</td>
<td>80.8</td>
<td>97.3</td>
<td>97.1</td>
</tr>
<tr>
<td>PCA plus LDA (Swets and Weng, 1996; Belhumeur et al., 1997)</td>
<td>86.0</td>
<td>83.4</td>
<td>98.0</td>
<td>95.7</td>
</tr>
<tr>
<td>Regularized LDA (Zhao et al., 2003)</td>
<td>91.5</td>
<td>75.4</td>
<td>97.9</td>
<td>97.3</td>
</tr>
<tr>
<td>Regularized LDA based on DLDA framework (Lu et al., 2005)</td>
<td>89.8</td>
<td>81.6</td>
<td>94.7</td>
<td>94.5</td>
</tr>
<tr>
<td>MLDA (Thomaz et al., 2005)</td>
<td>92.0</td>
<td>76.2</td>
<td>97.9</td>
<td>97.8</td>
</tr>
<tr>
<td>Eigenfeature regularization technique (Jiang et al., 2008)</td>
<td>92.3</td>
<td>81.8</td>
<td>98.6</td>
<td>97.7</td>
</tr>
<tr>
<td>Two-stage LDA technique</td>
<td>92.6</td>
<td>87.8</td>
<td>98.7</td>
<td>98.0</td>
</tr>
</tbody>
</table>
noticed (by comparing Tables 4 and 5) that the performance of the two-stage LDA technique with extrapolation procedure is in general better than that with the regularization procedure.

6. Conclusion

We have proposed a two-stage LDA technique that includes both the null space and range space information of between-class scatter and within-class scatter matrices. The regularization is done in parallel to give two orientation matrices. These orientation matrices are concatenated to form the final orientation matrix. The proposed technique is shown to provide better classification performance on several face recognition datasets than the other techniques.

Appendix I

In this appendix we describe (pragmatically) that all the four spaces namely, null space of $S_W (S_{null}^W)$, range space of $S_W (S^{range}_W)$, null space of $S_P (S_{null}^P)$ and range space of $S_P (S^{range}_P)$ contain significant discriminant information. In order to demonstrate this, first we project the original feature vectors onto the range space of total scatter matrix as a pre-processing step. Then all the spaces are utilized individually to do dimensionality reduction and to classify a test feature vector, the nearest neighbor classifier is used. For this experiment the datasets have been approximately equally divided into training samples and test sample. Table A1 depicts the classification accuracy. It can be observed from the table that individual spaces ($S_{null}^W, S^{range}_W$ and $S_{null}^P, S^{range}_P$) contain significant discriminant information. Though the $S_{null}^P$ is less effective, it still contains some information.

Appendix II

In this appendix we will show that by doing eigenvalue decomposition of $S_{null}^W, S_{null}^P$ could result in noisy eigenvectors (where the full rank scatter matrices $S_W$ and $S_P$ are the estimates of singular scatter matrices $S_W$ and $S_P$, respectively). From Eqs. (6) and (9) of the text, $Q = S_W^{-1} S_P$ can be expressed as

$$Q = \left[ U_W, U_{WN} \right] \begin{bmatrix} \Sigma_W^{-1} \gamma_{W} & 0 \\ 0 & \Sigma_P^{-1} \gamma_{P} \end{bmatrix} \left[ U_W^T U_{WN} \right] \begin{bmatrix} U_W^T & U_{BN} \end{bmatrix} \begin{bmatrix} \gamma_{W} & 0 \\ 0 & \gamma_{P} \end{bmatrix} \begin{bmatrix} U_W^T \\ U_{BN} \end{bmatrix}$$

where $P$ is the remaining sum of products. If the diagonal entries of $\Sigma_W$ is $\lambda_j > 0$ (for $j = 1,...,(r - r_c)$) and $\Sigma_P$ is $\gamma_j > 0$ (for $j = 1,...,(r - r_c)$), and the corresponding column vectors of $U_W$ is $u_j$ and $U_{BN}$ is $v_j$ then

$$Q = \left( \sum_{j=1}^{r-r_c} u_j u_j^T \right)^{1/2} \left( \sum_{j=1}^{r-r_c} v_j v_j^T \right)^{1/2}$$

Since $\lambda_j$ and $\gamma_j$ are lagging eigenvalues of $S_W$ and $S_P$, respectively, the eigenvalues will be small and noisy. It is reasonable to assume that the values of $\lambda_j$ (for all $j$) are closely equal and similarly the values of $\gamma_j$ (for all $k$) are closely equal; i.e., $\lambda_1 \approx \lambda_2 \approx \cdots \approx \lambda_{r-r_c}$ and $\gamma_1 \approx \gamma_2 \approx \cdots \approx \gamma_{r-r_c}$. Therefore, Eq. (A1) can be written as

$$Q \approx P = \left( \frac{\gamma_j}{\lambda_j} \left( \sum_{j=1}^{r-r_c} u_j u_j^T \right)^{1/2} \left( \sum_{j=1}^{r-r_c} v_j v_j^T \right)^{1/2} \right)$$

where $\gamma_j$ consists of true eigenvalue $\gamma_j$ and additive noise $\sigma_k$, where $|\sigma_k| < b$ and $b$ is a positive constant. Similarly, let $\lambda_j = \lambda_j + \sigma_k$, where $|\sigma_k| < W$ and $W$ is a positive constant. Let $\epsilon$ denotes the ratio $\gamma_j^2/\lambda_j^2$. The ratio $\epsilon$ could be in the range $0 < \epsilon < 1$ and if noise $\sigma$ and $\sigma_k$ are dominant factors then this will lead to serious erroneous value of $Q$ and the orientation matrix.

References


Belhumeur, P.N. (http://cvat.yale.edu/projects/yalefaces/yalefaces.html).


