A New Quadratic Classifier Applied to Biometric Recognition

Carlos E. Thomaz[1], Duncan F. Gillies[1] and Raul Q. Feitosa[2,3]

[1]Imperial College of Science Technology and Medicine, Department of Computing,
180 Queen's Gate, London SW7 2BZ, United Kingdom
{cet,dfg}@doc.ic.ac.uk
[2]State University of Rio de Janeiro, Department of Computer Engineering,
r. São Francisco Xavier, Rio de Janeiro 20559-900, Brazil
[3]Catholic University of Rio de Janeiro, Department of Electrical Engineering,
r. Marques de Sao Vicente 225, Rio de Janeiro 22453-900, Brazil
raul@ele.puc-rio.br

Abstract. In biometric recognition applications, the number of training examples per class is limited and consequently the conventional quadratic classifier either performs poorly or cannot be calculated. Other non-conventional quadratic classifiers have been used in limited sample and high dimensional classification problems. In this paper, a new quadratic classifier called Maximum Entropy Covariance Selection (MECS) is presented. This classifier combines the sample group covariance matrices and the pooled covariance matrix under the principle of maximum entropy. This approach is a direct method that not only deals with the singularity and instability of the maximum likelihood covariance estimator, but also does not require an optimisation procedure. In order to evaluate the MECS effectiveness, experiments on face and fingerprint recognition were carried out and compared with other similar classifiers, including the Reguralized Discriminant Analysis (RDA), the Leave-One-Out Covariance estimator (LOOC) and the Simplified Quadratic Discriminant Function (SQDF). In both applications, using the publicly released databases FERET and NIST-4, the MECS classifier achieved the lowest classification error.

1 Introduction

In most image recognition applications, especially in biometric ones, the number of training examples per class is limited. In such situations, the conventional maximum likelihood quadratic classifier either performs poorly or cannot be calculated when the group sample sizes are smaller than the number of features.

Other non-conventional quadratic classifiers have been used in limited sample and high dimensional classification problems [1,3,7,9]. All these quadratic approaches rely on optimisation techniques that are time consuming and do not necessarily lead to the highest classification accuracy for all circumstances.

In this paper, a new quadratic classifier called Maximum Entropy Covariance Selection (MECS) is presented. This classifier is based on combining covariance matrices under the principle of maximum entropy. It assumes that the sources of variation are similar from group to group and consequently a similar covariance shape may be

M. Tistarelli, J. Bigun, A.K. Jain (Eds.): Biometric Authentication, LNCS 2359, pp. 186-196, 2002.

expected for all classes. This has often been the case for biometric applications such as face recognition. The new classifier not only deals with the singularity and instability of the maximum likelihood covariance estimator, but also is computed directly not requiring an optimisation procedure.

In order to evaluate the MECS effectiveness compared with other classifiers, experiments on face and fingerprint applications were carried out using the corresponding publicly released databases FERET and NIST-4. In both applications the MECS classifier achieved the lowest classification error.

2 The Quadratic Discriminant Classifier

The Quadratic Discriminant classifier is based on the *p*-multivariate normal class-conditional probability densities, where *p* is the dimension of the feature vector. Assuming the symmetrical or zero-one loss function, the optimal quadratic discriminant (QD) rule stipulates that an unknown pattern *x* should be assigned to the class or group *i* that *minimises*:

$$d_i(x) = \ln|\Sigma_i| + (x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - 2\ln\pi_i\,, \tag{1}$$

where π_i is a prior probability associated with the *i*th group, and the parameters μ_i and Σ_i represent the true mean vector and covariance matrix respectively for the *i*th group.

In practical situations, the true values of the mean and covariance matrix in (1) are replaced by their respective maximum likelihood estimates $\bar{x}_i$ and S_i. Thus, the QD rule can be rewritten as: assign pattern *x* to class *i* that *minimises*:

$$d_i(x) = \ln|S_i| + (x-\bar{x}_i)^T S_i^{-1}(x-\bar{x}_i) - 2\ln\pi_i\,. \tag{2}$$

This is the standard or conventional quadratic discriminant function (QDF) classifier.

2.1 Sample Size Effects

As a general guideline, Jain and Chandrasekaran [6] have suggested that the class sample sizes n_i should be at least five to ten times the dimension of the feature space *p*. Indeed when n_i are small compared with *p*, the sample group covariance estimates S_i become highly variable or even not invertible when $n_i < p+1$.

The effect of that instability on the QD classifier can be explicitly seen by rewriting the quadratic discriminant distance described in (2) on its spectral decomposition form [2], as follows:

$$d_i(x) = \sum_{k=1}^{p} \ln\lambda_{ik} + \sum_{k=1}^{p} \frac{[\phi_{ik}^T(x-\bar{x}_i)]^2}{\lambda_{ik}} - 2\ln\pi_i\,, \tag{3}$$

where λ_{ik} is the kth eigenvalue of S_i and ϕ_{ik} is the corresponding eigenvector. As can be observed, a poor or unreliable estimation of the sample group covariance matrices tends to exaggerate the importance associated with the low-variance information and consequently distorts the quadratic discriminant analysis.

One method routinely applied is the so-called linear discriminant function (LDF). The LDF classifier is obtained by replacing the S_i in (2) with the pooled sample covariance matrix

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \cdots + (n_g - 1)S_g}{N - g}, \tag{4}$$

where n_i is the number of training observations from class i, g is the number of groups and $N = n_1 + n_2 + \cdots + n_g$. Theoretically, however, S_p is a consistent estimator of the true covariance matrices Σ_i only when $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g$.

3 Other Quadratic Classifiers

Other non-conventional quadratic classifiers have been used in limited sample and high dimensional classification problems. In the next sub-sections, three of the most important of these quadratic classifiers are briefly described. The new quadratic classifier proposed in this work is detailed in section 4.

3.1 Reguralized Discriminant Analysis (RDA) Classifier

The Friedman's RDA [1] approach is a two-dimensional optimisation method that shrinks both the S_i towards S_p and also the eigenvalues of the S_i towards equality by blending the first shrinkage with multiples of the identity matrix.

In this context, the sample covariance matrices S_i of the QD rule defined in (2) are replaced by the following $S_i^{rda}(\lambda,\gamma)$

$$\begin{aligned} S_i^{rda}(\lambda,\gamma) &= (1-\gamma)S_i^{rda}(\lambda) + \gamma\left(\frac{tr(S_i^{rda}(\lambda))}{p}\right)I, \\ S_i^{rda}(\lambda) &= \frac{(1-\lambda)(n_i - 1)S_i + \lambda(N-g)S_p}{(1-\lambda)n_i + \lambda N} \end{aligned}, \tag{5}$$

where the notation "tr" denotes the trace of a matrix. The mixing parameters λ and γ are restricted to the range 0 to 1 (optimisation grid) and are selected to maximise the leave-one-out classification accuracy regarding all groups [1]. Although RDA has the benefit of being directly related to the classification accuracy, it is the most computationally intensive method particularly when a large number of groups is considered.

3.2 Leave-One-Out Covariance (LOOC)Classifier

Hoffbeck and Landgrebe [3] proposed a covariance estimator that depends only on covariance optimisation of single classes.

The idea is to examine pair-wise mixtures of the sample group covariance estimates S_i and the unweighted common covariance estimate S, together with their diagonal forms [3]. The LOOC estimator has the following form:

$$S_i^{looc}(\alpha_i) = \begin{cases} (1-\alpha_i)\mathrm{diag}(S_i) + \alpha_i S_i & 0 \le \alpha_i \le 1 \\ (2-\alpha_i)S_i + (\alpha_i - 1)S & 1 < \alpha_i \le 2 \\ (3-\alpha_i)S + (\alpha_i - 2)\mathrm{diag}(S) & 2 < \alpha_i \le 3 \end{cases} . \quad \textbf{(6)}$$

The optimisation strategy is to evaluate several values of α_i over the grid $0 \le \alpha_i \le 3$, and then choose α_i that maximizes the average log likelihood of the corresponding p-variate normal density function [3]. Considering valid approximations of the covariance estimates [3] LOOC requires less computation than RDA estimator.

3.3 Simplified Quadratic Discriminant Function (SQDF) Classifier

The SQDF classifier has been proposed by Omachi et al. [7] and can be viewed as an approximation method of the standard quadratic discriminant function.

Basically, the SQDF classifier approximates the spectral decomposition form of the QD classifier described in (3) by the following function:

$$d_i(x) = \sum_{k=1}^{s} \ln \lambda_{ik} + \sum_{k=s+1}^{p} \ln \lambda + \sum_{k=1}^{s} \frac{[\phi_{ik}^T (x - \bar{x}_i)]^2}{\lambda_{ik}} + \sum_{k=s+1}^{p} \frac{[\phi_{ik}^T (x - \bar{x}_i)]^2}{\lambda} - 2\ln \pi_i , \quad \textbf{(7)}$$

where $(\lambda_{ik}, \phi_{ik})$ are the k-th eigenvalue-eigenvector pair of S_i, λ is a simplification constant and $s \le p$. The constant λ is determined by the mean value of $\lambda_{ik}\,(k = s+1,\ldots,p)$ and the parameter s can be defined arbitrarily, experimentally or by information criterion [7]. Since SQDF approximates the QD classification rule considering solely the information provided by each sample group covariance matrix, it seems to be more sensible to poor sample covariance estimation than other non-conventional QD classifiers.

4 A New Quadratic Classifier

In biometric recognition applications, the pattern classification task is commonly performed on pre-processed or well-framed images and the sources of variation are often the same from group to group. As a consequence, a similar covariance shape may be assumed for all groups. In such situations and when the sample group covariance matrices S_i are singular or not accurately estimated, linear combinations of S_i and the pooled covariance matrix S_p may lead to a “loss of covariance information”

[10]. This statement forms the basis of the new quadratic classifier based on the Maximum Entropy Covariance Selection method proposed.

4.1 The "Loss of Covariance Information"

The theoretical interpretation of the "loss of covariance information" can be described as follows. Let a matrix S_i^{mix} be given by the following linear combination:

$$S_i^{mix} = aS_i + bS_p , \tag{8}$$

where the mixing parameters a and b are positive constants, and the pooled covariance matrix S_p is a non-singular matrix. The S_i^{mix} eigenvectors and eigenvalues are given by the matrices Φ_i^{mix} and Λ_i^{mix}, respectively. From the covariance spectral decomposition formula [2], it is possible to write

$$(\Phi_i^{mix})^T S_i^{mix} \Phi_i^{mix} = \Lambda_i^{mix} = \begin{bmatrix} \lambda_1^{mix} & & & 0 \\ & \lambda_2^{mix} & & \\ & & \ddots & \\ 0 & & & \lambda_p^{mix} \end{bmatrix} = diag[\lambda_1^{mix}, \lambda_2^{mix}, ..., \lambda_p^{mix}] , \tag{9}$$

where $\lambda_1^{mix}, \lambda_2^{mix}, ..., \lambda_p^{mix}$ are the S_i^{mix} eigenvalues and p is the dimension of the measurement space considered. Using the information provided by equation (8), equation (9) can be rewritten as:

$$\begin{aligned} (\Phi_i^{mix})^T S_i^{mix} \Phi_i^{mix} &= diag[\lambda_1^{mix}, \lambda_2^{mix}, ..., \lambda_p^{mix}] \\ &= (\Phi_i^{mix})^T \left[aS_i + bS_p \right] \Phi_i^{mix} \\ &= a(\Phi_i^{mix})^T S_i \Phi_i^{mix} + b(\Phi_i^{mix})^T S_p \Phi_i^{mix} \\ &= a\Lambda^{i*} + b\Lambda^{p*} \\ &= diag[a\lambda_1^{i*} + b\lambda_1^{p*}, a\lambda_2^{i*} + b\lambda_2^{p*}, ..., a\lambda_p^{i*} + b\lambda_p^{p*}] \end{aligned} \tag{10}$$

where $\lambda_1^{i*}, \lambda_2^{i*}, ..., \lambda_p^{i*}$ and $\lambda_1^{p*}, \lambda_2^{p*}, ..., \lambda_p^{p*}$ are the corresponding variances of the sample and pooled covariance matrices spanned by the S_i^{mix} eigenvectors matrix Φ_i^{mix}. Then, the spectral decomposition form of the QD score described in (3) becomes:

$$d_i(x) = \sum_{k=1}^{p} \ln\left(a\lambda_k^{i*} + b\lambda_k^{p*}\right) + \sum_{k=1}^{p} \frac{[(\phi_{ik}^{mix})^T (x - \bar{x}_i)]^2}{a\lambda_k^{i*} + b\lambda_k^{p*}} - 2\ln\pi_i , \tag{11}$$

where ϕ_{ik}^{mix} is the corresponding k-th eigenvector of the matrix S_i^{mix}.

The discriminant score described in equation (11) considers the dispersions of sample group covariance matrices spanned by all the S_i^{mix} eigenvectors. However, when the group sample sizes n_i are small or not large enough compared with the dimension of the feature space p, the corresponding lower dispersion values are often estimated to be 0 or approximately 0, implying that these values are not reliable. Therefore, a linear combination of S_i and S_p that uses the same parameters a and b

as defined in (11) for the whole feature space fritters away some pooled covariance information.

The geometric idea of a hypothetical "loss of covariance information" on a three-dimensional feature space is illustrated in Figure 1. The constant probability density contour of S_i and S_p are represented by the two-dimensional (x_1, x_2) dark grey ellipse and three-dimensional (x_1, x_2, x_3) light grey ellipsoid, respectively.

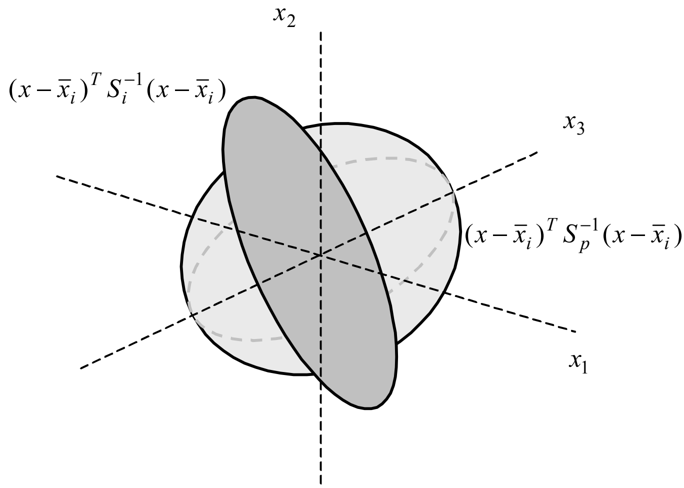

Figure 1. Geometric idea of a hypothetical "loss of covariance information".

As can be seen, S_i is well defined on the plane (x_1, x_2) but not defined at all on (x_1, x_2, x_3). In fact, there is no information from S_i on the x_3 axis. As a consequence, a linear combination of S_i and S_p that shrinks or expands both matrices equally all over the feature space simply ignores this evidence. Other covariance estimators have not addressed this problem.

4.2 Maximum Entropy Covariance Selection Method

The Maximum Entropy Covariance Selection (MECS) method considers the issue of combining the sample group covariance matrices and the pooled covariance matrix based on the maximum entropy (ME) principle [5].

Let a p-dimensional sample X_i be normally distributed with true mean μ_i and true covariance matrix Σ_i, i.e. $X_i \sim N_p(\mu_i, \Sigma_i)$. The entropy h of such multivariate distribution can be written as:

$$h(X_i) = \frac{p}{2} + \frac{1}{2}\ln|\Sigma_i| + \frac{p}{2}\ln 2\pi \, , \qquad \textbf{(12)}$$

which is simply a function of the determinant of Σ_i and is invariant under any orthonormal transformation [2]. Thus, when Φ_i consists of p eigenvectors of Σ_i

$$\ln\left|\Phi_i^T \Sigma_i \Phi_i\right| = \ln\left|\Lambda_i\right| = \sum_{k=1}^{p} \ln \lambda_k \ . \tag{13}$$

In order to maximise (13) or equivalently (12), we must select the covariance estimation of Σ_i that gives the largest eigenvalues.

Considering convex combinations between the sample group covariance S_i and S_p matrices, equation (13) can be rewritten (by using equation (10)) as

$$\ln\left|(\Phi_i^{mix})^T \left(aS_i + bS_p\right)\Phi_i^{mix}\right| = \sum_{k=1}^{p} \ln(a\lambda_k^{i*} + b\lambda_k^{p*}) , \tag{14}$$

where $\lambda_1^{i*}, \lambda_2^{i*}, ..., \lambda_p^{i*}$ and $\lambda_1^{p*}, \lambda_2^{p*}, ..., \lambda_p^{p*}$ are the corresponding variances of the sample and pooled covariance matrices spanned by Φ_i^{mix}, and the parameters a and b are nonnegative and sum to 1. Moreover, as the natural logarithm is a monotonic increasing function, we do not change the problem if instead of maximising equation (14) we maximise

$$\sum_{k=1}^{p} (a\lambda_k^{i*} + b\lambda_k^{p*}) . \tag{15}$$

However $a\lambda_k^{i*} + b\lambda_k^{p*}$ is a convex combination of two real numbers and the following inequality is valid [4]

$$a\lambda_k^{i*} + b\lambda_k^{p*} \le \max(\lambda_k^{i*}, \lambda_k^{p*}) \tag{16}$$

for any $1 \le k \le p$. Therefore, in order to maximise equation (15) and consequently the entropy given by the convex combination of S_i and S_p, we do not need to choose the best parameters a and b but simply select the maximum variances of the corresponding matrices.

Thus, the MECS estimator S_i^{mecs} is given by the following procedure:

i. Find the eigenvectors Φ_i^{me} of the covariance given by $S_i + S_p$.

ii. Calculate the variance contribution of both S_i and S_p on the Φ_i^{me} basis, i.e.

$$\begin{aligned} \Lambda_i^* &= diag[(\Phi_i^{me})^T S_i \Phi_i^{me}] = [\lambda_1^{i*}, \lambda_2^{i*}, ..., \lambda_p^{i*}] \\ \Lambda_p^* &= diag[(\Phi_i^{me})^T S_p \Phi_i^{me}] = [\lambda_1^{p*}, \lambda_2^{p*}, ..., \lambda_p^{p*}] \end{aligned} \tag{17}$$

iii. Form a new variance matrix based on the largest values, that is

$$\Lambda_i^{me} = diag[\max(\lambda_1^{i*}, \lambda_1^{p*}), \max(\lambda_2^{i*}, \lambda_2^{p*}), ..., \max(\lambda_p^{i*}, \lambda_p^{p*})] . \tag{18}$$

iv. Form the MECS estimator

$$S_i^{mecs} = \Phi_i^{me} \Lambda_i^{me} (\Phi_i^{me})^T . \tag{19}$$

The MECS quadratic classifier is constructed by substituting S_i^{mecs} for S_i on the QD rule defined in (2). This approach is a direct procedure that not only deals with the singularity and instability of S_i but also with the loss of information when similar covariance matrices are linearly combined. Furthermore, it does not require an optimisation procedure and consequently its computation cost is less severe than the RDA, LOOC and SQDF methods.

5 Experiments and Results

In order to investigate the performance of MECS compared with QDF, LDF, RDA, LOOC and SQDF classifiers, two biometric applications were considered: face and fingerprint recognition. In the face classification application, the training sample sizes were chosen to be extremely small compared to the dimensionality of the feature space. In contrast, large training sample sizes were considered for the fingerprint recognition. Both applications were analysed using publicly released databases.

5.1 Experiments

In the face recognition experiments the FERET Face Database [8] was used. Sets containing 4 "frontal b series" images for each of 200 total subjects were considered. Each image set is composed of a regular facial expression (referred as "ba" images in the FERET database), an alternative expression ("bj" images), and two symmetric images ("be" and "bf" images) taken with the intention of investigating 15 degrees pose angle effects. For implementation convenience all images were first resized to 96x64 pixels and transformed into eigenfeature vectors [11]. Each experiment was repeated 25 times using several of those eigenfeatures. Distinct training and test samples were randomly drawn, and the mean of the recognition rate was calculated. Since the LOOC computation requires at least three examples in each class [3], the recognition rate was computed utilising for each subject 3 images to train and 1 image to test.

The fingerprint classification was performed utilising the training and test feature vectors extracted from the grey scale images of the standard NIST-4 Special Database [12]. Each feature vector consists of 112 floating point numbers, made by a feature selection procedure that ends with the Karhunen-Loeve transform. The fingerprints were classified into one of five categories (L=left loop, W=whorl, R=right loop, T=tented arch, and A=arch) with an equal number of prints from each class (400). There are 2000 first-rolling fingerprint feature vectors for training and 2000 corresponding second-rolling ones for testing.

In both applications, the prior probabilities were assumed equal for all groups. Also the RDA optimisation grid was taken to be the outer product of $\lambda = [0, 0.125, 0.354, 0.650, 1.0]$ and $\gamma = [0, 0.25, 0.5, 0.75, 1.0]$, identically to the Friedman's work [1]. Analogously, the size of the LOOC mixture parameter was $\alpha_i = [0, 0.25, 0.5, \ldots, 2.75, 3.0]$. The SQDF parameter s, however, was defined differently in each application. Due to the extremely small training sample, s was 1 for the FERET classification – in this case, $rank(S_i) = 2$. In the fingerprint recognition, this parameter was selected experimentally as one of the following values: 5, 10, 20, 30 and 40.

5.2 Results

Figures 2 and 3 present the test average recognition error of the FERET face and NIST-4 fingerprint databases, respectively, over different number of features.

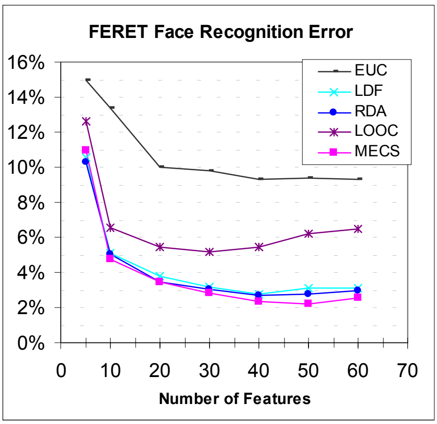

Figure 2. Feret Recognition Error

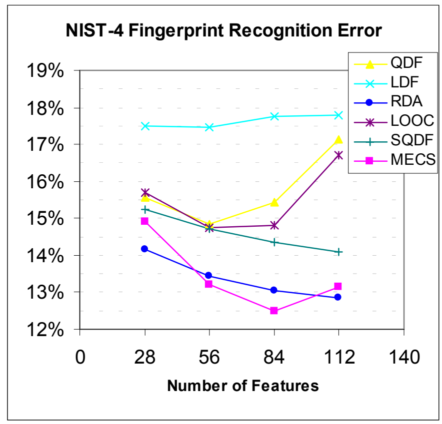

Figure 3. NIST-4 Recognition Error

Since only 3 face images were used to train the classifiers, the sample group covariance matrices S_i were singular and the QDF could not be calculated. Due to the same problem the SQDF classifier performed poorly (around 30% of classification error for all the features considered) and its results are not presented. Instead, the recognition rate of the Euclidean distance classifier (EUC) that corresponds to the classical Eigenfaces method proposed by Turk and Pentland [11] are displayed. Figure 2 shows that for all the feature components considered the MECS quadratic classifier performed as well or better than the other classifiers. The MECS quadratic classifier achieved the lowest classification error – 2.2% – on 50 eigenfeatures. In this application where S_i seem to be quite similar, favouring the LDF performance, the MECS classifier did better without loss of covariance information.

The recognition results of the NIST-4 fingerprint database are presented in figure 3. In the lowest and highest dimension spaces (28 and 112 features), RDA led to lower classification error than MECS estimator. However, for 56 and 84 features the MECS performed better than the other classifiers. Although in this application the ratio of the training sample size to the number of features is large, favouring the QDF, RDA, LOOC and SQDF classifiers, the MECS estimator achieved the lowest classification error – 12.5% – on 84 components. Putting this result in perspective, a classification error of 12% but with 10% rejection was reported on the same training and test sets [12].

6 Conclusion

In this work, a new quadratic classifier called Maximum Entropy Covariance Selection (MECS) was introduced. This classifier is based on combining the sample group covariance matrices and the pooled covariance matrix under the principle of maximum entropy. This approach is a direct procedure that not only deals with the singularity and instability of the maximum likelihood covariance estimators, but also with the loss of information when similar covariance matrices are convexly combined. Furthermore, it is not a time consuming method because it does not require an optimisation procedure.

The effectiveness of the MECS method compared with several classifiers (QDF, LDF, RDA, LOOC and SQDF) was evaluated on two biometric applications: face and fingerprint recognition. In the face classification application, the training sample sizes were chosen to be extremely small compared to the dimensionality of the feature space. In contrast, large training sample sizes were considered for the fingerprint recognition. In both applications, using the publicly released databases FERET and NIST-4, the MECS quadratic classifier achieved the lowest classification error.

These results indicate that the MECS classifier does increase the classification accuracy in biometric recognition applications where the sources of variation are commonly the same from group to group and limited training sample sizes are considered. In such situations and when concerns about the computation cost exists, the MECS should be preferable to the other aforementioned classifiers.

Acknowledgment

The first author was partially supported by the Brazilian Government Agency CAPES under grant No. 1168/99-1. Also, portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

1. J.H. Friedman, "Reguralized Discriminant Analysis", Journal of the American Statistical Association, vol. 84, no. 405, pp. 165-175, March 1989.
2. K. Fukunaga, Introduction to Statistical Pattern Recognition, second edition. Boston: Academic Press, 1990.
3. J.P. Hoffbeck and D.A. Landgrebe, "Covariance Matrix Estimation and Classification With Limited Training Data", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 7, pp. 763-767, July 1996.
4. R. A. Horn and C.R. Johnson, Matrix Analysis. Cambridge University Press, 1985.
5. E.T. Jaynes, "On the rationale of maximum-entropy methods", Proceedings of the IEEE, , vol. 70, pp. 939-952, 1982.
6. A.K. Jain and B. Chandrasekaran, "Dimensionality and Sample Size Considerations in Pattern Recognition Practice", Handbook of Statistics, P.R. Krishnaiah and L.N. Kanal Eds, vol. 2, pp. 835-855, North Holland, 1982.

7. S. Omachi, F. Sun, and H. Aso, "A New Approximation Method of the Quadratic Discriminant Function", SSPR&SPR 2000, Springer-Verlag LNCS 1876, pp. 601-610, 2000.
8. P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", Image and Vision Computing Journal, vol. 16, no. 5, pp. 295-306, 1998.
9. S. Tadjudin and D.A. Landgrebe, "Covariance Estimation With Limited Training Samples", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 4, July 1999.
10. C. E. Thomaz, D. F. Gillies and R. Q. Feitosa, "Small Sample Problem in Bayes Plug-in Classifier for Image Recognition", in proceedings of Int'l Conference on Image and Vision Computing New Zealand, pp. 295-300, Dunedin, New Zealand, November 2001.
11. M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, pp. 72-85, 1991.
12. C. L. Wilson, G. T. Candela, P. J. Grother, C. I. Watson, and R. A. Wilkinson, "Massively Parallel Neural Network Fingerprint Classification System", Technical Report NIST IR 4880, National Institute of Standards and Technology, July 1992.