Regularized Discriminant Analysis for Face Recognition

Itzik Pima , Mayer Aladjem
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev
P.O.Box 653, Beer-Sheva, 84105, Israel.

Abstract
This paper studies Regularized Discriminant Analysis (RDA) in the context of face recognition. We check RDA sensitivity to different photometric preprocessing methods and compare its performance to other classifiers. Our study shows that RDA is better able to extract the relevant discriminatory information from training data than the other classifiers tested, thus obtaining a lower error rate. Moreover, RDA is robust under various lighting conditions while the other classifiers perform badly when no photometric method is applied.

Keywords: face recognition, feature extraction, regularization, principal component analysis, discriminant analysis, photometric preprocessing.

1 Introduction

This study compares the performance of Regularized Discriminant Analysis [2] (RDA) with that of two classifiers: L2 (Euclidean distance) and angle (Normalized Correlation), usually used for face recognition. In saying L2 and angle we mean that we use a nearest center classifier using those distance metrics. The potential of extracting the relevant discriminatory information from a small amount of training data using RDA motivated us to explore RDA in the context of face recognition. We would like to state here that this is the first application of RDA [2] to face recognition.

In order to study the efficacy of the RDA for face recognition we have designed experiments in which a small number of faces are represented in both Principal Component Analysis (PCA) features using Eigenfaces [1] and Linear Discriminant Analysis (LDA) features using Fisherfaces [4]. Our goal was to study RDA's behavior with two completely different data types: one that is obtained by PCA, which is simply data compression, and the other, obtained by LDA, that yields highly separated data. Moreover we also studied the effect of image photometric preprocessing methods on the performances of the classifiers.

The paper is organized as follows. In the next section we explain the application of the RDA to face classification. Section 3 introduces the face database used for this work and describes the experiments carried out and their objectives. Finally, in Section 4, conclusions are drawn from the results obtained.

2. RDA Face Classification

Assume that we have training images sliced into column vectors $\mathbf{z}_i^{(j)}$ $(1 \le i \le n_j)$ for j=1,2,…,g. Each $\mathbf{z}_i^{(j)}$ belongs to one of the g classes $\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_g$ where $\mathbf{z}_i^{(j)}$ is an image taken from class $\mathbf{C}_j$. Here, n_j is the number of images from class $\mathbf{C}_j$. The dimension of $\mathbf{z}_i^{(j)}$ is $P = N \times M$ (the number of pixels in the face image). In this work we use the linear mapping $\mathbf{x}_i^{(j)} = \mathbf{V}^T \mathbf{z}_i^{(j)}$, where $\mathbf{V}$ is a $P \times S$ transformation matrix for $S \le P$ (usually $S << P$). The S- dimensional vector $\mathbf{x}_i^{(j)}$ is named “feature vector”. We compute $\mathbf{V}$ using PCA (Eigenfaces [1]) and LDA (Fisherfaces [4]). Then we use PCA and LDA features as input of the RDA classifier.

RDA [2] is a modification of the Quadratic Discriminant Analysis (QDA) [5]. QDA assigns an arbitrary face represented by feature vector $\mathbf{x}$ to class k if $D_k(\mathbf{x}) > D_m(\mathbf{x})$, $1 \le m, k \le g$, $k \ne m$, where

$$D_k(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) + \ln|\boldsymbol{\Sigma}_k| - 2\ln(P(k)).$$

Here $P(k)$, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are ML estimators [5] of the a-priori probability, mean and the covariance matrix of the feature vectors from class k. A problem with the QDA classifier occurs when the class sample sizes n_k are small compared with the dimension of the feature space S. In this case, the covariance matrices estimates become highly variable. In order to solve this problem, in RDA [2] the class-conditional covariance matrices $\boldsymbol{\Sigma}_k$ in $D_k(\mathbf{x})$ are replaced by a regularized estimate

$\mathbf{\Sigma}_{\mathrm{k}}(\lambda,\gamma)$. Following Friedman [2], we first compute the pooled (within-class) sample covariance matrix $\mathbf{\Sigma} = \sum_{\mathrm{k}=1}^{g} P(k)\,\mathbf{\Sigma}_{\mathrm{k}}$. Then using a regularization parameter λ we get

$$\mathbf{\Sigma}_{\mathrm{k}}(\lambda) = \frac{(1-\lambda)P(k)\Sigma_{\mathrm{k}} + \lambda\Sigma}{(1-\lambda)P(k) + \lambda} \quad , \quad 0 \le \lambda \le 1 .$$

Finally, using another regularization parameter γ, we have

$$\mathbf{\Sigma}_{\mathrm{k}}(\lambda,\gamma) = (1-\gamma)\mathbf{\Sigma}_{\mathrm{k}}(\lambda) + \frac{\gamma}{\mathrm{S}}\, trace[\mathbf{\Sigma}_{\mathrm{k}}(\lambda)]\,\mathbf{I} \quad \mathbf{,} \quad 0 \le \gamma \le 1 .$$

The parameter λ converts the class covariance matrix $\mathbf{\Sigma}_{\mathrm{k}}(\lambda)$ to a linear combination of $\mathbf{\Sigma}_{\mathrm{k}}$ and $\mathbf{\Sigma}$. The second parameter, γ, shrinks $\mathbf{\Sigma}_{\mathrm{k}}(\lambda)$ toward a multiple of the identity matrix. The suitable values of λ and γ are determined by the model selection procedure [2]. This procedure sets a 2-dimensional grid of points on the λ and γ plane ($0 \le \lambda \le 1,\ 0 \le \gamma \le 1$), evaluates the cross-validated estimate of misclassification risk at each prescribed point on the grid, and then chooses the point with the smallest estimated risk as our suitable values of the regularization parameters λ, γ. In our experiments we set the values λ, γ = 0.0001, 0.25, 0.50, 0.75, 1.00, and applied leave-one-out cross-validation procedure [5]. Note that for $\lambda = 1$ and $\gamma = 0.0001$ we get $\mathbf{\Sigma}_{\mathrm{k}}(\lambda,\gamma) \approx \mathbf{\Sigma}$ and we carry out linear discriminant classification [5]. For $\lambda = 1$ and $\gamma = 1$, RDA corresponds to the L2 classifier. Holding $\gamma = 0$ and varying λ produces classifiers between QDA and linear discriminant classification.

3. Experimental Study

All of our experiments are based on the *Olivetti Research Laboratory (ORL)* face database (retrieved from "ftp://ftp.uk.research.att.com:pub/data/att_faces.tar.Z"). Since we checked the performances of some preprocessing photometric methods we had changed the lighting in the database randomly.

The *ORL* database structure contains 10 different images of 40 distinct subjects (persons). As most researchers did, we used 5 images from every class for training and 5 images per test and a size of 48×48 pixels for each image.

We performed a number of experiments employing different photometric normalization, features (PCA and LDA), decision rules (RDA, L2 and angle). Following [3] we focused on the photometric methods based on image normalization and histogram equalization. The number of LDA features runs from 3 to 39 with steps of 3 (the last step is 2) where 39 is the maximal number available for the LDA. The number of PCA features was 10 to 199 with steps of 10 (the last step is 9), where 199 is the maximal number available for the PCA.

For every preprocessing photometric methods, and for every different feature dimension (for PCA and LDA) we ran RDA, L2 and angle classifiers. In Figs 1 and 2 we show the test error rates obtained.

4. Discussion and Conclusions

It is clear from looking at the results that RDA outperforms the L2 and angle classifiers when using the PCA feature extraction method (Fig. 1), but this phenomenon is not so obvious when using the LDA feature extraction method (Fig. 2). The best classification results are attained when using RDA with PCA (Fig. 1a) for histogram equalization when the features' dimension is between 30 and 50 (error rates are 10.5% to 11.5%). This is a remarkable feature of RDA because reports in the literature usually state that LDA is better than PCA since LDA extracts the relevant information while PCA only compresses it.

Looking at Fig. 1, presenting the results for PCA, we found that the RDA classifier does not need any preprocessing to achieve good results. This can save precious time when demanding real time applications are used.

An interesting point is the non-monotonic behavior of the RDA errors using PCA features for image normalization in the dimensions 60 – 120 and for histogram equalization in the dimensions 50 - 110. The reason for this is that the model selection procedure selected the values $\lambda = 0.0001$ and $\gamma = 0.0001$ as its suitable parameters. These values cause RDA to act as QDA (see end of Section 2 and [2]), which increases the risk of over-fitting, resulting in a large test error rate. For the higher dimension the parameters λ, γ are set to be other than 0.0001, RDA produces classifiers between QDA, LDA and L2, thus reducing the risk of over-fitting and decreasing the test error rate.

Finally the major features of RDA are its ability to extract relevant discriminatory information and its robustness to lighting changes. Support Vector machine (SVM) shares the same features [3]. It is interesting to compare RDA and SVM, which is an object of our future research.

References

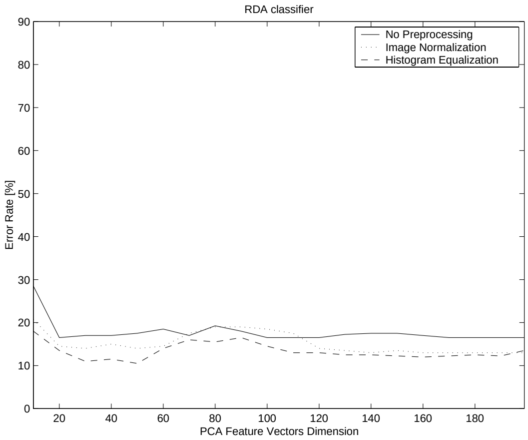
1.M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience, 3 (1), 1991, pp. 72-86.

2.J. H. Friedman, Regularized Discriminant Analysis, Journal of the American Statistical Association, 84 (405), 1989, pp. 165-175.

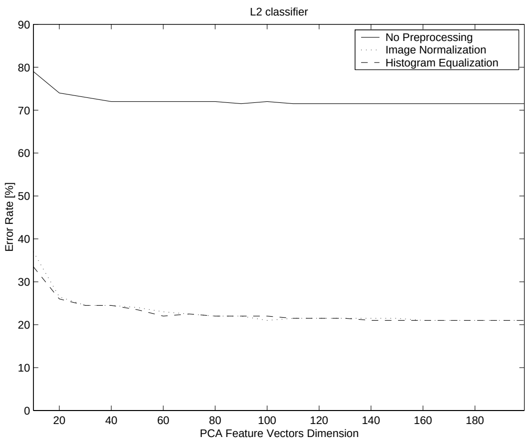
3.K. Jonsson, J. Kittler,YPLi, J. Matas, Support Vector Machines for Face Authentication, Image and Vision Computing, 20 (5-6), 2002, pp. 369-375.

4.K. Etemad, R. Chellappa, Discriminant Analysis for Recognition of Human Face Images, Journal of Optical Society of America A, 1997, pp. 1724-1733.

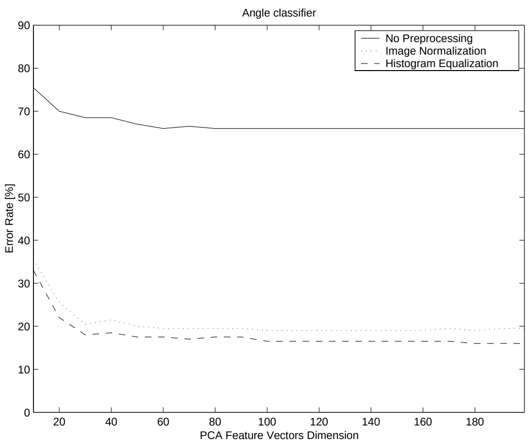
5. R.O. Duda, P.E. Hart, D.J. Stork, Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 2001.



(a)

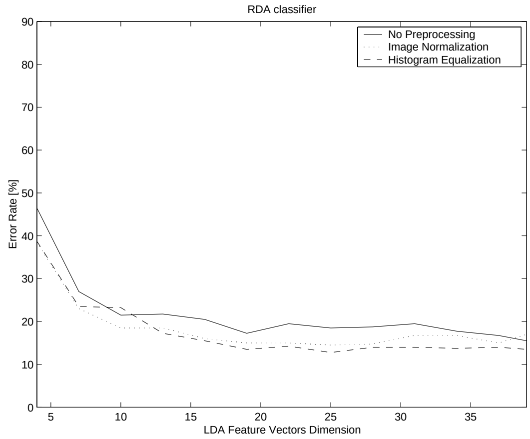


(b)

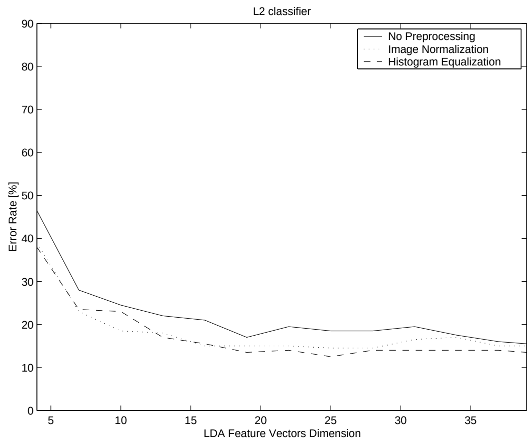


(c)

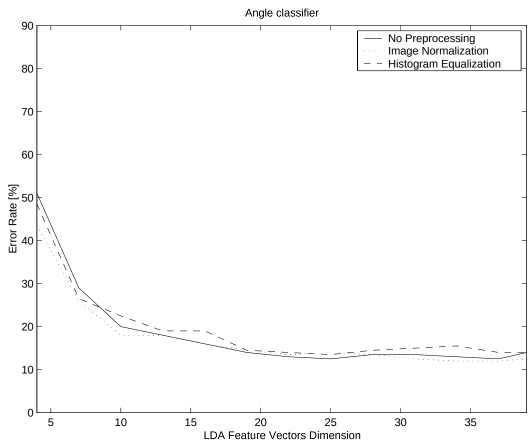
Figure 1. Classifiers' error rates with different preprocessing types versus different PCA features dimensions: (a) RDA classifier, (b) L2 classifier, (c) Angle classifier.



(a)



(b)



(c)

Figure 2. Classifiers' error rates with different preprocessing types versus different LDA features dimensions: (a) RDA classifier, (b) L2 classifier, (c) Angle classifier.