RBPCA MaxLike: A Novel Statistic Classifier for Face Recognition based on Block-Based PCA and Covariance Matrix Regularization

Denis H. P. Salvadeo, Nelson D. A. Mascarenhas, Jander Moreira
Computer Department
Federal University of São Carlos, UFSCar
São Carlos, Brazil
{denissalvadeo, nelson, jander}@dc.ufscar.br

Alexandre L. M. Levada, Débora C. Corrêa
Physics Institute of São Carlos
University of São Paulo, USP
São Carlos, Brazil
{alexandreluis, deboracorrea}@ursa.ifsc.usp.br

Abstract—This paper presents a novel statistical approach for face recognition, called *RBPCA MaxLike*, by combining covariance matrix regularization, block-based PCA (BPCA) for dimensionality reduction and maximum likelihood for classification. Typically, face recognition is an extremely ill-posed problem, since the data dimensionality is huge in comparison to the number of available samples (small sample size problem). The proposed method performs robust covariance matrices estimation by adopting the *Mixed-LOOC2* regularization technique, allowing a more accurate statistical modeling for parametric classification (Maximum Likelihood and Bayesian classifiers), since it avoids the presence of singular matrices. Experiments with two well known face databases (ORL and AR) show good results.

Keywords-Face Recognition; Block-based PCA; Covariance Matrix Regularization; Maximum Likelihood; PCA.

I. INTRODUCTION

In the last years, the human face has been widely explored in a variety of applications such as biometric systems, security and the development of smart digital cameras. Definitely, one of the most challenging tasks that are common to all these applications is the problem of face recognition.

It is known that a face recognition system can operate in two distinct modes: verification (authentication) or identification (recognition), which is the main focus of this work. Recognition involves one-to-many matches that compare a query face against all the template images in the database [1]. Face recognition methods can be divided in the following categorization: Holistic matching methods, Feature-based matching methods and Hybrid methods [1]. There are many technical challenges in face recognition and they are common to all existing approaches. According to [2], the key ones are: large variability, highly complex nonlinear manifolds, high dimensionality and small sample size, which is the main concern of this paper.

Holistic methods use the whole face as input. Statistical techniques based on Principal Component Analysis (Eigenfaces) [3], and other feature extraction methods as Linear Discriminant Analysis (Fisherfaces) [4] or Independent Component Analysis (ICA) [5], are used in this class of methods. PCA is a classical second order method that expands the observed feature vectors into the m eigenvectors associated to the m largest eigenvalues of the covariance matrix, creating orthogonal uncorrelated features [6]. LDA is a supervised mapping that founds basis vectors that maximize the data separability, while minimizing the intraclass scattering [6].

The proposed *RBPCA MaxLike* classifier is also an example of holistic method, since it uses features that are linear combination of non-overlapping blocks from the entire face. Feature-based methods are based on the extraction of local features such as statistics and geometrical measures. Finally, hybrid methods combine both local features and the global face representation during the classification stage.

In this work, our objective is to propose a novel classifier for face recognition, aiming the statistical modeling of human face images. Our motivation is the possibility of performing supervised parametric classification through Maximum Likelihood (ML) and Bayesian classifiers (*Maximum a Posteriori*) in face recognition. Comparisons among two other widely used supervised classifiers, PCA+NN (nearest neighbor) and PCA/LDA+NN, are also presented. To the best of our knowledge, it is the first time *Mixed-LOOC2* [7] is applied for covariance matrix regularization in face recognition.

The remaining of the paper is organized as follows: Section II describes the proposed BPCA for dimensionality reduction. Section III presents the *Mixed-LOOC2* criterion for covariance matrices regularization. Section IV shows the experiments and results and finally Section V presents the conclusions and final remarks.

II. BLOCK-BASED PCA

Feature extraction methods apply geometric transformations to the feature space in order to generate new features based on combinations of the original ones, aiming the dimensionality reduction. The motivation for the feature extraction stage on classification problems can be summarized on three main topics: the completely non-intuitive geometry of the hyperspaces (makes it hard to use similarity and distance measures), the curse of dimensionality, and last but not least, the computational cost. Also, it can be shown that the use of a low dimensional feature space is capable of improving the generalization of the classifier. Face recognition problems face an even worst scenario since the dimensionality is extremely high and the use of feature extraction methods becomes mandatory.

Within this context, we propose a PCA variant known as *Block-based PCA* or BPCA. Basically, the idea of BPCA is to divide each face image in several $k \times k$ blocks. Typically, all the blocks have same size (4×4, 8×8, etc.). Then, PCA is performed as if each block were a sample, that is, each $k \times k$ block is projected on m eigenvectors associated to the m largest eigenvalues. The detail here is that each $k \times k$ block is reduced to a smaller subset of pixels, decreasing the spatial dimensions of each face image. Thus, dimensionality reduction does not use all the information from different faces (classes).

The BPCA features are not considered as being linear combination of different faces, creating a completely different subspace than those obtained by PCA and LDA. Instead, they represent the mixture of small pieces from the same face. Fig. 1 shows the block diagram of the proposed BPCA method. First, the image is subdivided in r $k \times k$ blocks. By using the lexicographic notation it is possible to calculate both the mean 'block' vector and the covariance matrix for each face. In this work, we transform the image by reducing each $k \times k$ block to a single pixel, producing a final image representation in terms of r features (the same number of blocks).

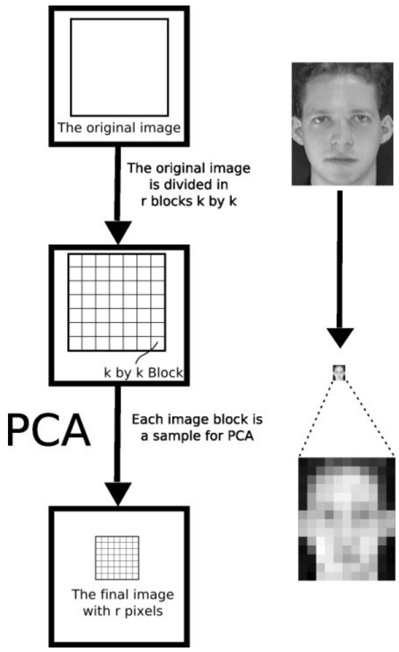

Figure 1. Illustrative diagram of the proposed Block-Based PCA.

III. MIXED-LOOC2 REGULARIZATION

The *Mixed-LOOC2* regularization technique was proposed by [7] to avoid singularity of covariance matrices in high dimensional data classification. Basically, it has two main principles. First, when dealing with small sample size problems, the common (or mixture) covariance matrix can replace the class covariance matrix, since although the later is also singular, it produces reduced estimation errors. Second, by inserting a small perturbation in the main diagonal of the covariance matrix, it can avoid singularity. The *Mixed-LOOC2* criterion proposes the following expression for regularizing the i-th class covariance matrix:

$$\hat{\Sigma}_i(\alpha_i) = \alpha_i A + (1-\alpha_i)B \qquad (1)$$

where $A = \frac{tr(S_i)}{p} I$, $diag(S_i)$, S_i, $\frac{tr(S)}{p} I$, $diag(S)$ or S, $B = S_i$ or $diag(S)$, p is the number of dimensions, tr is the matrix trace, $diag$ is an operator that returns a matrix with the main diagonal of the argument and replaces every element outside the diagonal with zeros, S is the common (pooled) covariance matrix, S_i is the i-th intraclass covariance matrix and α_i is a parameter that has to be estimated for each class, by maximizing the Leave-One-Out Log-Likelihood function (LOOL), defined as:

$$LOOL_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \ln\left[f\left(\vec{x}_k | \vec{m}_{i/k}, \hat{\Sigma}_{i/k}(\alpha_i)\right)\right] \qquad (2)$$

where N_i is the number of samples from i-th class, $\vec{m}_{i/k}$ is i-th class mean vector calculated without the sample k, $\hat{\Sigma}_{i/k}$ is the i-th class covariance matrix estimated by removing the sample k (Leave-One-Out).

Thus, the basic idea is to estimate the values of α_i, A and B for each class by maximizing the LOOL criterion to obtain a non-singular covariance matrix that better fits the data. Note that the final covariance matrix considers all the samples of a given class. With this kind of regularization we improve maximum likelihood and Bayesian face recognition under the Gaussian hypothesis. In this work, we combined BPCA and the *Mixed-LOOC2* regularization technique to define RBPCA, a regularized version of BPCA.

IV. METHODOLOGY

To test and evaluate our method, we propose a series of experiments using two well-known face databases: the ORL and AR databases [8].

A. *The AR face database*

This database consists of over 3200 768 x 576 color images of the frontal images of faces of 126 subjects (70 men and 56 women). There are 26 different images for each subject. For each subject, these images were recorded in two different sessions separated by two weeks, each session consisting of 13 images [9]. For the experiments we used the warped version of AR database, which is actually a subset of the original one, that is, the number of classes is reduced (100 subjects, being 50 men and 50 women). Each warped face image has dimensions of 165 x 120 pixels, corresponding to a 19800-D feature vector. The images contain variations in facial expressions, illumination, occlusions and age. For illustration, some of these images for one subject are shown in Fig. 2.

Figure 2. AR database face images from an individual: there are a total of 13 different variations.

B. *The ORL face database*

The ORL face database was collect between 1992 and 1994 and it has a total of 400 92 x 112 gray level facial images from 40 individuals (4 women and 36 men). There are 10 samples per class with variations in face expression and illumination [10]. Fig. 3 shows some examples of facial images from this database.

Figure 3. ORL database face images from an individual.

C. *Experiments and Results*

During the experiments we compared the classification performance of the proposed classifier (*RBPCA MaxLike*) against two widely used methods in face recognition: PCA + Nearest Neighbor and PCA/LDA + Nearest Neighbor classification. The classification error was estimated by using the holdout 50-50 rule.

For the AR dataset, PCA was first applied to the 19800-D space data to generate 48, 70 and 88 dimensional PCA subspaces. After that, LDA was used to generate 48, 70 and 88 dimensional LDA-subspaces from the 99-D PCA subspace. Then, classification was performed by a NN classifier. For the proposed method, we had the following transforms: to extract 48 features we first resized the face images to 160 x 120 dimensions. Then, BPCA was applied on 20 x 20 blocks, creating a 8 x 6 representation of each image. Similarly, for 70 and 88 features, 16 x 16 blocks and 15 x 15 blocks were used, resulting in 10 x 7 and 11 x 8 representations, respectively. After that, Maximum Likelihood classification is applied to the resulting feature vectors. Fig. 4 shows the obtained results. Note that the proposed method outperformed the PCA/NN and LDA/NN in all situations. Also, the obtained results show that by using 48 features with the proposed method leads to a performance similar to the case of 88 LDA features, showing the discriminative power of *RBPCA MaxLike* for face recognition problems.

For the ORL database, we first used both PCA and LDA to generate 39-D subspaces from the 10304-D data (since we are dealing with 40 classes, the maximum number of LDA features is 39). For RBPCA, the images were first resized to 112 x 88. Right after that, RBPCA was applied with 8 x 8 blocks, reducing the images dimensions to 14 x 11, that is, extracting 154 features. The holdout method was also applied to estimate the classification error, but considering different partitions in the training dataset (from 3 to 9). Fig. 5 compares the results of PCA/NN, LDA/NN and the proposed *RBPCA MaxLike* classifier. Again, the results clearly show the superiority of *RBPCA MaxLike* over traditional methods.

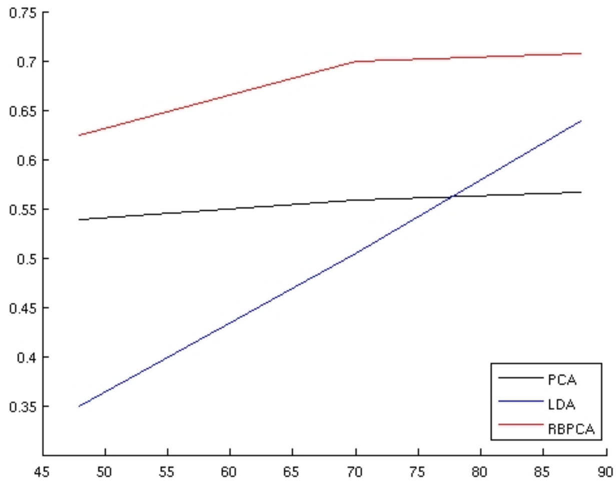

Figure 4. Comparison between Correct classification rates for PCA/LDA+NN and the proposed *RBPCA MaxLike* classifier in AR database face images.

As the AR database has 13 variations per individual, from pose and illumination conditions to occlusion, its is possible to perform a deeper analysis in terms of each one of these variations. Each individual has face images labeled from 1 to 13 according to the following: neutral expression, smile, anger, scream, left light on, right light on, both sides light on, wearing sunglasses, wearing sun glasses and left light on, wearing sun glasses and right light on, wearing scarf, wearing scarf and left light on and wearing scarf and right light on. Figs. 6, 7 and 8 show the classification errors for each possible variation using PCA/LDA+NN and the proposed *RBPCA MaxLike* method. Note that the proposed method provides a uniformly better performance, leading to smaller classification errors.

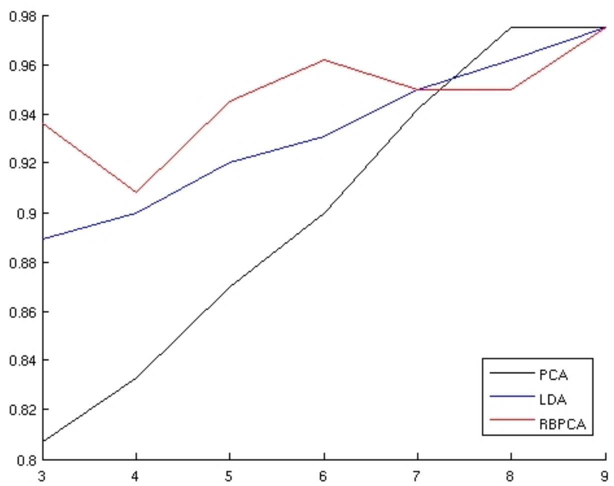

Figure 5. Comparison between correct classification rates for PCA/LDA+NN and the proposed *RBPCA MaxLike* classifier in ORL

database face images, varying the number of training and testing samples.

Besides improving the classification performance, the proposed method drastically reduced the computational cost. Regarding the AR database, while PCA and LDA took approximately 19 hours of processing, BPCA spent only 17 minutes. All tests were run in a Intel Core2Quad Q6600 2.40GHz machine with 8GB of DDR3 RAM. The implementation was done in MATLAB, using the PRTOOLS v4.1 toolbox [11].

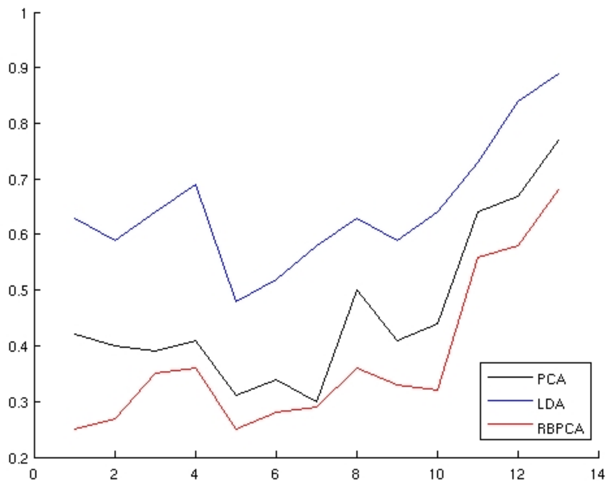

Figure 6. Classification errors for each possible variation using PCA/LDA+NN and the proposed *RBPCA MaxLike* method using 48-D feature vectors.

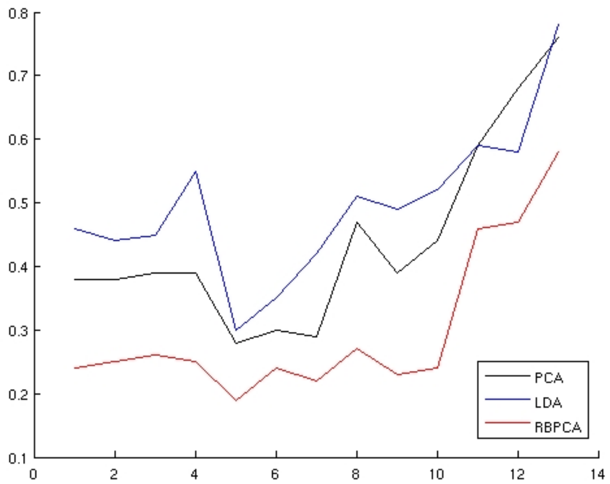

Figure 7. Classification errors for each possible variation using PCA/LDA+NN and the proposed *RBPCA MaxLike* method using 70-D feature vectors.

V. CONCLUSIONS

This paper presented a novel parametric pattern classifier for face recognition, called *RBPCA MaxLike*, by combining covariance matrix regularization, block-based PCA (BPCA) for dimensionality reduction and maximum likelihood for classification. Experimental results showed that the proposed method improves classification performance. Two main advantages of the proposed *RBPCA MaxLike* are: it is able to extract more features than LDA (since it is not restricted to the number of classes); and its features showed to be more discriminant for face recognition problems. Besides, covariance matrix regularization allows parametric classification through ML and Bayesian approaches. Future works include the use of classifier combination techniques and the incorporation of wavelets during feature extraction stage aiming for a further improvement on classification performance, reducing the computational cost by adopting more compact representations.

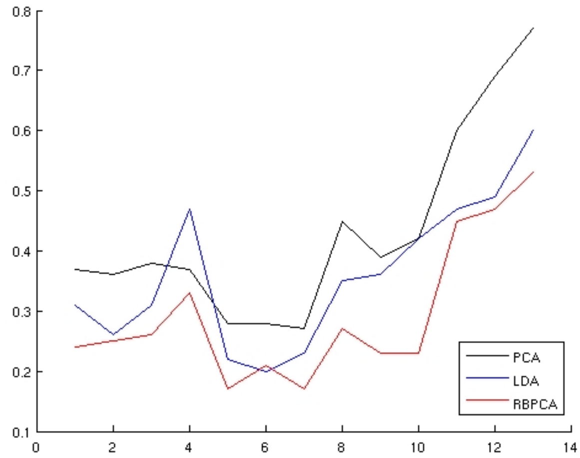

Figure 8. Classification errors for each possible variation using PCA/LDA+NN and the proposed *RBPCA MaxLike* method using 88-D feature vectors.

ACKNOWLEDGMENT

We would like to thank CNPq for Denis H. P. Salvadeo student scholarship and FAPESP for Alexandre L. M. Levada (grant n. 06/01711-4) and Debora C Correa (grant n. 2009/50142-0) student scholarships.

REFERENCES

[1] W. Zhao, R. Chellapa, P. J. Phillips and A. Rosenfeld (2003), 'Face Recognition: A Literature Survey', ACM Computing Surveys 35, 399-459.

[2] A. K. Jain and S. Z. Li (2005), 'Handbook of Face Recognition', Springer-Verlag.

[3] M. A. Turk and A. P. Pentland (1991), 'Eigenfaces for Recognition', Journal of Cognitive Neuroscience 3(1), 71-86.

[4] K. Etemad and R. Chellapa (1997), 'Discriminant Analysis for the Recognition of Human Face Images', Journal of the Opt. Society of America A 14, 1727-1733.

[5] B. A. Draper, K. Baek, M. S. Bartlett and J. R. Beveridge (2003), 'Recognizing Faces with PCA and ICA', Computer Vision and Image Understanding 91(1), 115-137.

[6] A. Webb (2002), 'Statistical Pattern Recognition', Wiley.

[7] B. C. Kuo and D. A. Landgrebe (2002), 'A Covariance Estimator for Small Sample Size Classification Problems and its Application to Feature Extraction', IEEE Trans. on Geoscience and Remote Sensing 40(4), 814-819.

[8] A.M. Martínez and R. Benavente (1998), 'The AR-face database', CVC Technical Report 24.

[9] A. M. Martínez and A. C. Kak (2001), 'PCA versus LDA', IEEE Trans. on Pattern Analysis and Machine Intelligence 23(2), 228-233.

[10] Gross, R. (2005) Face Databases. In: LI, S. Z. e JAIN, A. K. (Eds.). 'Handbook of Face Recognition', New York: Springer-Verlag, 2005. Chapter 13, 301-328.

[11] F. Heijden, R. P.W. Duin, D. Ridder and D. M.J. Tax (2004), 'Classification, parameter estimation and state estimation – an engineering approach using Matlab', John Wiley & Sons.