

Face Recognition from Color Images Using Sparse Projection Analysis

Vitomir Štruc and Nikola Pavešić

Faculty of Electrical Engineering,
University of Ljubljana,
Tržaška 25, SI-1000 Ljubljana, Slovenia
{vitomir.struc,nikola.pavesic}@fe.uni-lj.si.com
<http://luks.fe.uni-lj.si/>

Abstract. The paper presents a novel feature extraction technique for face recognition which uses sparse projection axes to compute a low-dimensional representation of face images. The proposed technique derives the sparse axes by first recasting the problem of face recognition as a regression problem and then solving the new (under-determined) regression problem by computing the solution with minimum L_1 norm. The developed technique, named Sparse Projection Analysis (SPA), is applied to color as well as grey-scale images from the XM2VTS database and compared to popular subspace projection techniques (with sparse and dense projection axes) from the literature. The results of the experimental assessment show that the proposed technique ensures promising results on un-occluded as well occluded images from the XM2VTS database.

Key words: Image processing, biometrics, face recognition, regression problem, sparse projection axes.

1 Introduction

It is a well known fact that the existing face recognition techniques are sensitive to external factors influencing the appearance of the human face in an image. Among these factors, glasses, scarfs, hats or any other objects occluding the facial region have an immense effect on the representation of the face image in the given feature space and consequently on the performance of the face recognition technique used [1].

The reason for this sensitivity can be linked to the way the face images are usually transformed into the feature space. This transform commonly involves computing a projection of the facial image onto a low-dimensional subspace. If the image contains occlusions even on a very small part of the face, the feature (or subspace) representation may differ significantly from the true feature representation of the un-occluded face image. To derive a stable low-dimensional representation and hence to overcome the sensitivity of the existing methods to occlusions in images, robust techniques (e.g., [2]) and local approaches (e.g.,

[3]) have been proposed in the literature. By computing facial features in a robust manner or by considering only a small part of the image at a time (when calculating the low-dimensional subspace representation) these approaches are capable of improving upon the recognition performance of the holistic methods, especially when parts of images are occluded or degraded in some way.

Recent advances in sparse signal recovery and compressed sensing opened up new possibilities for the design of local techniques as well as new ways of tackling the problem of occlusion in face recognition [4]. Based on these developments we present in this paper a novel technique for facial feature extraction called Sparse Projection Analysis (SPA). SPA derives a low-dimensional face representation in the form of projection coefficients computed by projecting face images onto a set of sparse projection axes. Different from other local appearance based methods, such as, for example, independent component analysis (architecture I) [3], the non-zero elements of the projection axes are not localized but are rather distributed over the entire projection axis. Due to this property, the proposed technique should exhibit even less sensitivity to image occlusions than traditional local appearance based feature extraction techniques. It should be noted that the proposed techniques exhibits similarities with established face recognition methods such as independent component analysis [3] or non-negative matrix factorization [5] (due to the sparse nature of the projection axes), while the relation to the method presented in [4] is given only by the fact that SPA relies on the L_1 norm instead of the commonly used L_2 norm. Hence, the work presented in this paper is related more to the work presented in [3], [5] or [6], rather than [4].

The developed SPA technique is ultimately assessed in a series of face verification experiments performed on the original images of the XM2VTS database and on degraded images with an artificially added occlusion. The results of the assessment and comparative evaluations with the popular principal component analysis and independent component analysis show the effectiveness of the proposed approach for face recognition.

2 The Sparse Projection Analysis

This section commences by formulating the problem of face recognition as a regression problem and then, based on this formulation, develops the novel sparse projection analysis (SPA) for face recognition.

2.1 Face Recognition as a Regression Problem

In its most basic form the problem of face recognition can be defined as a mapping f from the given face pattern vector \mathbf{x} to a class label (or identity) ω_i associated with the i -th class C_i , where $i \in \{1, 2, \dots, N\}$ and N denotes the number of identities enrolled in the recognition system. Formally, this can be written as:

$$f : \mathbf{x} \mapsto \omega_i, \text{ for } i \in \{1, 2, \dots, N\}. \quad (1)$$

While the most natural way for humans is to associate identities with names, the class labels are in general not restricted to textual descriptions, but can rather take an arbitrary form which uniquely describes the class.

To illustrate this concept let us assume that we have a set of three face pattern vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ each belonging to a different class C_1, C_2, C_3 . The corresponding class labels could then be written as vectors of the following form: $\omega_1 = [1\ 0\ 0]$, $\omega_2 = [0\ 1\ 0]$ and $\omega_3 = [0\ 0\ 1]$. From this example we can see that the class labels can be chosen as an orthonormal basis of a vector space, whose dimensionality is defined by the number of classes that need to be labeled.

Let us now consider a more general case and presume a set of n d -dimensional face pattern vectors $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^d, i = 1, 2, \dots, n\}$ belonging to N classes C_1, C_2, \dots, C_N with associated class labels $\omega_1, \omega_2, \dots, \omega_N$. If we arrange the face pattern vectors in \mathcal{X} into the $n \times d$ row data matrix \mathbf{X} and use the same principle as illustrated in the above example to construct our labels, then we can build the $n \times N$ label matrix \mathbf{Y} , whose rows represent labels of the pattern vectors in \mathbf{X} . In this case the label matrix \mathbf{Y} takes the following form [7], [8], [9]:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{n_{C_1}} & \mathbf{0}_{n_{C_1}} & \cdots & \mathbf{0}_{n_{C_1}} \\ \mathbf{0}_{n_{C_2}} & \mathbf{1}_{n_{C_2}} & \cdots & \mathbf{0}_{n_{C_2}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_{C_N}} & \mathbf{0}_{n_{C_N}} & \cdots & \mathbf{1}_{n_{C_N}} \end{bmatrix}, \quad (2)$$

where n_{C_i} represents the number of face pattern vectors corresponding to class C_i , $\mathbf{1}_{n_{C_i}}$ denotes a $n_{C_i} \times 1$ vector of all ones and $\mathbf{0}_{n_{C_i}}$ stands for a $n_{C_i} \times 1$ vector of all zeros.

Since we have encoded the class labels in the form of a (label) data matrix \mathbf{Y} , we can use this matrix to define the mapping f . We implement f using a simple linear transformation of the pattern vectors in \mathbf{X} as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{W}, \quad (3)$$

where \mathbf{W} denotes the $d \times N$ transformation matrix.

Clearly, the above expression can be thought of as being a regression problem with the goal of finding the regression matrix \mathbf{W} capable of mapping the input variables in \mathbf{X} to the response variables in \mathbf{Y} .

2.2 Computing the Projection Basis

While there are several ways to determine the regression matrix \mathbf{W} (see, for example, [8]), we present in this paper a method which results in the N d -dimensional column vectors comprising the regression matrix \mathbf{W} being sparse.

Let us first have a closer look at our problem defined by Eq. (3). It is easy to notice that each of the columns in \mathbf{Y} represents the projection coefficients of all pattern vectors in the data matrix \mathbf{X} on one of the N column vectors comprising \mathbf{W} . Hence, the problem of determining the regression matrix \mathbf{W} can be broken down into a set of N independent sub-problems of the following form:

$$\mathbf{y}_i = \mathbf{X}\mathbf{w}_i, \text{ for } i = 1, 2, \dots, N, \quad (4)$$

where the $n \times 1$ vector \mathbf{y}_i denotes the i -th column of the response matrix \mathbf{Y} , and similarly the $d \times 1$ vector \mathbf{w}_i stands for the i -th column of \mathbf{W} . It has to be noted that the vector \mathbf{y}_i does not correspond to an encoded label, but is rather composed of the i -th elements of all encoded labels.

In the field of face recognition the dimensionality of the face pattern vectors in \mathbf{X} is usually much larger than the available number of training pattern vectors, i.e., $d \gg n$. The equation defined by (4) is, therefore, under-determined and its solution is not unique. We can, however, overcome this problem by selecting the solution with minimum norm [4]. While usually the L_2 norm is used for this purpose, recent research in the field of compressed sensing suggests that, when adopting the L_1 norm rather than the L_2 norm, the found solution exhibits a number of desirable properties, e.g., sparseness. To solve Eq. (4) for \mathbf{w}_i , i.e., to find the i -th projection axis, we therefore recast the problem as follows:

$$\mathbf{w}_i = \arg \min \|\mathbf{w}\|_{L_1}, \quad \text{subject to } \mathbf{y}_i = \mathbf{X}\mathbf{w}, \quad (5)$$

where \mathbf{w} denotes the (non-unique) solution of the above problem. As stated in [4] the above problem can be solved in polynomial time using standard linear programming techniques (see [10],[11]).

If we examine the projection matrix \mathbf{W} more closely, we can notice that the individual projection axes \mathbf{w}_i are not orthogonal and, hence, do not necessary form a basis of our N -dimensional vector space. As the final processing step we, therefore, orthogonalize the projection matrix and use the result as our final mapping from \mathbf{X} to \mathbf{Y} , i.e.:

$$\mathbf{W} = \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-0.5}. \quad (6)$$

Clearly, the orthogonalization procedure also influences the label data matrix \mathbf{Y} , which now turns into

$$\mathbf{Y}' = \mathbf{Y}(\mathbf{W}^T \mathbf{W})^{-0.5}. \quad (7)$$

This matrix, i.e., \mathbf{Y}' , is ultimately employed to construct client models, which are then stored in the system's database. Note that the client model (or template) for the i -th subject is computed as the mean vector of all encoded class labels (i.e., rows of \mathbf{Y}') corresponding to the subject labeled as ω_i .

When a new query image arrives at the input of the face recognition system it is simply projected into the sparse subspace using the orthogonalized version of \mathbf{W} .

As already indicated above, the exploited algorithm for finding the projection basis results in the computed projection axes being sparse. This fact is especially important for the feature extraction technique using this projection basis, since sparse axes insure that only a few pixels affect the value of each feature component. Such an approach should be robust to a number of image degradations including occlusion. A visual example of the sparseness of the projection basis is presented in Fig. 1, where a sample orthogonalized vector \mathbf{w}_i in image form (Fig. 1 left) and its 3D surface plot (Fig. 1 right) are depicted.

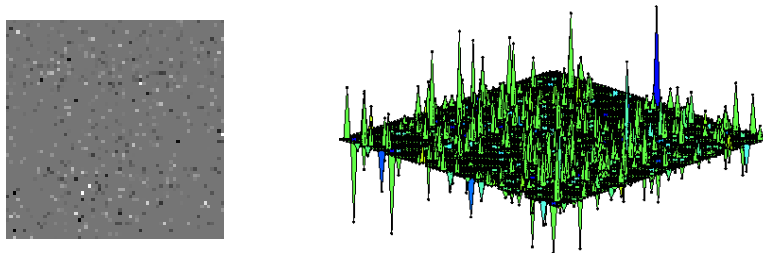


Fig. 1. A visual example of a sparse projection axis: in image form (left), as a surface (right).

3 Experiments

The feasibility of the SPA technique was empirically assessed using the XM2VTS database, which has long been the standard face-image database for evaluating novel authentication algorithms [12]. The database comprises 2360 images of 295 distinct subjects of different gender, age and race. The images were recorded in four separate sessions over a period of five months with the recording setup featuring controlled conditions, i.e., uniform background, controlled illumination, etc. Due to this recording setup, the variability in the images is induced mainly by the temporal factor. Thus, images of the same subjects differ in appearance due to changes in hairstyle, head-pose, presence or absence of make-up, glasses, etc.

For the experiments we followed the first configuration of the established experimental protocol associated with the XM2VTS database [12]. The protocol, known as the Lausanne protocol, partitions the subjects of the database into two disjoint groups of clients (200 subjects) and impostors (95 subjects) and further divides the images of these two groups into image sets used for: *(i) training and enrollment* - images which are used to train feature extractors and build client models/templates in form of mean feature vectors, *(ii) evaluation* - images which are used to determine the operating point, i.e., the decision threshold, of the face verification system and to define any potential parameters of the feature extractor (e.g., number of features, selection of features, etc.), and *(iii) testing* - images which are used to determine the verification error rates in real operating conditions.

While the first image set features only images belonging to the client group, the latter two image sets comprise images belonging to both the client and the impostor groups. These images are used in our assessment to determine the two error rates commonly exploited to quantify the performance of a given face verification system, namely, the false rejection and false acceptance error rates (FRR and FAR). These two error rates are defined as the relative frequency with which a face verification system falsely rejects a legitimate- and falsely accepts an impostor-identity-claim [7].

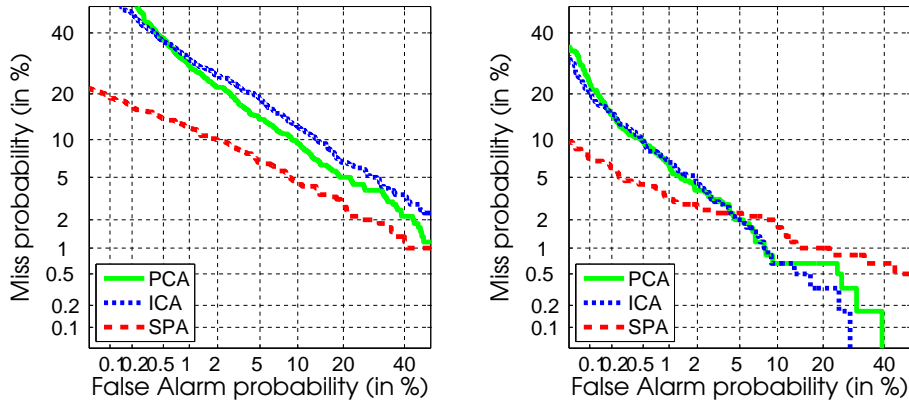


Fig. 2. DET curves generated during the experiments on the evaluation image set.

Unfortunately, both the FAR and the FRR depend on the value of the so-called decision threshold Δ and, hence, selecting a threshold which ensures a small value of FRR inevitably results in a high value of the FAR and vice versa, selecting a threshold which ensures a small value of the FAR results in a high value of the FRR. To compare the performance of two face verification systems an operating point has to be determined or the error rates should be plotted against a number of values of the decision threshold, thus generating a performance curve. Here we choose the latter approach and present our verification results as performance curves [7].

The employed protocol results in 600 client and 40000 impostor verification attempts in the evaluation stage to determine the decision thresholds needed to construct the performance curves, and 400 client and 112000 impostor verification attempts in the test stage to determine the final performance of the assessed techniques.

In all experiments we assume that the facial images are already localized and scaled to a standard size of 64×64 pixels. To compensate for any potential illumination induced appearance changes we further normalize the images by applying histogram equalization followed by a zero mean and unit variance normalization. The presented procedure is employed on all three color components of the YIQ color space into which the images of the XM2VTS database are transformed. It should be noted that the YIQ color space [13] rather than the RGB color space is used in the experiments to reduce the correlations between the individual color components and to make additional discriminatory information not present in the commonly adopted grey-scale images available to the feature extraction techniques. As a final step, we construct the SPA feature vector (of size $1 \times 3N$) of each face image by simply concatenating the feature vectors of the three color components. In addition to the experiments on the color images, we also provide baseline comparisons with experiments performed on grey-scale images.

The goal of our first series of verification experiments is to assess the performance of the proposed SPA feature extraction technique on the evaluation image sets and to compare the obtained error rates to that of some baseline feature extraction techniques. To this end, we implement the principal component analysis (PCA)[14] and independent component analysis (ICA - architecture I)[3] face verification techniques and test them on the XM2VTS database. For all assessed techniques we employ the nearest neighbor (to the mean) classifier in conjunction with the cosine similarity measure for the matching stage. The classifier is chosen for the comparison, as it represents a non-parametric classifier and hence does not introduce any classifier-dependant bias to the results.

From the DET curves in Fig. 2¹, which plot the false acceptance error rate against the false rejection error rate at various values of the decision threshold, we can find that the proposed technique significantly outperforms the PCA and ICA methods on the grey scale images (Fig. 2 (left)), and performs better at the lower values of the FAR on the color images (Fig. 2 (right)). The reason for the good performance of the SPA technique can be found in the way the projection axes are computed. As the label data matrix encodes the class membership of the training face pattern vectors and all training samples from each class are mapped to the same encoded label, the proposed technique compresses the intra-class scatter similar to the popular linear discriminant analysis technique.

It was argued in [15] that DET curves cannot be used to effectively compare two face recognition systems, since in real operating conditions an operational threshold has to be set and the performance with this threshold on unseen data might differ from the performance achieved when setting the threshold. To this end, the so-called expected performance curves (EPC) were introduced in [15]. To construct these curves two data sets of impostor and client images are needed. The first, the evaluation data set, is used to set a decision threshold which minimizes the following weighted error rate: $\alpha FAR(\Delta) + (1 - \alpha)FRR(\Delta)$, where the parameter α controls the relative importance of the two error rates FAR and FRR. This threshold is then used on the second, the test image set, to determine the value of the half total error rate (HTER) defined as $HTER = (FAR + FRR)/2$. When plotting the obtained HTER against a number of values of α , we obtain an example of the EPC.

In our second series of experiments we produce EPC curves for all three tested techniques. The generated curves are shown in the graphs labeled as Fig. 3(a) and Fig. 3(b) for the grey-scale and color images, respectively. We can see that as with the evaluation image set, the proposed technique performs the best, followed by a similar performance of the remaining PCA and ICA techniques. On the remaining two graphs of Fig. 3 (i.e., on the graphs labeled as Fig. 3(c) and Fig. 3(d)) we present the results of the assessment on grey-scale and color images, where the goal was to test the robustness of the two techniques to partial occlusion of the facial region. Here, the facial images were occluded by setting

¹ Note that the axis labels in Fig. 2 *Miss probability* and *False Alarm Probability* correspond to the false acceptance and false rejection error rates, respectively.

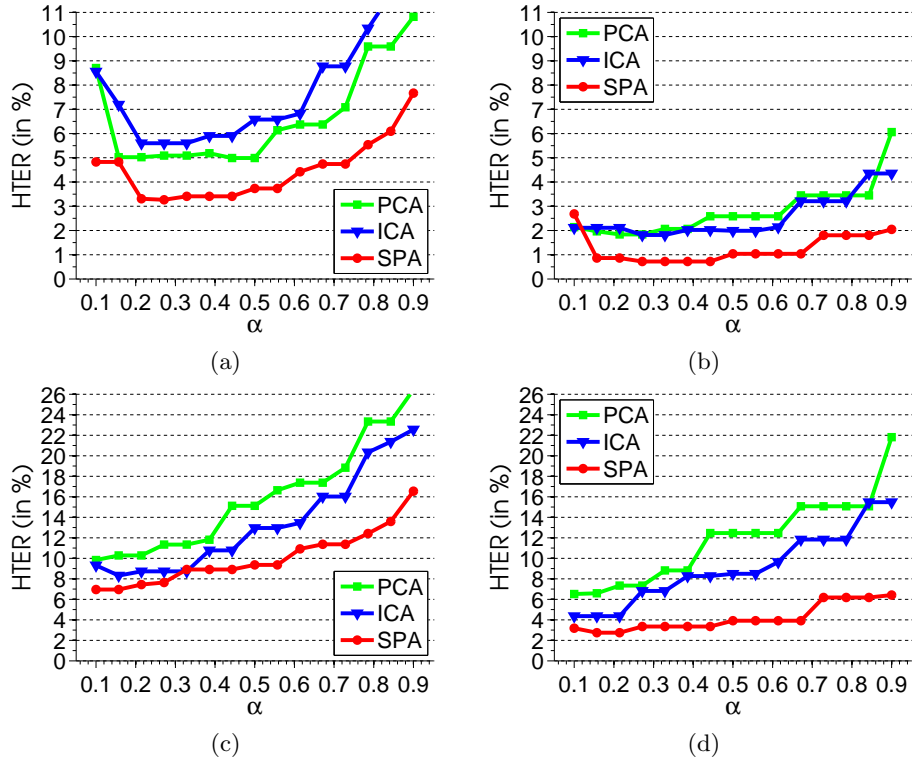


Fig. 3. EPC curves generated during the experiments: (a) on grey-scale images of the original test set, (b) on color images of the original test set, (c) on grey-scale images of the occluded test set, (d) on color images of the occluded test set.



Fig. 4. Examples of occluded images.

at most 30% of the pixels in each image to zero. The location and the size of the occlusion were chosen randomly as shown in the examples presented in Fig. 4.

From the results we can see that the proposed technique results in much better verification rates than the PCA technique, and even the ICA technique, which, similarly as SPA, represents a local method. While all methods deteriorate in their performance, the deterioration is more extreme for the PCA and ICA techniques.

4 Conclusion

In the paper we have presented a novel feature extraction technique for face recognition. The technique, which uses a sparse projection basis to reduce the dimensionality of the face pattern vectors, was assessed in a series of face verification experiments performed on the XM2VTS database. In the experiments it was shown to outperform the popular PCA and ICA techniques and to perform reasonably well even if parts of the facial images are occluded.

References

1. Zhao, W., Chellappa, R., Phillips, J., Rosenfeld, A.: Face Recognition: A Literature Survey. *ACM Computing Surveys* 35(4), pp. 399–458, (2003)
2. Leonardis, A., Bischof H.: Robust Recognition Using Eigenimages. *Computer Vision and Image Understanding* 78(1), 99–118, (2000)
3. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face Recognition by Independent Component Analysis. *IEEE Transactions on Neural Networks* 13(6), 1450–1464, (2002)
4. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2), 210–227, (2009)
5. Guillaumet, D., Vitria, J., Non-negative Matrix Factorization for Face Recognition. In: Proc. of the 5th Catalanian Conference on AI: Topics in Artificial Intelligence, LNCS, Springer, Berlin, Heidelberg, pp. 336–344, (2002)
6. Bociu, I., Pitas, I.: A New Sparse Image Representation Algorithm Applied to Facial Expression Recognition. In: Proc. of the IEEE Workshop on Machine Learning for Signal Processing, pp. 539–548, (2004).
7. Štruc, V., Mihelič, F., Pavešić, N.: Face Authentication Using a Hybrid Approach. *Journal of Electronic Imaging* 17(1), 1–11, (2008)
8. Štruc, V., Gajšek, R., Mihelič, R., Pavešić, N.: Using Regression Techniques for Coping with the One-sample-size Problem of Face Recognition. *Electrotechnical Review* 76(1-2), 7–12, (2009)
9. Rosipal, R.: Kernel Partial Least Squares for Nonlinear Regression and Discrimination. *Neural Network World* 13(3), 291–300, (2003).
10. Chen, S., Donoho, D., Saunders, M.: Atomic Decomposition by Basis Pursuit. *SIAM Review* 43(1), 129–159, (2001)
11. Candes, E., Romberg, J., Tao, T.: Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Comm. Pure and Applied Math.* 59(8), 1207–1223, (2006)
12. Messer, K., Matas, J., Kittler, J., Luettin, J.: XM2VTSDB: the Extended M2VTS Database. In: Proc. of AVBPA, pp. 72–77, (1999).
13. Liu, Z., Liu, C.: Fusion of the Complementary Discrete Cosine Features in the YIQ Color Space for Face Recognition. *Computer Vision and Image Understanding* 111(3), 249–262, (2008)
14. Turk, M., Pentland, A.: Eigenfaces for Recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86, (1991)
15. Bengio, S., Marithoz, J.: The Expected Performance Curve: A New Assessment Measure for Person Authentication. In: Proc. of Odyssey, pp. 279–284, (2004)