

# Symmetric Subspace Learning for Image Analysis

Konstantinos Papachristou, Anastasios Tefas and Ioannis Pitas

Aristotle University of Thessaloniki  
Department of Informatics  
Box 451, 54124 Thessaloniki, Greece  
{tefas,pitas}@aiia.csd.auth.gr

**Abstract**—Subspace Learning is one of the most useful tools for image analysis and recognition. A large number of such techniques has been proposed utilizing a-priori knowledge about the data. In this paper, new Subspace Learning techniques are presented that use symmetry constraints in their objective functions. The rational behind this idea is to exploit the a-priori knowledge that geometrical symmetry appears in several types of data, such as images, objects, faces, etc. Experiments on artificial, facial expression recognition, face recognition and object categorization databases highlight the superiority and the robustness of the proposed techniques, in comparison to standard Subspace Learning techniques.

**Index Terms**—Subspace learning, symmetry constraints, principal component analysis (PCA), linear discriminant analysis (LDA), clustering based discriminant analysis (CDA).

## I. INTRODUCTION

Visual information is an important part of our daily life. As a result, visual information is exploited in many applications such as robotics, multimedia retrieval and face or object recognition. Many of these applications exploit Subspace Learning (SL) techniques, which have been employed in many computer vision and pattern recognition tasks [1], [2], such as image retrieval, face detection, face recognition, facial expression recognition and object recognition. Such techniques project the high-dimensional data on low-dimensional spaces and, thus, can be employed for dimensionality reduction, data visualization and compression, as well as a main preprocessing step in pattern classification and clustering. The adoption of SL techniques leads to methods which are faster, use less memory and have improved classification performance. All SL techniques are based on the optimization of a given objective function to calculate projection matrices. Various objective functions have been proposed to this end. Many of them exploit a-priori knowledge related to data nature (e.g., illumination, geometry) and/or dataset structure (e.g., classes and subclasses).

A fundamental category of the SL algorithms is unsupervised. The most well known SL algorithm is Principal Component Analysis (PCA) [3], [4]. PCA projects the data along the directions of maximum data variance. Independent Component Analysis [5], [6] does not require the projection directions are orthogonal to each other. Instead, it tries to find directions, where the data projections are independent

from each other. Locality Preserving Projection (LPP) [7], [8] tries to preserve the local structure of the data space during dimensionality reduction. Non-negative Matrix Factorization (NMF) [9] is a data decomposition algorithm, which requires that both the projection vectors and the related coefficients to be non-negative. Most of the above SL algorithms have been represented in the graph embedding framework [10].

Another SL category called Discriminant Analysis (DA) methods is supervised and uses the data class label information. The most popular supervised SL technique is Linear Discriminant Analysis (LDA) [11], [12], [13], which seeks for a linear subspace, where the projected data classes are optimally discriminated. LDA-based techniques have been successfully used in many computer vision applications, such as human face recognition [14] and facial expression classification [15]. In [16], a discriminant NMF algorithm has been proposed, which achieves data decomposition, while increasing the between-class data dispersion and decreasing the within-class dispersion. In [17], a LPP algorithm was proposed, which takes into account discriminant information to compute the optimal discriminant projection subspace. Two normalized versions of PCA and LDA have been proposed in [18] that use class information to preserve the local data structure. The above algorithms assume that each class consists of a single cluster (subclass). However, if a class is divided into two or more subclasses, the above algorithms do not achieve the desired performance. To overcome this deficiency, two LDA-like algorithms, called Clustering based Discriminant Analysis (CDA) and Subclass Discriminant Analysis, were presented in [19] and [20], respectively. They take into account cluster information to compute the between-class and within-class scatter matrices, in order to enhance the class separability in the new projection space. Other techniques [21], [22] aim to identify a low dimensional projection subspace which ensures that the margin between the projected data of different classes is maximal.

All the above algorithms handle vector data and ignore the fact that such data often have specific geometric structure. Tensor object can be used to describe such structures. For example, an image is a second-order tensor, whereas a video is a third-order tensor. Algorithms similar to PCA [23] and to LDA [24], [25] have been proposed for representing  $k$ th-order tensor objects. Also, new algorithms which use two-

dimensional data (images) as input and are extensions of PCA [26], LDA [27] and LPP [28] have been proposed. In [29], SL algorithms based on tensor representation were proposed, which use the spatial information of the images, adding Laplacian constraints in objective functions, in order to generate projection vectors that are spatially smooth.

In practise, data class information is often unavailable. Furthermore, the training set may consist of both labeled and unlabeled data. In [30], [31], [32], [33], [34], semi-supervised extensions of SL algorithms have been proposed, which use all available data (both labeled and unlabeled) to learn optimal discriminant subspaces. Other common problems, which deteriorates the performance of SL algorithms, are data noise (e.g., due to occlusion, illumination variations) and lack of image registration (alignment). To overcome such problems, robust extensions of SL algorithms have been proposed in [35], [36], [32], [37], [38], [39].



Fig. 1. Symmetry is in machines, textures, crystals, molecules, flowers.

The images depicted in Figure 1 have a common feature exhibited in nature: geometrical symmetry. Symmetry is everywhere: from natural objects such as crystals, leaves, flowers, animals, to human constructions e.g., buildings, machines, computer graphics. Thus, there is no surprise that human vision has evolved to detect symmetric patterns [40].

Symmetry has been used in many scientific fields. Symmetries have been used in differential equations [41], in order to reduce the total number of independent variables. By introducing symmetry constraints in eigenanalysis, frequency estimation is improved, as has been shown in [42]. In the design of analog circuits, some modules are required to be placed symmetrically, in order to avoid the high offset voltages, because of the bad layout from parasitics and to reduce the circuit sensitivity to thermal gradients. In order to restrict the overlaps between modules, various optimization methods have been developed, which use the symmetry constraint as a penalty term in the cost function in [43] and [44]. Symmetry is used in search algorithms, so that only unique solutions are returned. Symmetry constraints are used, so that the symmetric versions of the failed part of the search space will not be considered in future [45].

The symmetry of the human face and other objects has also been used in computer vision. One symmetry measure has been proposed in [46] that computes a symmetry map of an image, which can be used for extracting and grouping interest points. A gait recognition method employing human

motion symmetry, has been proposed in [47]. Shape reconstruction, which exploits the fact that the occluded surfaces of a symmetric object can be reconstructed using object symmetries has been presented in [48]. A method for 3D face authentication and recognition has been proposed in [49] which computes three interest points of the nose of a face, using both the facial surface and symmetry. It was shown that these points can determine a unique face. In [50], the authors have defined two types of facial asymmetry measures to select automatically local facial regions, which are stable to facial expression variations. Furthermore, they combined asymmetry information with PCA and LDA dimensionality reduction, for human identification. A similar method for facial expression recognition has been presented in [51]. A human face and facial feature detection method, which uses symmetry-based cost functions in order to find the eye/nose/mouth locations for face detection was described in [52].

Such methods use the knowledge that the image structure can be symmetric for computing interest points or specific regions. In the case of SL algorithms, we would expect that the image data symmetries are preserved during projection. However, this is not always true, either due to the small number of data samples, or data noise (occlusion, illumination), or both. In this paper, the main goal is to use the a-priori knowledge on data symmetry, in order to learn projection subspaces equipped with more robustness and generalization ability. This is performed in a novel way, by adding symmetry constraints in the objective functions of the SL algorithms. Thus, we propose new SL algorithms based on the standard PCA, LDA, CDA and their combinations.

The rest of this paper is organized as follows: In Section II, the standard SL algorithms are reviewed. Section III introduces the symmetry constraint and the proposed symmetric extensions of the SL algorithms. In Section IV, the experimental results of the proposed algorithms are compared with those of the standard SL ones. Finally, some concluding remarks are given in Section V.

## II. SUBSPACE LEARNING

In this section, we provide a brief review of standard SL techniques, namely PCA, LDA, CDA. In the following, the image set  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  contains sample images  $\mathbf{x}_i \in \mathcal{R}^{m \times 1}$  in vectorized form. The projection vectors are denoted by  $\mathbf{w} \in \mathcal{R}^{m \times 1}$ . The total number of samples in the image dataset, the total number of image classes and the mean vector of the entire image dataset are denoted by  $N$ ,  $c$  and  $\boldsymbol{\mu}$ , respectively. The initial dimensionality of the samples is denoted by  $m$ , while the dimensionality of the projection space is denoted by  $m'$  (typically  $m' \ll m$ ).

### A. Principal Component Analysis

PCA [3], [12] tries to find projection vectors  $\mathbf{w}$  that maximize the variance of the projected samples  $y_i = \mathbf{w}^T \mathbf{x}_i$  along  $\mathbf{w}$ , for better data representation. If we define the total scatter matrix  $\mathbf{S}_T$  as:

$$\mathbf{S}_T = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T, \quad (1)$$

the objective of PCA is to find the transformation matrix  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{m'}]$  that maximizes the trace of  $\mathbf{S}_T$ :

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}} \text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}]. \quad (2)$$

The solution of (2) is equal to the solution of following generalized eigenvalue decomposition problem:

$$\mathbf{S}_T \cdot \mathbf{w} = \lambda \cdot \mathbf{w}, \quad (3)$$

when retaining the  $m'$  eigenvectors of  $\mathbf{S}_T$  that correspond to the  $m'$  largest eigenvalues. The eigenvector that corresponds to the largest eigenvalue projects the samples along the direction of maximal variance. Any sample  $\mathbf{x}_i$  from the initial space can now be approximated by a linear combination of the  $m'$  first eigenvectors to produce a new  $m'$ -dimensional vector. We can choose  $m'$  such that the sum of the  $m'$  largest eigenvalues is more than a percentage  $P\%$  of the total eigenvalue sum.

### B. Linear Discriminant Analysis

In contrast to PCA, LDA [53], [54] determines projection vectors  $\mathbf{w}$  along which the data classes are well separated. For this reason, the between-class scatter matrix:

$$\mathbf{S}_B^{LDA} = \sum_{i=1}^c (\boldsymbol{\mu}_i - \boldsymbol{\mu}) (\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \quad (4)$$

and the within-class scatter matrix:

$$\mathbf{S}_W^{LDA} = \sum_{i=1}^c \sum_{k=1}^{n_i} (\mathbf{x}_k^i - \boldsymbol{\mu}_i) (\mathbf{x}_k^i - \boldsymbol{\mu}_i)^T, \quad (5)$$

are defined.  $\mathbf{x}_k^i$  is the  $k$ -th sample in the class  $i$  and  $\boldsymbol{\mu}_i$ ,  $n_i$  are the mean vector and the number of samples in class  $i$ , respectively. The objective of LDA is to find the transformation matrix  $\mathbf{W}$  that maximizes the ratio of the trace of the between-class scatter to the trace of the within-class scatter matrix:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_B^{LDA} \mathbf{W}]}{\text{tr}[\mathbf{W}^T \mathbf{S}_W^{LDA} \mathbf{W}]} \quad (6)$$

The solution of (6) is approximated [55], [56] by the following generalized eigenvalue decomposition problem:

$$\mathbf{S}_B^{LDA} \cdot \mathbf{w} = \lambda \cdot \mathbf{S}_W^{LDA} \cdot \mathbf{w}, \quad (7)$$

by keeping the  $m'$  eigenvectors that correspond to the  $m'$  largest eigenvalues. Because  $\mathbf{S}_B^{LDA}$  is the sum of  $c$  matrices in (4) of rank one or less and only  $c-1$  of these are independent, the maximum number of nonzero eigenvalues is equal to  $c-1$ . Consequently, the upper bound on  $m'$  is  $c-1$ .

### C. Clustering based Discriminant Analysis

LDA is optimal when the samples of each class are generated from a normal distribution [57]. However, in many applications the classes consist of several clusters (also called subclasses) and follow multimodal data distributions. In such cases, LDA is not the optimal projection. CDA [19] determines projection vectors  $\mathbf{w}$ , so that the corresponding clusters are

well discriminated after the projection. The between-cluster scatter and the within-cluster scatter matrices are defined by:

$$\mathbf{S}_B^{CDA} = \sum_{i=1}^{c-1} \sum_{l=i+1}^c \sum_{j=1}^{d_i} \sum_{h=1}^{d_l} (\boldsymbol{\mu}_j^i - \boldsymbol{\mu}_h^l) (\boldsymbol{\mu}_j^i - \boldsymbol{\mu}_h^l)^T, \quad (8)$$

$$\mathbf{S}_W^{CDA} = \sum_{i=1}^c \sum_{j=1}^{d_i} \sum_{k=1}^{n_{ij}} (\mathbf{x}_k^{ij} - \boldsymbol{\mu}_j^i) (\mathbf{x}_k^{ij} - \boldsymbol{\mu}_j^i)^T, \quad (9)$$

where  $d_i$  is the number of clusters in class  $i$ ,  $\mathbf{x}_k^{ij}$  denotes the  $k$ th sample of the  $j$ th cluster in class  $i$  and,  $\boldsymbol{\mu}_j^i$ ,  $n_{ij}$  are the mean vector and the number of samples in the  $i$ th cluster of class  $i$ , respectively. The objective of CDA is to find the transformation matrix  $\mathbf{W}$  that maximizes (6), where  $\mathbf{S}_B^{LDA}$ ,  $\mathbf{S}_W^{LDA}$  are replaced by  $\mathbf{S}_B^{CDA}$ ,  $\mathbf{S}_W^{CDA}$ , respectively. In this case, the solution of (6) is again approximated by (7). If the total number of the sample clusters is  $d$  the upper bound on  $m'$  is  $d-1$ .

### D. Principal Component Analysis plus Discriminant Analysis

The Discriminant Analysis (DA) techniques (LDA, CDA) are very prone to the 'small sample size' problem [1]. In order to overcome this problem an alternative has been proposed [58], [12]. Firstly, the samples are projected to a subspace of dimensionality lower than  $N-l$  using PCA, where  $l$  denotes the number of classes or clusters for LDA or CDA techniques respectively. Finally, LDA or CDA is applied for the determination of final projection vectors.

More formally, the final transformation matrix  $\mathbf{W}_f$  is given by:

$$\mathbf{W}_f = \mathbf{M}_P \mathbf{M}_D, \quad (10)$$

where

$$\mathbf{M}_P = \arg \max_{\mathbf{W}} \text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}], \quad (11)$$

$$\mathbf{M}_D = \arg \max_{\mathbf{W}} \frac{\text{tr}[\mathbf{W}^T (\mathbf{M}_P^T \mathbf{S}_B \mathbf{M}_P) \mathbf{W}]}{\text{tr}[\mathbf{W}^T (\mathbf{M}_P^T \mathbf{S}_W \mathbf{M}_P) \mathbf{W}]} \quad (12)$$

$\mathbf{S}_B$ ,  $\mathbf{S}_W$  denote  $\mathbf{S}_B^{LDA}$ ,  $\mathbf{S}_W^{LDA}$  or  $\mathbf{S}_B^{CDA}$ ,  $\mathbf{S}_W^{CDA}$ , respectively.

## III. SYMMETRIC SUBSPACE LEARNING

Many computer vision problems operate on symmetric image patterns, such as facial images or facial details (eyes, mouth, nose) as illustrated in Figure 2. In such cases, symmetry is a main characteristic of the image data. Therefore, it would be expected for the generated SL outputs to be symmetric, in order to achieve a higher generalization capability and not to suffer from the over-training phenomenon. Therefore, the generated projection vectors should be symmetric as well.

However, this does not usually happen for the following reasons. Firstly, the image training set very often consists of a small number of image samples, resulting in a poor pattern representation and, therefore, in a bad pattern learning and generalization. A second reason is that the image samples are not strictly symmetric since, e.g., facial expressions or poses may be unsymmetric. As a result, the SL output is not symmetric either. There are several solutions to overcome



Fig. 2. Facial images under various illumination conditions, expressions and facial details ([59]).

these problems. The first one is to increase the number of samples in the image training set and, thus, to obtain a better image representation. A second possible solution is to perform manual training data selection, in order to exclude image samples which are not quite symmetric. However, these solutions are neither efficient nor easy to be applied, because on one hand it is difficult to increase the number of samples, while, on the other hand, it is quite tedious to select the symmetric samples. Also, the decision concerning what samples are symmetric is ambiguous, since all the samples are usually inevitably asymmetric.

In this section, we modify the SL techniques, so that the property of symmetry is taken into account. Firstly, we refer to the necessary preprocessing steps so that image symmetry to be retained when vectorizing the images. The symmetry constraint is presented in Subsection III-B. Finally, in Subsections III-C to III-F, we describe symmetric extensions of the SL techniques presented in Section II.

#### A. Image symmetry to vector symmetry

Symmetry is usually encountered in image classification problems, which are two-dimensional data. Since the presented SL techniques use vectors as input, it is necessary to provide an appropriate scanning method corresponding to the order that a set of pixels is read, which is able to retain the symmetry when vectorizing the images. The two known scanning methods used to convert an image to a vector, namely horizontal (row-wise) and vertical (column-wise) scanning, are not appropriate for the conversion of an image to a vector, when we want to retain data symmetry. An alternative scanning method, which maintains the image symmetry and, therefore, can be used in symmetric SL techniques, is shown in Figure 3. By applying such a scanning, the symmetry along the vertical axis is retained. Similarly, we can apply a corresponding scanning to convert a symmetric image into a vector on the horizontal axis or on any directional axis.

When symmetry is not a-priori known, it can be automatically detected using a symmetry detection algorithm [60],

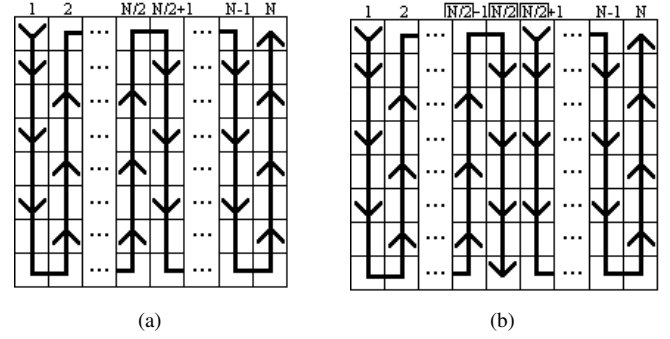


Fig. 3. The alternative scanning method of an image for a) even and (b) odd column number  $N$ .

[61], [62]. By performing such an algorithm, the images can be normalized to have e.g., vertical symmetry. The images can be transformed in symmetric vectors then by applying the corresponding scanning method.

In both cases, either the symmetry is known a-priori (e.g., in facial images) or it is detected using a symmetry detection algorithm, the proposed symmetric subspace analysis is applied afterwards.

#### B. Symmetry Constraint

In the proposed SL techniques, the goal is the determination of projection vectors that are symmetric, so that the samples are projected in symmetric discriminant subspaces. The symmetry error of a vector  $\mathbf{w} = [w_1, w_2, \dots, w_{d-1}, w_d]^T$  is given by the following equation:

$$s_{\mathbf{w}} = \sum_{i=1}^{d/2} (w_i - w_{d+1-i})^2. \quad (13)$$

The objective functions used in SL techniques use projection matrices. Therefore, we define a symmetry matrix  $\mathbf{A}$ , whose multiplication with a projection vector ( $\mathbf{w}^T \mathbf{A} \mathbf{A}^T \mathbf{w}$ ) should measure the projection vector symmetry. If the  $d \times d$  symmetry matrix takes the form:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \dots & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \dots & -\frac{1}{\sqrt{2}} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -\frac{1}{\sqrt{2}} & \dots & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \dots & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad (14)$$

it can be easily proven that:

$$s_{\mathbf{w}} = \mathbf{w}^T \mathbf{A} \mathbf{A}^T \mathbf{w} = \sum_{i=1}^{d/2} (w_i - w_{d+1-i})^2. \quad (15)$$

It is straightforward to prove that an equivalent, but computationally superior, form of  $\mathbf{A}$  is the following  $d \times \frac{d}{2}$  matrix:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \dots \\ 0 & \frac{1}{\sqrt{2}} & \dots \\ \vdots & \vdots & \ddots \\ 0 & -\frac{1}{\sqrt{2}} & \dots \\ -\frac{1}{\sqrt{2}} & 0 & \dots \end{bmatrix}. \quad (16)$$

Our goal is to introduce this constraint in the objective functions of the SL techniques presented in Section II, by minimizing the quantity  $\text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]$ .

### C. Symmetric Principal Component Analysis

The objective of Symmetric Principal Component Analysis (SPCA) is to calculate those projection vectors  $\mathbf{w}$ , which both maximize the projected sample variance and are symmetric.

Specifically, we want to maximize  $\text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}]$  in order to maximize the variance of the projected samples, while, on the other hand, we want to minimize  $\text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]$ , in order to minimize the symmetry error of the projection vectors. Thus, we define the following objective function:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}]}{(1-s) \text{tr}[\mathbf{W}^T \mathbf{I} \mathbf{W}] + s \text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]}, \quad (17)$$

where  $\mathbf{I}$  is the  $m \times m$  identity matrix and  $s \in [0, 1]$  is a symmetry factor that controls the symmetry of  $\mathbf{w}$ . The identity matrix  $\mathbf{I}$  is used to control the level of symmetry error to overcome the singularity of  $\mathbf{A} \mathbf{A}^T$ . It is obvious that, for  $s = 0$ , SPCA is equivalent to PCA, while, as  $s \rightarrow 1$ , the level of symmetry of the projection vectors is maximized. For avoiding the trivial solution,  $\mathbf{W}$  is constrained by imposing the orthogonal constraint [63]. The solution of (17) is given by the solution of following generalized eigenvalue decomposition problem:

$$\mathbf{S}_T \cdot \mathbf{w} = \lambda \cdot ((1-s)\mathbf{I} + s\mathbf{A} \mathbf{A}^T) \cdot \mathbf{w}, \quad (18)$$

by keeping the  $m'$  eigenvectors of  $\mathbf{S}_T$  that correspond to the  $m'$  largest eigenvalues.

To show the effectiveness of PCA and the symmetric extension, we applied them to ETH-80 database [64]. It contains images from eight categories: apple, pear, tomato, cow, horse, dog, cup and car. We have used the object categories corresponding to symmetric objects (apple, cup, pear, tomato). The database was randomly divided into two subsets. The first subset was used as training set, while the second one as test set. Figure 4 illustrates the 2-D projections of the test samples using the projection spaces determined by PCA and SPCA, respectively. It can be easily seen that 2-D projection corresponding to SPCA has improved discriminant ability in comparison to PCA.

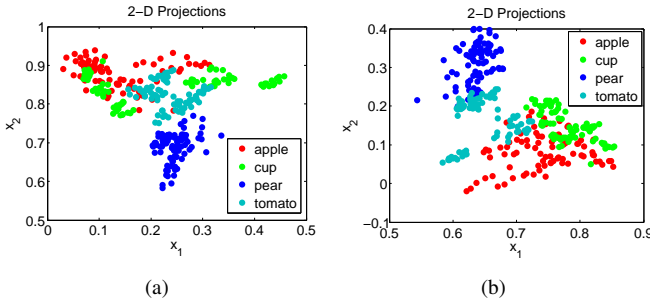


Fig. 4. a) PCA and b) SPCA 2-D projections for the ETH-80 database [64].

### D. Symmetric Linear Discriminant Analysis

The objective of Symmetric Linear Discriminant Analysis (SLDA) is to exploit symmetry for the determination of projection vectors  $\mathbf{w}$ , which both increase class discrimination and are symmetric.

To this end, we want to maximize  $\text{tr}[\mathbf{W}^T \mathbf{S}_B^{LDA} \mathbf{W}]$ , so that the dispersion of samples from different classes will be maximized after the projection, while, at the same time, we want to minimize a) the trace of the  $\mathbf{W}^T \mathbf{S}_W^{LDA} \mathbf{W}$  so that samples from the same classes will come as close as possible to their mean vector after the projection and b) to minimize the trace of the  $\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}$ , so that the symmetry error becomes as small as possible after the projection:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_B^{LDA} \mathbf{W}]}{(1-s) \text{tr}[\mathbf{W}^T \mathbf{S}_W^{LDA} \mathbf{W}] + s \text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]}. \quad (19)$$

The solution of (19) is approximated by the following generalized eigenvalue decomposition problem:

$$\mathbf{S}_B^{LDA} \cdot \mathbf{w} = \lambda \cdot ((1-s)\mathbf{S}_W^{LDA} + s\mathbf{A} \mathbf{A}^T) \cdot \mathbf{w}, \quad (20)$$

by keeping the  $m'$  eigenvectors that correspond to the  $m'$  largest eigenvalues. The upper bound on  $m'$ , as in the case of LDA, is  $c - 1$ , where  $c$  is the total number of classes.

To highlight the effectiveness of SLDA we used the ETH-80 database. LDA and SLDA projection matrices were determined by a part of the database, while the remaining images were projected onto the two determined projection spaces. Figure 5 illustrates the two 2-D projections as obtained by LDA and SLDA, respectively. In the case of LDA, projected samples are not separated well. On the other hand, it is obvious that SLDA achieves a better projection: samples belonging to “pear” and “cup” classes are clearly separated, while a better separation is achieved between samples belonging to “apple” and “tomato” classes.

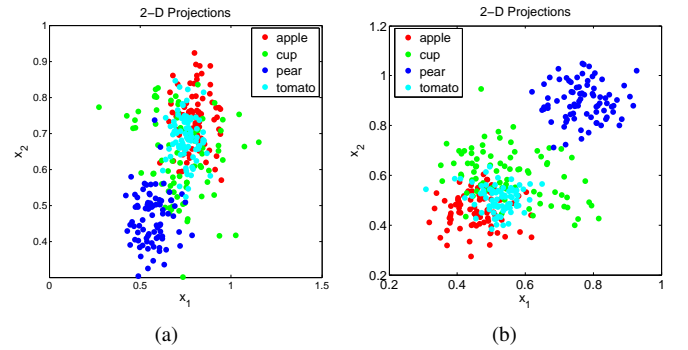


Fig. 5. a) LDA and b) SLDA 2-D projections for the ETH-80 database [64].

### E. Symmetric Clustering based Discriminant Analysis

In the same context, the objective of Symmetric Clustering based Discriminant Analysis (SCDA) is to determine the projection vectors  $\mathbf{w}$  that increase cluster discrimination and are symmetric, by optimizing:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_B^{SCDA} \mathbf{W}]}{(1-s) \text{tr}[\mathbf{W}^T \mathbf{S}_W^{SCDA} \mathbf{W}] + s \text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]}. \quad (21)$$



The solution of (21) is approximated by solving the following generalized eigenvalue decomposition problem:

$$\mathbf{S}_B^{CDA} \cdot \mathbf{w} = \lambda \cdot ((1-s)\mathbf{S}_W^{CDA} + s\mathbf{A}\mathbf{A}^T) \cdot \mathbf{w}, \quad (22)$$

by keeping the  $m'$  eigenvectors that correspond to the  $m'$  largest eigenvalues. Because the maximum number of nonzero eigenvalues is equal to  $d-1$ , as described in Subsection II-C, the upper bound on  $m'$  is  $d-1$ , where  $d$  is the total number of clusters of the samples.

Like LDA and PCA, two projection matrices were determined by CDA and SCDA using the ETH-80 database. Figure 6 illustrates the two obtained 2-D projections. As can be clearly seen, SCDA achieves to determine a projection space with improved discriminant ability in comparison to CDA. That is, “pear” and “cup” classes are well separated, while a better separation is achieved between “apple” and “tomato” classes.

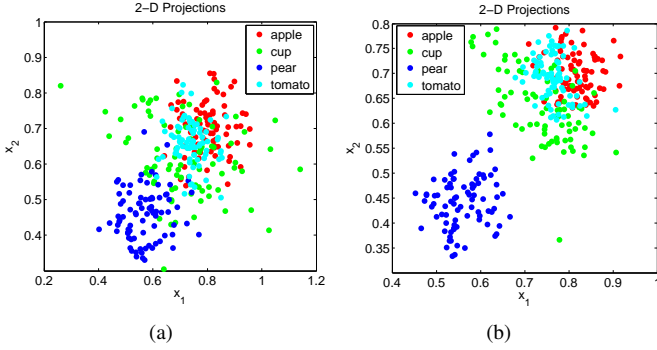


Fig. 6. a) CDA and b) SCDA 2-D projections for the ETH-80 database [64].

#### F. Symmetric Principal Component Analysis plus Symmetric Discriminant Analysis

The same approach can be followed for PCA+LDA or PCA+CDA, for employing symmetry constraints. Firstly, SPCA is applied to map the samples' space into the SPCA space. SPCA is used instead of PCA technique, because PCA would cause the loss of symmetry of the projected samples in the PCA space, and thus applying the SDA techniques would be meaningless. Finally, in SPCA space, a SDA technique (SLDA or SCDA) is applied using a modified symmetry constraint to compute the final projection vectors. A modified symmetry constraint should be used because our goal is to determine projection vectors that are symmetric with respect to the initial space and not the SPCA space. Specifically, a final projection vector  $\mathbf{w}_f$  is derived as follows

$$\mathbf{w}_f = \mathbf{M}_{SP}\mathbf{w}_{SD}, \quad (23)$$

where  $\mathbf{M}_{SP}$  is the SPCA transformation matrix and  $\mathbf{w}_{SD}$  is a generated projection vector from an SDA technique. Using (15) and (23), the symmetry of  $\mathbf{w}_f$  is computed as follows

$$\begin{aligned} S_{\mathbf{w}_f} &= \mathbf{w}_f^T \mathbf{A} \mathbf{A}^T \mathbf{w}_f \\ &= (\mathbf{M}_{SP} \mathbf{w}_{SD})^T \mathbf{A} \mathbf{A}^T \mathbf{M}_{SP} \mathbf{w}_{SD} \\ &= \mathbf{w}_{SD}^T \mathbf{M}_{SP}^T \mathbf{A} \mathbf{A}^T \mathbf{M}_{SP} \mathbf{w}_{SD}. \end{aligned}$$

Therefore, in the case of SDA techniques the symmetry constraint is  $\mathbf{M}_{SP}^T \mathbf{A} \mathbf{A}^T \mathbf{M}_{SP}$ .

Finally, the final transformation matrix  $\mathbf{W}_f$  is given by

$$\mathbf{W}_f = \mathbf{M}_{SP} \mathbf{M}_{SD}, \quad (24)$$

where  $\mathbf{M}_{SP}$  and  $\mathbf{M}_{SD}$  are given by (25) and (26) at the bottom of the page, respectively.

#### IV. EXPERIMENTS

In this section, we evaluate the performance of the proposed symmetric SL techniques and we compare it with the performance of the standard SL techniques in real-databases, which contain symmetric data. We selected three pattern recognition problems: 1) facial expression recognition, 2) face recognition and 3) object categorization.

In all the experiments, the recognition process is performed as follows. First, we convert each image into a vector by using the scanning method presented in Subsection III-A. Then, we calculate the projection subspace, by applying a SL technique to the training set. During testing, the samples are projected into the corresponding subspace. Finally, the projected samples were classified using the Nearest Centroid (NC) and k-Nearest Neighbor (kNN) classifiers. kNN was used for  $k = 1, 3, 5, 7, 9, 11$ . In all classifiers, the Euclidean distance measure is adopted. The SL techniques applied are PCA, LDA, CDA, PCA+LDA, PCA+CDA together with the corresponding symmetric extensions. The new dimensionality of PCA has been defined by maintaining the 99% of the total eigenvalue sum of the training set, while in DA techniques the new dimensionality was  $l-1$ , where  $l$  is the number of classes or clusters for LDA or CDA techniques, respectively. For clustering-based SL techniques (CDA, SCDA), the training set is divided into a number of clusters, by applying various clustering techniques. For ease of representation, we will follow the notation  $\text{CDA}(a,n)$ , where  $a$  denotes the clustering technique (km for k-means and fcm for fuzzy c-means) and  $n$  is the number of clusters. SPCA was used for  $s = 0.0, 0.1, \dots, 1.0$ , while SLDA and SCDA were used for  $s = 0.0, 0.1, \dots, 0.9999$ .

The results of our experiments on facial expression recognition, face recognition and object categorization are presented in Subsections IV-A, IV-B and IV-C, respectively. Subsection IV-D describes the effect of the symmetry factor in the

$$\mathbf{M}_{SP} = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}]}{(1-s) \text{tr}[\mathbf{W}^T \mathbf{I} \mathbf{W}] + s \text{tr}[\mathbf{W}^T \mathbf{A} \mathbf{A}^T \mathbf{W}]} \quad (25)$$

$$\mathbf{M}_{SD} = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}[\mathbf{W}^T (\mathbf{M}_{SP}^T \mathbf{S}_B \mathbf{M}_{SP}) \mathbf{W}]}{(1-s) \text{tr}[\mathbf{W}^T (\mathbf{M}_{SP}^T \mathbf{S}_W \mathbf{M}_{SP}) \mathbf{W}] + s \text{tr}[\mathbf{W}^T (\mathbf{M}_{SP}^T \mathbf{A} \mathbf{A}^T \mathbf{M}_{SP}) \mathbf{W}]} \quad (26)$$

performance of the presented techniques, the symmetry error of generated projection vectors and the new dimensionality of the obtained subspaces. Finally, a different usage of symmetry is presented and is compared with the results of our proposed techniques in Subsection IV-E.

#### A. Experiments on Facial Expression Recognition

The COHN-KANADE [65], BU [66], JAFFE [67] and FER-AIIA [68] face databases were used in our experiments for facial expression recognition. Each facial image belongs to one of the following seven facial expressions: anger, disgust, happiness, fear, sadness, surprise and neutral expression. The COHN-KANADE database contains 210 subjects of age between 18 and 50 years (69% female, 31% male, 81% Euro-American, 13% Afro-American and 6% other groups). We used 35 images of each facial expression. The BU database consists of 100 subjects (60% female and 40% male) with a variety of ethnic/racial background, including White, Black, East-Asian, Middle-east Asian, Hispanic Latino and others. All expressions, except the neutral one, are expressed at four intensity levels. The most expressive intensity of each facial expression was used in our experiments. The JAFFE database contains 213 images depicting 10 Japanese female subjects. Each subject has 3 or 4 images of each facial expression. We used 3 images for subject of each facial expression in our experiments. The FER-AIIA database has been constructed from the Artificial Intelligence and Information Analysis Laboratory, Aristotle University of Thessaloniki, Greece. It consists of image sequences depicting five persons. Each person interprets each of the 7 expressions, starting from the neutral expression and ending with the expression of maximum intensity. In our experiments, we used 20 images of maximum intensity for each expression for each of the five persons. All facial images were cropped to include only the subject's facial region containing the main facial features (as eyebrows, eyes, nose and mouse). The cropped face images were resized to  $40 \times 30$  pixels. In Figures 7, 8, 9 and 10, a cropped facial image for all facial expressions of the COHN-KANADE, BU, JAFFE and FER-AIIA databases is shown, respectively.

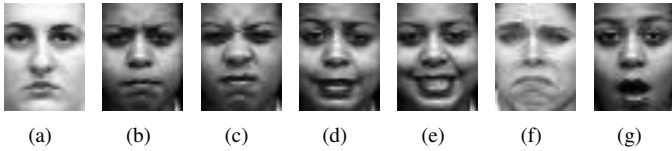


Fig. 7. A cropped image for all facial expressions of the Cohn-Kanade database: (a) neutral, (b) angry, (c) disgusted, (d) feared, (e) happy, (f) sad, and (g) surprised.

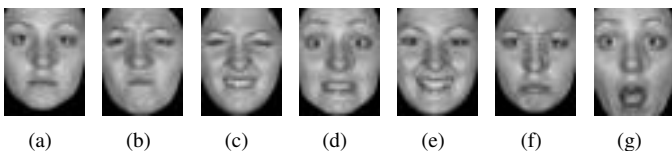


Fig. 8. A cropped image for all facial expressions of a subject of the BU database: (a) neutral, (b) angry, (c) disgusted, (d) feared, (e) happy, (f) sad, and (g) surprised.

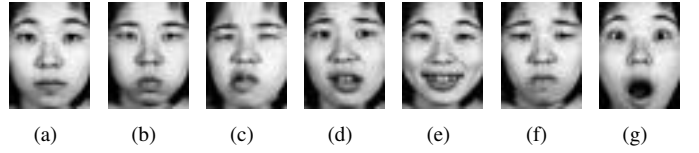


Fig. 9. A cropped image for all facial expressions of a subject of the JAFFE database: (a) neutral, (b) angry, (c) disgusted, (d) feared, (e) happy, (f) sad, and (g) surprised.

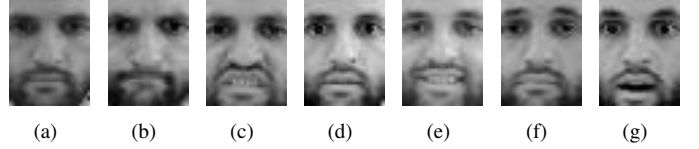


Fig. 10. A cropped image of each facial expression for a subject from the FER-AIIA database: (a) neutral, (b) angry, (c) disgusted, (d) feared, (e) happy, (f) sad, and (g) surprised.

We begin the description of experimental results with the COHN-KANADE, BU and JAFFE databases. The application of DA techniques (LDA, CDA, SLDA, SCDA) on the above databases encounters computational difficulties due to the 'small sample size' problem. For this reason, in order to examine the behaviour of the proposed techniques on bigger databases, we constructed enriched versions of above three databases following the process described in [68]. Specifically, each facial image was translated, rotated and scaled according to 25 different geometric transformations. Each eye has five possible positions: original position, one-pixel left, right, up and down (resulting into 25 pairs). Therefore, the enriched version of a database consists of the original database (centered images) and the 24 shifted images for each centered image.

We conducted two series of experiments, one by using the original data version and the second one by using the enriched versions. In the second case, testing was performed by using the centered facial images. To estimate the recognition accuracy, we used the 5-fold cross validation procedure. More specifically, each database was divided into 5 non-overlapping subsets. Each experiment includes five training-test procedures (folds). In each fold, the techniques were trained by using 4 subsets and testing was performed on the remaining subset. Recognition accuracy was measured by using the mean classification rate over all five folds. For the COHN-KANADE experiments, each subset contained 20% of the facial images for each class based on random selection. For the BU and JAFFE databases, we performed person-independent experiments: each subset contained the entire set of the facial images from 20% of the persons. Thus, the facial images of each person were either in the training or in the test set. We used the same partition of the databases in the experiments performed on the original and the enriched databases. The results obtained for the COHN-KANADE, BU and JAFFE databases, are shown in Tables I, II and III, respectively. In the first column of these tables, the applied SL techniques are given. The second and third columns illustrate the best recognition accuracies obtained by applying the standard SL and symmetric SL techniques on the original version of the databases. The next two columns illustrate the

best recognition accuracies obtained for the enriched version of the databases. The best results are shown in bold.

TABLE I

COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %) OF STANDARD SL VERSUS SYMMETRIC SL ON THE COHN-KANADE DATABASE

technique	Original database		Enriched database	
	Standard	Symmetric	Standard	Symmetric
PCA	33.88 $\pm$ 3.20	<b>37.55<math>\pm</math>5.15</b>	37.55 $\pm$ 4.87	<b>37.96<math>\pm</math>4.65</b>
LDA	-	-	73.88 $\pm$ 6.24	<b>75.92<math>\pm</math>4.68</b>
CDA(km,2)	-	-	65.31 $\pm$ 10.4	<b>68.57<math>\pm</math>7.00</b>
CDA(fcm,2)	-	-	66.53 $\pm$ 6.98	<b>70.20<math>\pm</math>8.42</b>
CDA(km,3)	-	-	60.82 $\pm$ 9.27	<b>65.71<math>\pm</math>10.75</b>
CDA(fcm,3)	-	-	66.53 $\pm$ 6.32	<b>69.80<math>\pm</math>6.01</b>
PCA+LDA	68.98 $\pm$ 6.45	<b>72.24<math>\pm</math>6.01</b>	76.73 $\pm$ 4.32	<b>78.37<math>\pm</math>4.13</b>
PCA+CDA(km,2)	63.67 $\pm$ 4.54	<b>67.35<math>\pm</math>6.12</b>	68.98 $\pm$ 10.26	<b>71.43<math>\pm</math>10.05</b>
PCA+CDA(fcm,2)	64.49 $\pm$ 3.99	<b>66.94<math>\pm</math>3.54</b>	69.80 $\pm$ 10.01	<b>71.84<math>\pm</math>10.16</b>
PCA+CDA(km,3)	61.63 $\pm$ 6.27	<b>63.27<math>\pm</math>6.54</b>	63.27 $\pm$ 5.97	<b>66.94<math>\pm</math>6.42</b>
PCA+CDA(fcm,3)	60.41 $\pm$ 5.64	<b>65.31<math>\pm</math>5.69</b>	65.31 $\pm$ 9.21	<b>68.98<math>\pm</math>7.74</b>

TABLE II

COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %) OF STANDARD SL VERSUS SYMMETRIC SL ON THE BU DATABASE

technique	Original database		Enriched database	
	Standard	Symmetric	Standard	Symmetric
PCA	44.00 $\pm$ 1.88	<b>46.29<math>\pm</math>2.79</b>	<b>51.86<math>\pm</math>3.15</b>	51.71 $\pm$ 1.51
LDA	-	-	67.86 $\pm$ 6.42	<b>69.29<math>\pm</math>4.12</b>
CDA(km,2)	-	-	67.29 $\pm$ 4.65	<b>69.00<math>\pm</math>4.86</b>
CDA(fcm,2)	-	-	68.14 $\pm$ 5.54	<b>69.86<math>\pm</math>4.67</b>
CDA(km,3)	-	-	68.71 $\pm$ 6.78	<b>69.71<math>\pm</math>2.24</b>
CDA(fcm,3)	-	-	68.86 $\pm$ 6.12	<b>69.00<math>\pm</math>5.45</b>
PCA+LDA	64.86 $\pm$ 4.54	<b>66.43<math>\pm</math>4.69</b>	70.14 $\pm$ 5.79	<b>70.57<math>\pm</math>5.97</b>
PCA+CDA(km,2)	63.57 $\pm$ 3.87	<b>65.57<math>\pm</math>4.21</b>	<b>68.00<math>\pm</math>6.15</b>	67.86 $\pm$ 5.89
PCA+CDA(fcm,2)	63.43 $\pm$ 4.54	<b>65.71<math>\pm</math>5.19</b>	69.14 $\pm$ 6.42	<b>71.14<math>\pm</math>4.65</b>
PCA+CDA(km,3)	61.57 $\pm$ 2.94	<b>64.57<math>\pm</math>1.01</b>	69.29 $\pm$ 4.71	<b>69.57<math>\pm</math>4.51</b>
PCA+CDA(fcm,3)	62.29 $\pm$ 4.81	<b>65.43<math>\pm</math>4.11</b>	69.14 $\pm$ 5.76	<b>70.00<math>\pm</math>6.32</b>

TABLE III

COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %) OF STANDARD SL VERSUS SYMMETRIC SL ON THE JAFFE DATABASE

technique	Original database		Enriched database	
	Standard	Symmetric	Standard	Symmetric
PCA	38.10 $\pm$ 9.71	<b>43.81<math>\pm</math>11.21</b>	49.05 $\pm$ 11.75	<b>50.00<math>\pm</math>11.81</b>
LDA	-	-	52.86 $\pm$ 12.46	<b>55.24<math>\pm</math>12.25</b>
CDA(km,2)	-	-	49.52 $\pm$ 12.20	<b>51.90<math>\pm</math>11.10</b>
CDA(fcm,2)	-	-	48.57 $\pm$ 11.10	<b>52.38<math>\pm</math>12.39</b>
CDA(km,3)	-	-	46.19 $\pm$ 11.82	<b>51.43<math>\pm</math>11.54</b>
CDA(fcm,3)	-	-	48.10 $\pm$ 10.94	<b>52.86<math>\pm</math>11.12</b>
PCA+LDA	51.90 $\pm$ 11.72	<b>58.57<math>\pm</math>11.14</b>	52.38 $\pm$ 13.84	<b>60.48<math>\pm</math>11.71</b>
PCA+CDA(km,2)	52.86 $\pm$ 12.81	<b>56.67<math>\pm</math>9.36</b>	46.67 $\pm$ 14.79	<b>58.57<math>\pm</math>9.26</b>
PCA+CDA(fcm,2)	54.29 $\pm$ 12.41	<b>58.10<math>\pm</math>12.47</b>	50.95 $\pm$ 13.21	<b>61.43<math>\pm</math>8.37</b>
PCA+CDA(km,3)	53.81 $\pm$ 13.91	<b>57.14<math>\pm</math>12.45</b>	55.71 $\pm$ 11.41	<b>61.43<math>\pm</math>10.46</b>
PCA+CDA(fcm,3)	54.57 $\pm$ 12.21	<b>57.14<math>\pm</math>11.21</b>	54.29 $\pm$ 13.88	<b>61.43<math>\pm</math>9.06</b>

We observe that:

- In small training sets (original databases) symmetric techniques achieve much better performance than the standard ones. In larger training sets (enriched databases), the symmetric SL techniques also outperform the standard ones, but not at the same as in the original database case. This confirms our claim that the proposed symmetric extensions achieve a higher generalization in the cases where the training data do not represent the real class

structure by exploiting the knowledge that the problem is formed by symmetrical patterns.

- In databases with a small number of samples, the use of symmetry improves the performance of SL techniques to a larger extent compared to larger databases. Thus, when the training set is small (JAFFE), the standard techniques achieve poor generalization, while the corresponding symmetric techniques overcome this problem by using the a-priori knowledge of symmetry. On the other hand, on databases with a large training set, such as (BU), there is a better representation of samples. Therefore, the possibility that standard and symmetric techniques will achieve a similar degree of generalization is bigger.

In order to compare the generalization ability of the standard techniques and the corresponding symmetric extensions we used the FER-AIIA database. More specifically, the training set consists of all the centered and shifted facial images of the three databases (COHN-CANADE, BU and JAFFE), while the test set consists of the FER-AIIA facial images. In this experiment, only the NC classifier was used. The results are presented in Table IV.

TABLE IV

COMPARISON OF THE BEST RECOGNITION ACCURACIES (%) OF STANDARD SL VERSUS SYMMETRIC SL ON THE FER-AIIA DATABASE.

technique	Standard	Symmetric
PCA	22.71	<b>24.00</b>
LDA	31.86	<b>33.43</b>
CDA(km,2)	33.14	<b>35.14</b>
CDA(fcm,2)	32.14	<b>33.14</b>
CDA(km,3)	30.71	<b>32.29</b>
CDA(fcm,3)	31.71	<b>33.86</b>
CDA(km,4)	28.29	<b>30.14</b>
CDA(fcm,4)	30.86	<b>32.43</b>
PCA+LDA	34.29	<b>35.00</b>
PCA+CDA(km,2)	34.43	<b>35.86</b>
PCA+CDA(fcm,2)	33.86	<b>37.57</b>
PCA+CDA(km,3)	31.14	<b>35.29</b>
PCA+CDA(fcm,3)	33.00	<b>35.14</b>
PCA+CDA(km,4)	29.57	<b>36.71</b>
PCA+CDA(fcm,4)	33.43	<b>33.57</b>

At first, we observe that, in this experiment, the recognition accuracy is quite smaller than in previous experiments. This is due to the fact that different databases have been used for training and for testing. This means that there are different illumination conditions and different geometric transformations (scaling, rotation, translation) applied to facial images of the training and test sets. We notice that, in all the cases, the symmetric techniques achieve better recognition accuracies when compared to the standard ones. Therefore, we can conclude that under these conditions, the symmetric SL techniques achieve greater generalization on unknown data.

Finally, we conducted experiments with images rotated in random angles for assessing the proposed techniques in cases where symmetry axis is not known. That is, we constructed three new datasets based on the original data versions of the COHN-CANADE, BU and JAFFE databases by rotating each facial image in a random angle. An automatic symmetry detection algorithm [60] has been applied in order to find the symmetry axis and then the images have been normalized to



become vertically symmetric. The alternative scanning method presented in Subsection III-A has been applied for keeping symmetry to vector representation and, then, the proposed and the standard SL techniques have been applied. We used the 5-fold cross validation procedure to estimate the recognition accuracy. Some examples of rotated images and the corresponding ones after normalization using the symmetry detection algorithm are shown in Figure 11. The results are presented in Table V. As can be seen, the increase in the performance between the standard SL techniques and the proposed ones ranges from 0.95% to 6.20%, highlighting the superiority of the proposed techniques and the ability them to cope with the small symmetry detection errors that occur using the imposed symmetry constraint.

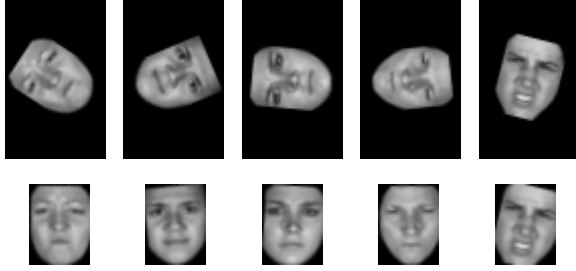


Fig. 11. Examples of rotated images (first row) and the corresponding ones (second row) after applying symmetry detection [60].

TABLE V  
COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %)  
OF STANDARD SL VERSUS SYMMETRIC SL ON THE ROTATED  
VERSIONS OF THE COHN-KANADE, BU AND JAFFE DATABASES AFTER  
SYMMETRY DETECTION

database	technique	Standard	Symmetric
COHN-KANADE	PCA	35.92 $\pm$ 3.85	<b>37.55<math>\pm</math>4.35</b>
	PCA+LDA	64.08 $\pm$ 8.79	<b>68.98<math>\pm</math>5.07</b>
	PCA+CDA(km,2)	58.78 $\pm$ 3.41	<b>63.67<math>\pm</math>5.25</b>
	PCA+CDA(fcm,2)	58.37 $\pm$ 3.99	<b>63.27<math>\pm</math>5.04</b>
	PCA+CDA(km,3)	55.92 $\pm$ 6.36	<b>60.00<math>\pm</math>7.71</b>
	PCA+CDA(fcm,3)	57.14 $\pm$ 5.79	<b>61.63<math>\pm</math>5.01</b>
BU	PCA	39.14 $\pm$ 4.77	<b>42.43<math>\pm</math>3.76</b>
	PCA+LDA	59.00 $\pm$ 6.14	<b>61.14<math>\pm</math>6.04</b>
	PCA+CDA(km,2)	58.00 $\pm$ 5.62	<b>59.86<math>\pm</math>5.39</b>
	PCA+CDA(fcm,2)	58.71 $\pm$ 3.12	<b>60.57<math>\pm</math>4.07</b>
	PCA+CDA(km,3)	57.57 $\pm$ 3.45	<b>59.14<math>\pm</math>3.55</b>
	PCA+CDA(fcm,3)	59.57 $\pm$ 3.67	<b>60.71<math>\pm</math>3.68</b>
JAFFE	PCA	38.57 $\pm$ 7.14	<b>39.52<math>\pm</math>5.35</b>
	PCA+LDA	39.52 $\pm$ 12.79	<b>41.90<math>\pm</math>7.04</b>
	PCA+CDA(km,2)	38.57 $\pm$ 15.54	<b>42.86<math>\pm</math>9.05</b>
	PCA+CDA(fcm,2)	38.57 $\pm$ 5.85	<b>44.77<math>\pm</math>6.02</b>
	PCA+CDA(km,3)	37.62 $\pm$ 12.12	<b>43.33<math>\pm</math>12.59</b>
	PCA+CDA(fcm,3)	41.43 $\pm$ 6.51	<b>43.33<math>\pm</math>7.12</b>

### B. Experiments on Face Recognition

We used the ORL [69], AR [59] and Extended YALE-B [70] face databases in our experiments for face recognition. The ORL database contains 400 images of 40 distinct persons (10 images each). The images were captured at different times and at different condition, including illumination conditions, facial expressions (smiling/not smiling) and facial details (open/closed eyes, with/without glasses). Furthermore, the

images were taken in frontal position with a tolerance for some tilting and rotation of the face of up to 20 degrees. All images are grayscale and normalized to a resolution of 112 $\times$ 92 pixels. The AR database contains over 4000 color images corresponding to 70 male and 56 female faces. The images were taken in frontal position with different facial expressions (anger, smiling and screaming), illumination conditions (left and/or right light on) and occlusions (sun glasses and scarf). Each person participated in two recording sessions, separated by two weeks (14 days) time. Each session contains 13 color images. The same images were taken in both sessions. In our experiments, we used a subset from the AR database, which contains cropped images from 100 persons (50 men and 50 women) [71]. The Extended YALE-B database contains images of 38 persons in 9 poses and under 64 illumination conditions. We used only the frontal cropped images [72]. All images were resized to 40  $\times$  30 pixels, in our experiments. Some example facial images from the ORL, the AR and the Extended YALE-B databases are displayed in Figures 12, 13 and 14, respectively.



Fig. 12. Sample images from the ORL database.



Fig. 13. Sample images from the AR database.



Fig. 14. Sample images from the Extended YALE-B database.

We conducted five series of experiments, for the above databases. In each experiment, the 10%, 20%, 30%, 40% and 50% of images per person were randomly selected for training. The remaining images were used for testing. For each experiment the standard PCA, PCA+LDA, PCA+CDA(km,2) and PCA+CDA(fcm,2) techniques, and the corresponding symmetric extensions were applied. The direct application of the DA techniques in all the databases was impossible because of the 'small sample size' problem. Also, for the ORL database, the application of the DA techniques was impossible, because of the fact that the training set size is quite small. The results obtained for the ORL, AR and Extended YALE-B databases, are illustrated in Tables VI, VII and VIII, respectively.

From these results, we can observe the following:

- For the ORL and the Extended YALE-B databases the symmetric SL techniques outperform the standard SL ones. Therefore, we can conclude that for symmetric

patterns (such as a human face), where the samples have symmetry error, either because the facial pose is not exactly frontal (ORL database case), or due to illumination variations (Extended YALE-B database), the proposed techniques achieve better data representation and are not affected by the symmetry noise of the images.

- In some cases (AR database) PCA and PCA+LDA outperformed SPCA and SPCA+SLDA. A possible explanation for this is that AR database comprises of facial images varying in facial expressions and occlusion. Such conditions do not effect image symmetry. Thus, the symmetric techniques can not achieve a better degree of generalization, in comparison to the standard ones.
- We can observe that SPCA+SCDA achieves better results when compared to PCA+CDA. This can be explained by the fact that some facial images varied in illumination conditions. Thus, the clustering technique in SPCA+SCDA results to better clusters than in the case of PCA+CDA. This can be observed in Figure 15, where the results of k-means based clustering are illustrated. In the case of SPCA+SCDA, the facial images are clustered in two clusters according to the illumination conditions, which is not the case for PCA+CDA. Therefore, SPA+SCDA can achieve better generalization by exploiting data symmetry.

### C. Experiments on Object Categorization

In the experiments on object categorization we used the ETH-80 [64] database. It contains images from eight categories: apple, pear, tomato, cow, horse, dog, cup and car. For

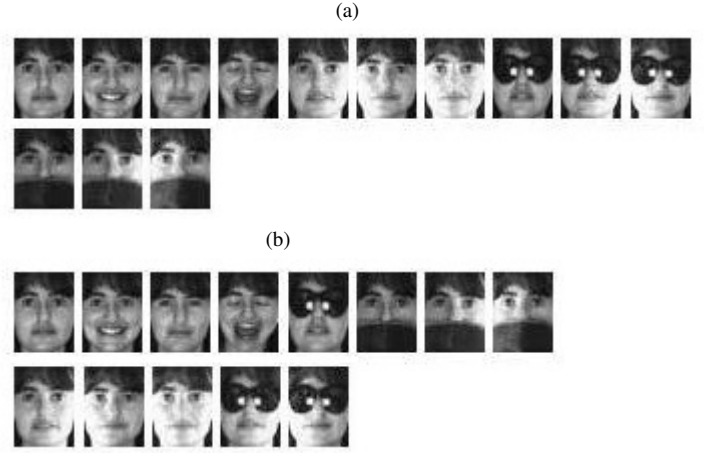


Fig. 15. Results of k-means clustering in (a) PCA+LDA, and (b) SPCA+SCDA.

each category there are images of ten different objects. Each object has been captured by 41 different views. In our experiments we used the version, in which entire objects appear. The images were resized to  $32 \times 32$  pixels. In our experiments we have used the object categories that correspond to symmetric objects (apple, cup, pear, tomato). Examples for each category are shown in Figure 16.

We evaluated the performance of the proposed techniques using the 5-fold cross validation procedure. Specifically, images of each object were either in the training set or the test set. The results are shown in Table IX. As can be seen, the symmetric techniques outperform the corresponding standard

TABLE VI  
COMPARISON OF THE BEST RECOGNITION ACCURACIES (%) OF STANDARD SL VERSUS SYMMETRIC SL ON THE ORL DATABASE

Training set technique	10%		20%		30%		40%		50%	
	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric
PCA	71.67	<b>73.33</b>	81.88	<b>85.00</b>	85.36	<b>89.29</b>	88.75	<b>90.83</b>	90.50	<b>92.00</b>
PCA+LDA	-	-	80.94	<b>87.81</b>	85.36	<b>91.07</b>	88.33	<b>90.83</b>	88.50	<b>93.50</b>
PCA+CDA(km,2)	-	-	-	-	85.00	<b>93.57</b>	85.83	<b>92.50</b>	85.00	<b>95.00</b>
PCA+CDA(fcm,2)	-	-	-	-	85.00	<b>93.21</b>	84.17	<b>92.92</b>	87.00	<b>95.00</b>

TABLE VII  
COMPARISON OF THE BEST RECOGNITION ACCURACIES (%) OF STANDARD SL VERSUS SYMMETRIC SL ON THE AR DATABASE

Training set technique	10%		20%		30%		40%		50%	
	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric
PCA	23.91	<b>24.13</b>	27.81	<b>32.38</b>	<b>39.89</b>	39.67	44.69	<b>45.38</b>	59.54	<b>60.15</b>
PCA+LDA	<b>48.17</b>	47.39	48.86	<b>52.52</b>	<b>65.94</b>	65.17	<b>60.94</b>	60.88	80.85	<b>81.54</b>
PCA+CDA(km,2)	41.00	<b>45.91</b>	49.62	<b>52.81</b>	64.39	<b>65.61</b>	59.31	<b>62.94</b>	72.46	<b>82.08</b>
PCA+CDA(fcm,2)	41.00	<b>44.43</b>	49.24	<b>52.33</b>	63.89	63.89	59.44	<b>60.56</b>	79.62	<b>82.23</b>

TABLE VIII  
COMPARISON OF THE BEST RECOGNITION ACCURACIES (%) OF STANDARD SL VERSUS SYMMETRIC SL ON THE EXTENDED YALE-B DATABASE

Training set technique	10%		20%		30%		40%		50%	
	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric	Standard	Symmetric
PCA	41.15	<b>47.23</b>	55.06	<b>60.53</b>	60.58	<b>63.45</b>	67.59	<b>73.06</b>	74.01	<b>82.73</b>
PCA+LDA	68.83	<b>73.23</b>	81.53	<b>84.00</b>	85.20	<b>86.02</b>	91.14	<b>92.24</b>	94.82	<b>95.64</b>
PCA+CDA(km,2)	71.82	<b>76.27</b>	79.98	<b>83.80</b>	81.75	<b>85.73</b>	90.10	<b>92.17</b>	94.24	<b>95.39</b>
PCA+CDA(fcm,2)	71.19	<b>77.00</b>	81.53	<b>84.00</b>	83.68	<b>85.96</b>	90.79	<b>92.04</b>	94.16	<b>95.48</b>

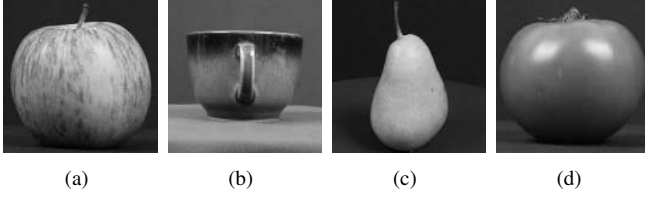


Fig. 16. Example objects from the ETH-80 database: (a) apple, (b) cup, (c) pear, and (d) tomato.

ones. Indeed, a large improvement in recognition accuracy is observed when symmetry constraints are exploited. For example, in the CDA(km,2) case the improvement is about 7.68%.

TABLE IX

COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %) OF STANDARD SL VERSUS SYMMETRIC SL ON THE ETH-80 DATABASE.

technique	Standard	Symmetric
PCA	85.43 $\pm$ 8.56	<b>86.22<math>\pm</math>8.42</b>
LDA	74.88 $\pm$ 3.84	<b>81.52<math>\pm</math>5.18</b>
CDA(km,2)	75.67 $\pm$ 6.87	<b>83.35<math>\pm</math>7.12</b>
CDA(fcm,2)	76.52 $\pm$ 5.89	<b>83.23<math>\pm</math>6.47</b>
CDA(km,3)	76.34 $\pm$ 4.41	<b>83.66<math>\pm</math>6.28</b>
CDA(fcm,3)	76.46 $\pm$ 5.49	<b>83.41<math>\pm</math>6.66</b>
CDA(km,4)	76.95 $\pm$ 7.01	<b>83.84<math>\pm</math>8.02</b>
CDA(fcm,4)	76.34 $\pm$ 5.87	<b>82.68<math>\pm</math>7.26</b>
PCA+LDA	85.00 $\pm$ 5.62	<b>87.20<math>\pm</math>6.59</b>
PCA+CDA(km,2)	85.73 $\pm$ 9.28	<b>87.50<math>\pm</math>8.84</b>
PCA+CDA(fcm,2)	85.98 $\pm$ 8.92	<b>87.38<math>\pm</math>9.04</b>
PCA+CDA(km,3)	86.52 $\pm$ 8.89	<b>88.00<math>\pm</math>9.74</b>
PCA+CDA(fcm,3)	86.59 $\pm$ 8.47	<b>88.05<math>\pm</math>9.01</b>
PCA+CDA(km,4)	85.98 $\pm$ 8.84	<b>88.11<math>\pm</math>9.41</b>
PCA+CDA(fcm,4)	86.40 $\pm$ 7.91	<b>87.93<math>\pm</math>7.29</b>

In order to examine the performance of the proposed SL techniques on asymmetric data, we used the entire ETH-80 dataset which has both symmetric and asymmetric objects and also multiple views from which some objects are not symmetric, as shown in Figure 17. The 5-fold cross validation procedure is used for evaluation. In this case, the optimal value of  $s$  has been learned from the training set for each fold using cross-validation and this value was used in testing. In our experiments, the optimal value of  $s$  found was 0.7 on average. For fully asymmetric datasets, it is expected that the optimal value of  $s$  learned during training using cross-validation to be close to zero. The results, as shown in Table X, illustrate that the proposed techniques are able to cope with partial symmetry giving improved performance. Indeed, the performance of the proposed approaches is about 0.5%-5.5% better than the standard SL techniques. Therefore, we can notice that in some cases the symmetry constraint can act as a general regularizer that prevents overfitting even for asymmetric objects.



Fig. 17. Example objects of the category "car" from the ETH-80 database.

TABLE X

COMPARISON OF THE BEST RECOGNITION ACCURACIES (MEAN  $\pm$  STD %) OF STANDARD SL VERSUS SYMMETRIC SL ON THE ENTIRE ETH-80 DATABASE.

technique	Standard	Symmetric
PCA	71.80 $\pm$ 5.22	<b>74.36<math>\pm</math>5.94</b>
LDA	67.04 $\pm$ 5.78	<b>70.40<math>\pm</math>6.48</b>
CDA(km,2)	68.93 $\pm$ 5.28	<b>71.95<math>\pm</math>5.01</b>
CDA(fcm,2)	68.63 $\pm$ 4.66	<b>72.32<math>\pm</math>4.09</b>
CDA(km,3)	69.54 $\pm$ 5.98	<b>72.77<math>\pm</math>5.26</b>
CDA(fcm,3)	68.51 $\pm$ 4.62	<b>72.17<math>\pm</math>4.88</b>
PCA+LDA	69.48 $\pm$ 7.34	<b>69.76<math>\pm</math>6.12</b>
PCA+CDA(km,2)	71.46 $\pm$ 5.82	<b>74.06<math>\pm</math>5.95</b>
PCA+CDA(fcm,2)	72.29 $\pm$ 5.22	<b>73.96<math>\pm</math>5.33</b>
PCA+CDA(km,3)	71.71 $\pm$ 6.49	<b>76.04<math>\pm</math>6.61</b>
PCA+CDA(fcm,3)	71.22 $\pm$ 5.95	<b>76.77<math>\pm</math>5.12</b>

#### D. The Effect of Symmetry Factor

The symmetry factor  $s$  is a critical parameter in symmetric SL techniques. In this subsection, we show the effect of the symmetry factor in the performance of symmetric SL techniques, the symmetry error of generated projection vectors and the subspace dimensionality in the case of SPCA. For this reason, we performed a series of experiments on the facial expression recognition databases. Specifically, we applied the SPCA, SLDA, SCDA(km,2) and SCDA(fcm,2) techniques in the enriched versions of the COHN-KANADE, BU and JAFFE databases. In these experiments, we did not use all 24 shifted images for each centered image, but we randomly selected 12 shifted images. We used enriched databases to overcome the small sample size problem. On the other hand, we did not use a large number of shifted images to show the importance of the symmetry factor. In all the experiments, we used the 5-fold cross validation procedure.

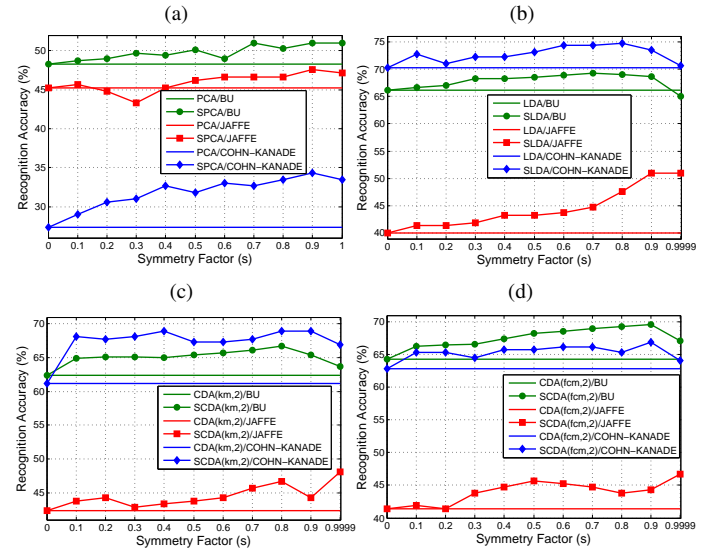


Fig. 18. The recognition accuracy of (a) SPCA, (b) SLDA, (c) SCDA(km,2), and (d) SCDA(fcm,2) techniques, on COHN-KANADE, BU and JAFFE databases.

The performance of SPCA, SLDA, SCDA(km,2) and SCDA(fcm,2) techniques and the comparison with the corresponding standard SL techniques are displayed in Figure

18. In SPCA and DA techniques we used the kNN and the NC classifiers, respectively. It is easy to see that the most appropriate region of  $s$  is between 0.6 and 0.9 for the COHN-KANADE and the BU databases, while in the case of JAFFE database it is between 0.9 and 1.0.

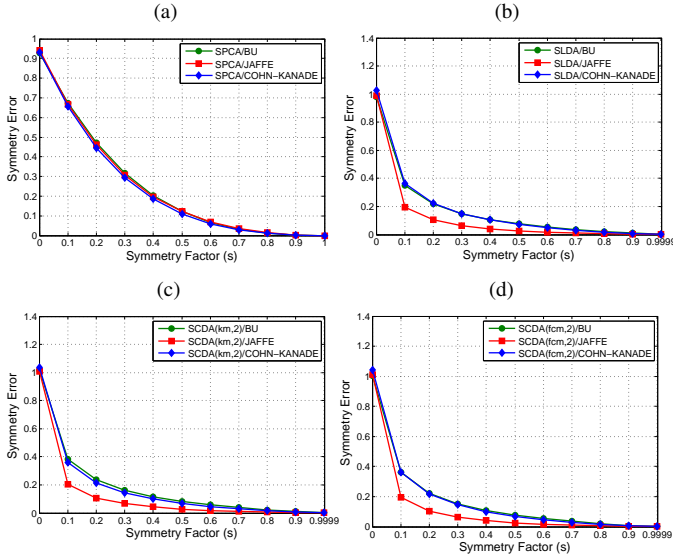


Fig. 19. The projection vectors symmetry error of (a) SPCA, (b) SLDA, (c) SCDA(km,2), and (d) SCDA(fcm,2) techniques, on COHN-KANADE, BU and JAFFE databases.

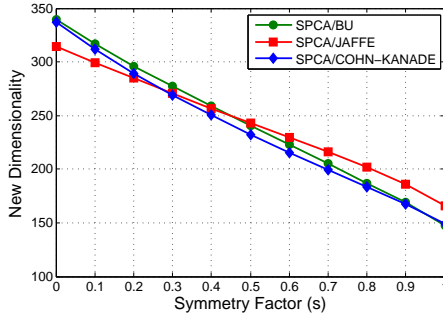


Fig. 20. The new dimensionality of SPCA, on COHN-KANADE, BU and JAFFE databases.

Figure 19 illustrates the average symmetry error of the projection vectors. As expected, the symmetry error of the generated projection vectors decreases as the factor increases. Figure 20 illustrates the new dimensionality of SPCA by maintaining the 99% of the total eigenvalue sum of the training set. As can be seen, a better dimensionality reduction is achieved as the factor increases. For example, the 99% of the total eigenvalue sum corresponds to maintaining 320 eigenvectors in the case of PCA for the COHN-KANADE database, whereas SPCA ( $s=1$ ) needs 150 eigenvectors only, respectively. Therefore, in the SPCA ( $s=1$ ) case a compression of about 50% is achieved.

Figure 21 shows the reconstructed versions of test images using the SPCA projection vectors of the training images. We clearly see that the reconstructed versions of the images are more symmetric as the symmetry factor increases, resulting

in a better image representation by keeping the main characteristics of the face. Specifically, SPCA corrects the image registration errors and the direction of eye gaze in the first two images and discards the effects of illumination variations in the next two ones.

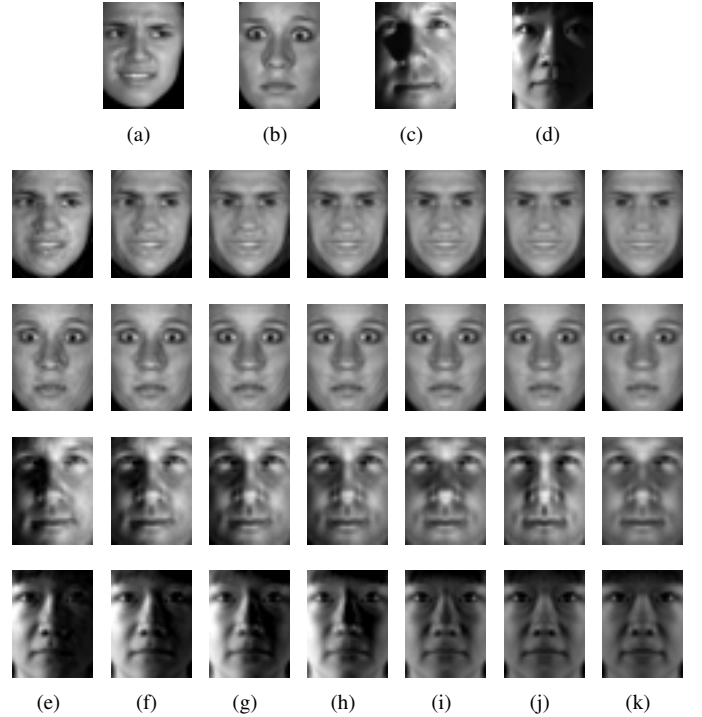


Fig. 21. Original images from the BU ((a),(b)) and YALE-B ((c),(d)) databases (first row) and their reconstructed versions (next rows) using the projection vectors determined by SPCA for  $s$ : (e) 0.0, (f) 0.2, (g) 0.4, (h) 0.5, (j) 0.6, (i) 0.8, and (k) 1.0.



Fig. 22. PCA (first row) and SPCA (second row) Eigenfaces from the ORL database.

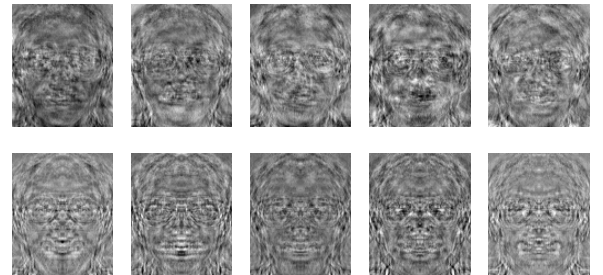


Fig. 23. PCA+LDA (first row) and SPCA+SLDA (second row) Fisherfaces from the ORL database.

Finally, Figure 22 shows the Eigenfaces obtained for the ORL database. The first row illustrates Eigenfaces generated by applying PCA technique, while the second row contains Eigenfaces generated by applying the symmetric extension of PCA. Figure 23 shows the corresponding Fisherfaces obtained for the ORL database. It is obvious that the use of symmetric constraints leads to the correction of the pattern symmetry.

#### E. A different usage of symmetry

Until now, we used symmetry to modify the objective functions of SL techniques and to produce symmetric projection vectors. Alternatively, we can use this property to produce symmetric projection vectors, by doubling the training set [73]: for each sample, we add its inverted version. Thus, the standard SL techniques will be able to learn the symmetry errors of the training samples.

To test this approach, we created inverted versions for COHN-KANADE, BU and JAFFE databases and we applied the PCA, PCA+LDA and PCA+CDA(km,2) techniques. We used the 5-cross validation procedure and the same partition in training and test sets, used in Subsection IV-A, so that the comparison of the results will be valid. The obtained results are displayed in Figure 24 and are compared with the corresponding results of Subsection IV-A. We can observe that in some cases the application of standard SL techniques in the inverted versions achieves better results than for the original databases. Thus, the hypothesis that a better generalization can be achieved is confirmed. In addition, comparing these results with the results of the corresponding symmetric techniques leads us to conclude that bigger generalization is achieved with the proposed use of symmetry constraints, than with the standard techniques plus image inversion.

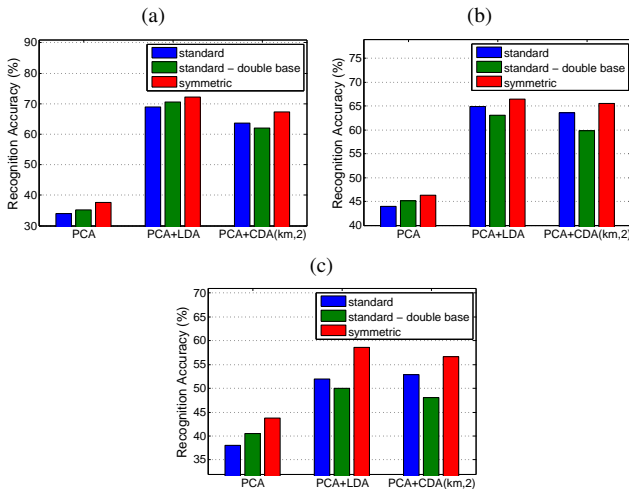


Fig. 24. Recognition accuracy on (a) COHN-KANADE, (b) BU, and (c) JAFFE databases.

Figure 25 also illustrates the corresponding average symmetry error of obtained projection vectors. It can be easily observed that the symmetry error of the projection vectors always decreases, as in the symmetric extensions of SL techniques. Thus, the enrichment of the training set can produce symmetric projection vectors. However, they are not appropriate for good discrimination in symmetric databases, as described above.

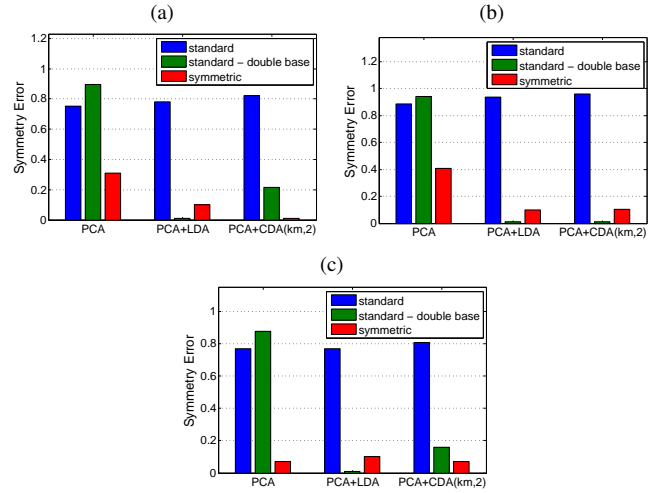


Fig. 25. The projection vectors symmetry error on (a) COHN-KANADE, (b) BU, and (c) JAFFE databases.

## V. CONCLUSION

In this paper, we studied whether and how image symmetry property can be exploited by SL techniques to increase their performance. We showed that this is feasible by modifying the objective functions of SL techniques with the introduction of symmetry constraints. For this purpose, we proposed extensions for PCA, LDA and CDA techniques which use symmetry constraints in order to determine symmetric projection vectors. The application of these techniques on artificial data belonging to symmetric patterns and having high symmetry error showed that they generate better projection vectors than standard SL techniques. Similar results were obtained by applying the proposed techniques on real datasets for facial expression and face recognition and object categorization. The above results lead to the conclusion that the proposed techniques improve the robustness and the generalization ability of standard SL techniques, even if the training set is small or the samples contain a big symmetry error measure. Similar extensions could be exploited for other SL and classification techniques.

## ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IM-PART). This publication reflects only the authors views. The European Union is not liable for any use that may be made of the information contained therein.

## REFERENCES

- [1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press Professional, 1990.
- [2] A. Jain, R. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [3] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.
- [4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [5] T.-W. Lee, *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, 1998.



- [6] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [7] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. MIT Press, 2003, pp. 153–160.
- [8] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Orthogonal locality minimizing globality maximizing projections for feature extraction," *Optical Engineering*, vol. 48, no. 1, pp. 017 202–017 202–5, 2009.
- [9] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [10] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [11] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 7, pp. 179–188, 1936.
- [12] P. Belhumeur, P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [13] A. Iosifidis, A. Tefas, and I. Pitas, "On the optimal class representation in linear discriminant analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1491–1497, 2013.
- [14] M. Kyperountas, A. Tefas, and I. Pitas, "Dynamic training using multistage clustering for face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 894 – 905, 2008.
- [15] —, "Salient feature and reliable classifier selection for facial expression classification," *Pattern Recognition*, vol. 43, no. 3, pp. 972 – 986, 2010.
- [16] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in non-negative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.
- [17] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image and Vision Computing*, vol. 24, no. 3, pp. 239–248, 2006.
- [18] Y. Koren and L. Carmel, "Robust linear dimensionality reduction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 4, pp. 459–470, 2004.
- [19] X.-W. Chen and T. Huang, "Facial expression recognition: a clustering-based approach," *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1295–1302, 2003.
- [20] M. Zhu and A. Martínez, "Subclass discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274–1286, 2006.
- [21] O. Zoidi, A. Tefas, and I. Pitas, "Multiplicative update rules for concurrent nonnegative matrix factorization and maximum margin classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 3, pp. 422–434, 2013.
- [22] S. Nikitidis, A. Tefas, and I. Pitas, "Maximum margin projection subspace learning for visual data analysis," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4413–4425, 2014.
- [23] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.
- [24] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212–219, 2007.
- [25] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Extracting the optimal dimensionality for local tensor discriminant analysis," *Pattern Recognition*, vol. 42, no. 1, pp. 105 – 114, 2009.
- [26] J. Yang, D. Zhang, A. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.
- [27] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems 17*, L. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005, pp. 1569–1576.
- [28] S. Chen, H. Zhao, M. Kong, and B. Luo, "2D-LPP: A two-dimensional extension of locality preserving projections," *Neurocomputing*, vol. 70, no. 4–6, pp. 912–921, 2007.
- [29] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2007, pp. 1–7.
- [30] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *IEEE 11th International Conference on Computer Vision*. IEEE Computer Society, 2007, pp. 1–7.
- [31] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proceedings of the 10th European Conference on Computer Vision*. Springer-Verlag, 2008, pp. 725–738.
- [32] J. Yang, S. Yan, and T. Huang, "Ubiquitously supervised subspace learning," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 241–249, 2009.
- [33] F. Nie, D. Xu, I. Tsang, and C. Zhang, "Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1921–1932, 2010.
- [34] B.-K. Bao, G. Liu, R. Hong, S. Yan, and C. Xu, "General subspace learning with corrupted training data via graph embedding," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4380–4393, 2013.
- [35] F. Torre and M. Black, "Robust principal component analysis for computer vision," in *8th IEEE International Conference on Computer Vision*. IEEE Computer Society, 2001, pp. 362–369.
- [36] F. Nie, S. Xiang, and C. Zhang, "Neighborhood minmax projections," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 2007, pp. 993–998.
- [37] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, 2011.
- [38] F. Nie, S. Xiang, Y. Liu, C. Hou, and C. Zhang, "Orthogonal vs. uncorrelated least squares discriminant analysis for feature extraction," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 485 – 491, 2012.
- [39] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [40] D. Washburn and D. Crowe, *Symmetries of culture: Theory and practice of plane pattern analysis*. University of Washington Press, 1988.
- [41] G. Bluman and S. Anco, *Symmetry and Integration Methods for Differential Equations*. Springer-Verlag, YEAR = 2002.
- [42] G. Smyth, "Employing symmetry constraints for improved frequency estimation by eigenanalysis methods," *Technometrics*, vol. 42, no. 3, pp. 277–289, 2000.
- [43] E. Malavasi, E. Charbon, E. Felt, and A. Sangiovanni-Vincentelli, "Automation of IC layout with analog constraints," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, no. 8, pp. 923–942, 1996.
- [44] F. Balasa and K. Lampaert, "Symmetry within the sequence-pair representation in the context of placement for analog design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 7, pp. 721–731, 2000.
- [45] R. Backofen and S. Will, "Symmetry-breaking predicates for search problems," in *In Proceedings of the 5th International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann, 1996, pp. 148–159.
- [46] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context-free attentional operators: the generalized symmetry transform," *International Journal of Computer Vision*, vol. 14, no. 2, pp. 119–130, 1995.
- [47] J. Hayfron-Acquah, M. Nixon, and J. Carter, "Automatic gait recognition by symmetry analysis," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2175–2183, 2003.
- [48] S. Thrun and B. Wegbreit, "Shape from symmetry," in *Proceedings of the 10th IEEE International Conference on Computer Vision 2*, N. Sebe, S. Michael, and S. Thomas, Eds. IEEE Computer Society, 2005, pp. 1824–1831.
- [49] L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, and C. Lockwood, "3D face authentication and recognition based on bilateral symmetry analysis," *The Visual Computer*, vol. 22, no. 1, pp. 43–55, 2006.
- [50] Y. Liu, K. Schmidt, J. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *Computer Vision and Image Understanding*, vol. 91, no. 1–2, pp. 138–159, 2003.
- [51] S. Mitra and Y. Liu, "Local facial asymmetry for expression classification," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, D. Jacobs, Ed. IEEE Computer Society, 2004, pp. 889–894.
- [52] E. Saber and A. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 669–680, 1998.
- [53] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, 2001.
- [54] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2004.



- [55] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [56] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 729–735, 2009.
- [57] C. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley-Interscience, 2002.
- [58] D. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.
- [59] A. Martínez and R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998.
- [60] G. Loy and J.-O. Eklundh, "Detecting symmetry and symmetric constellations of features," in *Proceedings of the 9th European Conference on Computer Vision*. Springer-Verlag, 2006, pp. 508–521.
- [61] H. Cornelius and G. Loy, "Detecting bilateral symmetry in perspective," in *Conference on Computer Vision and Pattern Recognition*, 2006, pp. 191–191.
- [62] S. Tsogkas and I. Kokkinos, "Learning-based symmetry detection in natural images," in *European Conference on Computer Vision*, 2012, pp. 41–54.
- [63] F. Nie, S. Xiang, Y. Jia, and C. Zhang, "Semi-supervised orthogonal discriminant analysis via label propagation," *Pattern Recognition*, vol. 42, no. 11, pp. 2615 – 2627, 2009.
- [64] B. Leibe and B. Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2003, pp. 409–415.
- [65] T. Kanade, Y. Tian, and J. Cohn, "Comprehensive database for facial expression analysis," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2000, pp. 46–53.
- [66] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2006, pp. 211–216.
- [67] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proceedings of the 3rd International Conference on Face and Gesture Recognition*. IEEE Computer Society, 1998, pp. 200–205.
- [68] A. Maronidis, D. Bolis, A. Tefas, and I. Pitas, "Improving subspace learning for facial expression recognition using person dependent and geometrically enriched training sets," *Neural Networks*, vol. 24, no. 8, pp. 814–823, 2011.
- [69] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*. IEEE Computer Society, 1994, pp. 138–142.
- [70] A. Georgiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
- [71] A. Martínez and A. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [72] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.
- [73] K. Papachristou, A. Tefas, and I. Pitas, "Subspace learning with enriched databases using symmetry," in *Intelligent Data analysis and its Applications, Volume I*, ser. Advances in Intelligent Systems and Computing, J.-S. Pan, V. Snasel, E. S. Corchado, A. Abraham, and S.-L. Wang, Eds., vol. 297. Springer International Publishing, 2014, pp. 113–122.



**Konstantinos Papachristou** received the B.Sc. degree in informatics and the M.Sc. degree in digital media in 2009 and 2011 respectively, both from the Aristotle University of Thessaloniki, Greece. Since 2009 he has been a researcher and developer at the Artificial Intelligence and Information Analysis Lab of the same Department. He has participated in 4 research projects financed by European funds. He has co-authored 6 papers in international journals and conferences. His research interests are image/video processing, computer vision and pattern recognition.



**Anastasios Tefas** (S97-M04) received the B.Sc. in informatics in 1997 and the Ph.D. degree in informatics in 2002, both from the Aristotle University of Thessaloniki, Greece. Since 2013 he has been an Assistant Professor at the Department of Informatics, Aristotle University of Thessaloniki. From 2008 to 2012, he was a Lecturer at the same University. From 2006 to 2008, he was an Assistant Professor at the Department of Information Management, Technological Institute of Kavala. From 2003 to 2004, he was a temporary lecturer in the Department

of Informatics, University of Thessaloniki. From 1997 to 2002, he was a researcher and teaching assistant in the Department of Informatics, University of Thessaloniki. Dr. Tefas participated in 12 research projects financed by national and European funds. He has co-authored 46 journal papers, 130 papers in international conferences and contributed 7 chapters to edited books in his area of expertise. Over 2320 citations have been recorded to his publications and his H-index is 25 according to Google scholar. His current research interests include computational intelligence, pattern recognition, statistical machine learning, digital signal and image processing and computer vision.



**Ioannis Pitas** (SM94-F07) received the Diploma and PhD degree in Electrical Engineering, both from the Aristotle University of Thessaloniki, Greece. Since 1994, he has been a Professor at the Department of Informatics of the same University. He served as a Visiting Professor at several Universities. His current interests are in the areas of image/video processing, intelligent digital media, machine learning, human centered interfaces, affective computing, computer vision, 3D imaging and biomedical imaging. He has published over 750 papers, contributed in 39 books

in his areas of interest and edited or (co-)authored another 9 books. He has also been an invited speaker and/or member of the program committee of many scientific conferences and workshops. In the past he served as Associate Editor or co-Editor of eight international journals and General or Technical Chair of four international conferences (including ICIP2001). He participated in 68 R&D projects, primarily funded by the European Union and is/was principal investigator/researcher in 40 such projects. He has 19000+ citations (Source Publish and Perish), 6900+ (Scopus) to his work and h-index 66+ (Source Publish and Perish), 42+ (Scopus).