

Inverted Sparse Discriminant Preserving Projection for Face Recognition

Kiril Kirilov

kiril.st.kirilov@gmail.com

*Faculty of Economics and Business Administration
Sofia University "St. Kliment Ohridski"*

Abstract

Image classification and face recognition has been a popular subject matter for the last several decades. Images are usually handled as transformed as vectors which makes their classification a dimensionality reduction task. Some of the well-known algorithms in the area, such as the Sparsity Preserving Projection (SPP), create new theoretical concepts families, while other successfully modify or combine useful properties of the former ones. Compiled algorithms like Sparse Discriminant Preserving Projections (SDPP) employ the properties of the Sparse Representation (SR) as in their objective functions they include a supervised modification of the sparse weight matrix that considers the intra-class relations. By examining the construction of the SDPP algorithm and by providing some arguments on the supervised SR, in this paper we propose a new subspace learning algorithm, called Inverted Sparse Discriminant Preserving Projection (ISDPP). Likewise SDPP, ISDPP integrates supervised SR with the Fisher's criterion. In contrast to SDPP, ISDPP incorporates a *between-class* SR with the Fischer's within-class scatter matrix. A preliminary round of experiments support the initiative and provide an expectation for possible superior performance of the proposed ISDPP that is confirmed in the next round of empirical examinations.

Keywords: Dimensionality reduction, Sparse Representation, Face Recognition

1. Introduction

Image classification and face recognition have received a lot of attention for the last several decades already. The interest in it stems from its wide practical applicability in numerous fields in science and life like medical research, public safety, space science, documents manipulations and many others. The images' pixels matrices, converted as vectors – one of the main approach of handling images objects – are classical examples of high-dimensional data objects. In addition, the improving technical devises produce more and more qualitative and sizable images. This makes the task of image classification a dimensionality reduction problem, which is the essence of the subject of the algorithms in the area, prior to a final classification with a standard classifier. As a differentiated machine learning subject matter, the dimensionality reduction (DR) includes numerous theoretical concepts and algorithms that are applied for the solution of various problems in both unsupervised and supervised context. In general, the DR does simultaneously 1) work for the detection

of a new valuable variables-subspace that is much smaller than the original one, and 2) dramatically reduce the required computational recourses.

Since the images are large and complex data objects for a DR, the problem resolution has many aspects. Therefore, the numerous methods in the stream can be analysed and classified by some of their main characteristics (properties) and/or by the data structure they consider, like the following proposed: A) globality-locality reflection (representation) B) linearity-non-linearity C) presence (requirement) of parameter(s) D) unsupervised-supervised E) having the small-sample-size problem (SSS) F) having the out-of-sample problem.

Principal Component Analysis (PCA) [1]-[4] and Linear Discriminant Analysis (LDA) [3],[4] are fundamental bases of the DR subject matter. They have natural intuitive concepts that are a ground of the composition of many of the following DR algorithms. PCA and LDA are based on the Euclidian distances and reflect the global structure of the data. They have immediate and quite simple application. PCA, however, does not consider the input data label information, which may lead to a loss of a critical pattern information. Contrarily, LDA takes into account the classes labels, which makes it more effective than PCA in many applications. Unfortunately, LDA cannot be applied directly to a data with the so called small sample size problem (SSS) when the number of the features is greater than the training observations – something that is very common in the image classification). To address this, Belhumeur et al. [3] propose a two stage PCA+LDA method, in which PCA is exploited to make the within-class scatter matrix nonsingular. However, this causes a loss of an important discriminative information. On the other side, the proposed by Li et al. [5] maximal margin criterion (MMC) overcome the SSS more efficiently.

All the above methods capture the global Euclidian structure of the images and cannot characterize their local structure. To address this many manifold learning algorithms have been proposed based on the idea that the data points are samples from a low-dimensional manifold embedded in a high-dimensional space. Some representative algorithms are locally linear embedding (LLE) [6], Isometric Feature Mapping (ISOMAP) [7], Laplacian eigenmaps (LE) [8], etc, whereas each of them attempt to preserve a different geometrical property of the underlying manifold. Unfortunately all these methods suffer from the so called out of sample problem, consisting in the difficulty to find new sample images in the embedding space by utilizing the low-embedding results of the training data set.

To overcome this, some improved methods like the introduced by He et al. Neighbourhood preserving embedding (NPE) [9] and Locality preserving projection (LPP) [10] have been proposed. NPE and LPP work with Euclidian distances as well, which makes them sensitive to data noise and outliers. To address this, Yu et al. [11] present the ILPP-L1 that is based on the l_1 -norm-distances reaching a better robustness towards undesired outliers' impact on the cost of a bit more complicated subspace extraction, whereas the new space's vectors are extracted one by one consecutively. As NPE and LPP consider only the local structure of the data, some extended methods like the presented by Yu et al. Discriminant Locality Preserving Projections (DLPP) [12] integrate "global" with "locality" information by taking into account the data labels. However, DLPP suffers from the small sample size problem

as well. To overcome this Lu et al. [13] propose the DLPP/MMC that embed the concept of the maximal margin criterion (MMC) and thus elegantly eliminates the SSS.

The introduced by Qiao et al. Sparsity Preserving Projection (SPP) [14], sets the creation of a very attractive DR theory family. SPP seeks for a projection that preserves the sparse reconstruction relationship of the data points. Its main idea is imbedded in the computation of a *sparse reconstruction weight matrix* (also shortened as *sparse weight matrix*), whereat each row of the matrix is the observation x_i presented as a reconstruction from the remaining data points with the requirement for *sparsity*.

A lot of DR algorithms combine verified beneficial properties and techniques of different models. For instance, the Discriminant Sparse Neighbourhood Preserving Embedding (DSNPE) of Gui et al. [15] compiles parts of SPP and MMC. DSNPE tries to simultaneously preserve the sparsity property and to maximise the margin between the classes, whereas the sparse weight matrix is calculated within the classes. Similarly, the Sparse Locality Preserving Discriminative Projections (SLPDP), proposed by Zhang et al. [16], assembles LPP with the MMC and aims preservation of the neighbourhood together with a between-classes margin maximisation, whereas, in the computation of the Laplacian matrix, the adjacency weight matrix is replaced by the within-class sparse weight matrix. On the other hand, the Sparse Discriminant Preserving Projections (SDPP), suggested by Zhang et al. [17], employs the relationship of the within-class sparse representation, to the between-class scatter matrix, seen is a carrier of the global structure of the data.

The remainder of this paper is organized as follows. Section 2 introduces related works. The proposed Inverted Sparse Discriminant Preserving Projection algorithm is introduced in section 3. We show our experimental results in section 4 and the paper is concluded in section 5.

2. Related works

2.1 Linear Discriminant Analysis (LDA)

The objective function of LDA is defined as:

$$\max_w \frac{w^T S_B w}{w^T S_W w}$$

where:

$$S_B = \sum_{i=1}^C n_i (m^{(i)} - m)(m^{(i)} - m)^T \tag{1}$$

$$S_W = \sum_{i=1}^C n_i \sum_{j=1}^{n_i} n_i (x_j^{(i)} - m)(x_j^{(i)} - m)^T \tag{2}$$

The \mathbf{S}_B and \mathbf{S}_w are called between-class scatter and within-class scatter matrices, respectively, $m^{(i)}$ is the mean vector of class i , m is the total mean of all observations and n_i is the number of observations from class i .

2.2 Sparsity preserving projection (SPP)

Sparsity Preserving Projection (SPP) belongs to the generalized graph embedding framework [17]. It aims to preserve the sparse reconstruction relationship of the data [16].

Given a data set of n samples $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$, SPP seeks for a reconstruction of each sample x_i by as few as possible from the remaining observations from \mathbf{X} , or:

$$\min \|s_i\|_0, \text{ s.t. } x_i = \mathbf{X}s_i, \mathbf{1} = \mathbf{1}^T s_i \quad (3)$$

$\|\cdot\|_0$ is the l_0 -norm, denoting the number of non-zeros in the coefficients' vector, $s_i = [s_{i1}, s_{i2}, \dots, s_{i-1}, 0, s_{i+1}, \dots, s_{in}]^T, s_i \in \mathbb{R}^n$, where its i -th coefficient equals zero as an observation can not be presented by itself and $\mathbf{1}^T \in \mathbb{R}^n$ is a vector of all ones. Unfortunately Eq. (3) is NP-hard to solve as the function is not convex [14]. This difficulty is overcome as the optimization is translated into a l_1 -minimisation problem or:

$$\min \|s_i\|_1, \text{ s.t. } x_i = \mathbf{X}s_i, \mathbf{1} = \mathbf{1}^T s_i \quad (4)$$

Further, in many practical problems when the data \mathbf{X} is noisy, the constraint $x_i = \mathbf{X}s_i$ does not always hold [14]. Therefore, Qiao et al. [14] propose two stable extensions of the optimization. One of them is to relax the constraint $\mathbf{x} = \mathbf{X}\mathbf{s}$ as $\|\mathbf{x} - \mathbf{X}\mathbf{s}\| < \varepsilon$, where ε can be viewed as an error tolerance. Thus Eq. (4) is modified as:

$$\min \|s_i\|_1, \text{ s.t. } \|\mathbf{x} - \mathbf{X}\mathbf{s}\| < \varepsilon, \mathbf{1} = \mathbf{1}^T s_i \quad (5)$$

The objection function of SPP seeks to best preserve each optimal vector \mathbf{s}_i by:

$$\min_w \sum_{i=1}^n \|w^T x_i - w^T X s_i\|^2 = \min_w \frac{w^T X S_\alpha X^T w}{w^T X X^T w},$$

where:

$$\mathbf{S}_\alpha = \mathbf{I} - \mathbf{S} - \mathbf{S}^T + \mathbf{S}^T \mathbf{S} \quad (6)$$

and $\mathbf{S} = [s_1, s_2 \dots s_n]^T$, where $s_i, i = 1, 2, \dots, n$, is the optimal solution of Eq. (5). SPP owns some important characteristics:

- it is linear and doesn't suffer from the out of sample problem;
- it is parameter-free;

- it does naturally preserve some discriminant information of the data;
- it is robust to data noise;

2.3 Sparse Discriminant Preserving Projection (SDPP)

Sparse Discriminant Preserving Projection (SDPP) combines features of SPP and LDA. The discriminant ability of the sparse representation is further strengthened as the vectors of the sparse weight matrix \mathbf{S} are constructed within the classes. According to the authors of SDPP, Zhang et al. [17], SPP does not consider the global structure of the data. Therefore in SDPP they involve the Fisher’s between-class scatter matrix \mathbf{S}_B as a carrier of information of the global structure of the data and construct the model’s objective function as following:

$$\min \frac{\text{tr } w^T S_\alpha w}{\text{tr } w^T S_B w} \tag{7}$$

where \mathbf{S}_α is obtained by Eq. (6) after Eq. (5) has been modified as:

$$\min \|s_i^k\|_1, \text{ s.t. } \|\mathbf{x}_i - \mathbf{X}_k s_i^k\| < \varepsilon, \mathbf{1} = \mathbf{1}^T s_i^k, \text{ label}(x_i) = k, \tag{8}$$

where if n_k is the number of samples with label k ,

$$s_i^k = [s_{i,1}^k, \dots, s_{i,i-1}^k, 0, s_{i,i+1}^k, \dots, s_{i,n_k}^k]^T \text{ and } n = \sum_{k=1}^c n_k.$$

Thus the sparse weight matrix is:

$$\mathbf{S} = \text{diag}(\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^c)$$

where

$\mathbf{S}^k = [s_1^k, s_2^k, \dots, s_{n_k}^k]$, and $k = 1, 2, \dots, c$, are the available classes. \mathbf{S}_B in Eq. (7) is the LDA’s between-class scatter matrix from Eq. (1).

3. The proposed Inverted Sparse Discriminant Preserving Projection

3.1 Motivation

The motivation of the new proposed algorithm comes with an observation of the construction idea of SDPP and has two main aspects. First, we challenge the extend of a discriminative information that the sparsity preserving element in that model contributes and provide a suggestion of the increase of its potential. Secondly, we try to keep the construction idea of SDPP in which a supervised sparsity element is integrated with a supervised “global” part.

On the first aspect: In the unsupervised sparse representation an observation is reconstructed by the remaining samples where it is excluded from its own reconstruction. Analogically, when the classes’ labels are available, a class should be reconstructed by the remaining classes from the dataset. One can expect that a sample,

representing a class, will be “described” better by the observations from the different classes than by the observations from its own class as its reconstruction will reflect its relation to samples that defer from it a priori (as belonging to another classes). In contrast, in the within-class sparse representation case, an observation will be represented only by the same group members which may lead to an omission of discriminative information.

Another reason for the expectation for a better discriminative information preservation when the sparse weight matrix is defined between the classes, is related to the number of observations that participate in the obtainment of the sparsity weight vectors \mathbf{s}_i . In general, it is more likely that the count of the observations from all remaining classes will exceed the same class’ one and therefore would contribute more discriminative information to the particular \mathbf{s}_i . We can demonstrate this with the following illustrative example: if we consider a training dataset that has fully equally distributed classes – e.g. containing 10 classes with 10 samples each – in the computation of the between-class sparsity vector s_i of each sample x_i will participate 90 instead of only nine observations.

On the second aspect: As LDA is a “global” method we expect that the within-class scatter matrix \mathbf{S}_w Eq. (2) would similarly work for the inclusion of global structure preservation, the way the between-class scatter matrix \mathbf{S}_B Eq. (1) does in SDPP.

By taking into account the above considerations we propose a new subspace learning algorithm called Inverted Sparse Discriminant Preserving Projection (ISDPP). Similarly to SDPP, ISDPP simultaneously utilize the properties of SPP with the addition of an information on the global structure of the data contained in the Fisher’s criterion. *In contrast* to SDPP, ISDPP embeds a SR with a sparse weight matrix, the sparse weight vectors of which are computed *between the classes*. On the other hand, the preservation of the global structure of the data is achieved by the inclusion of the LDA’s *within-class scatter matrix* Eq. (2). Thus, while the optimization problem of SDPP is to minimize the objective function Eq. (7), ISDPP seeks a projection that maximises the ratio between the Sparse Representation element – derived by sparsity weight vectors computed between the classes – and the Fisher’s within-class scatter matrix.

Thus, the objective function of the proposed ISDPP is formulated as:

$$\min \frac{\text{tr } w^T \mathbf{S}_\alpha w}{\text{tr } w^T \mathbf{S}_w w} \quad (9)$$

Where \mathbf{S}_α is obtained from Eq. (6) and the sparse weight matrix \mathbf{S} and its vectors are constructed as following:

$$\min \|s_i^k\|_1, \text{ s.t. } \|x_i - X_{\bar{k}} s_i\| < \varepsilon, 1 = 1^T s_i^k, \text{ label}(x_i) = k. \quad (10)$$

The \bar{k} stands for *is not class k*. The notation $X_{\bar{k}}$ means that to the obtainment of the sparse weight vector s_i^k , corresponding to sample i from class k , contribute the samples from all classes other than k . If n_k is the number of samples with label k and

$n_{\bar{k}}$ is the number of samples with different label (\bar{k}), the length of s_i^k will be $n_{\bar{k}}$. The vector s_i^k , the optimal solution of Eq. (10), is completed with n_k -zero coefficients on the places of the respective samples from class k . If we denote the final reconstruction weight vector corresponding to sample i with $s_i^{\hat{k}}$ and $n_k + n_{\bar{k}} = n$, then:

$s_i^{\hat{k}} = [0_{(m)}, s_{i(j)}^k]^T$, where m denotes all indexes of the observations from class k as they are ordered in the dataset \mathbf{X} and j denotes all indexes of the remaining observations from the other classes with an order in $s_i^{\hat{k}}$ the same as the order of $\mathbf{X}_{\bar{k}}$ in \mathbf{X} . The sum of the m -elements is n_k and the sum of the j -elements is $n_{\bar{k}}$, respectively. Then for the sparse weight matrix \mathbf{S} we have:

$$\mathbf{S} = [s_1^{\hat{k}}, s_2^{\hat{k}} \dots, s_n^{\hat{k}}]^T \quad (11)$$

4. Experimental results

4.1 Experiments on the sparse representation

In this experimental round we examine if the between-class sparse weight matrix does add a valuable discriminative information to the SPP, when supervised, how the two alternated supervised SPP's compare to each other and also to the unsupervised SPP.

To evaluate this, we compare the classification results after a subspace learning done by SPP with sparse weight matrices' vectors computed by Eqs. (5), (8) and (10), respectively. We set the value of the parameter ε as 10. For a better differentiation, we denote the algorithm with Eq. (8) as SPP-W and the one with Eq. (10) as SPP-B ("W" and "B" for a within- or between-class notation, respectively).

For the above proposition we involve the public face images database ORL. ORL contains 400 grey scale images of 40 persons with a resolution of 112×92 pixels. Each person has 10 images that have variations in the lightning, facial expressions (smiling/not smiling, open/closed eyes) and/or other details (glasses/no glasses) – **Fig. 1** shows the images of the first one person. All images are cropped and resized to 32×32 pixels. The experiment is conducted as the SPP, SPP-B and SPP-W are trained over the sampled training set for features extractions. The test samples are projected over the generated subspaces and evaluated with the application of kNN classifier. We choose to use 1-nearest-neighbor which is consistent with the literature in the area. The procedure is performed with a random sample of three and five images per person for training with 10 runs for each case. As a pre-processing step we perform PCA and retain 98% of the images' energy. **Tables 1-2** show the obtained average recognition accuracies (RA) – up to the 150th dimension – and their standard deviations over the 10 independent runs for the two scenarios. These are also plotted on **Fig. 2** and **Fig. 3**.



Figure 1. All 10 images of person 1 from the ORL database.

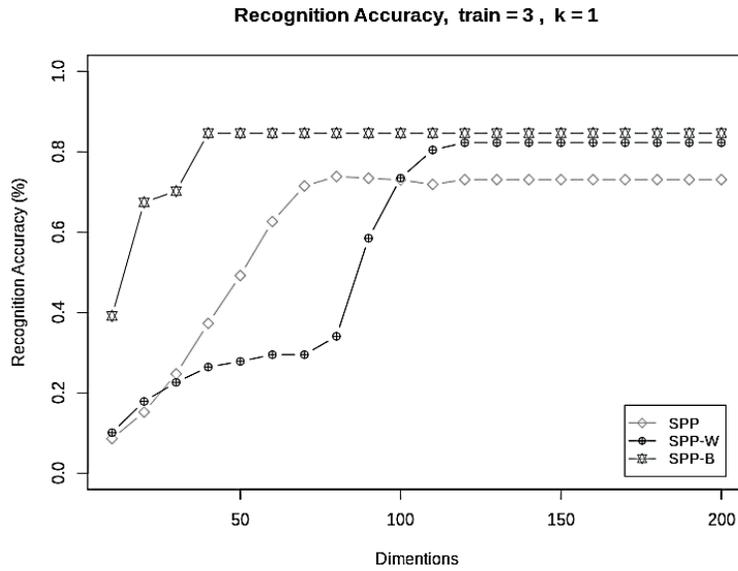


Figure 2. Recognition accuracy (%) vs. number of projected vectors of the SPP's algorithms on the ORL database with a sample of 3 images per subject for training.

Method/ Dim.	Avg./ Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
SPP	Avg.	8.6	15.3	24.8	37.3	49.2	62.6	71.5	73.9	73.4	73.0	71.9	73.1	73.1	73.1	73.1
	Std.	1.7	2.1	3.6	3.4	3.4	2.8	1.7	3.1	3.5	3.5	3.9	3.5	3.5	3.5	3.5
SPP-W	Avg.	10.1	17.9	22.6	26.5	27.9	29.5	29.5	34.1	58.5	73.5	80.5	82.3	82.3	82.3	82.3
	Std.	1.9	2.0	1.5	3.0	3.3	2.9	3.4	3.1	5.2	2.5	2.6	2.5	2.5	2.5	2.5
SPP-B	Avg.	39.2	67.4	70.2	84.6											
	Std.	3.2	2.9	2.6	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Table 1. Recognition accuracy (%) vs. number of projected vectors of the SPP's algorithms on the ORL database with a sample of 3 images per subject for training.

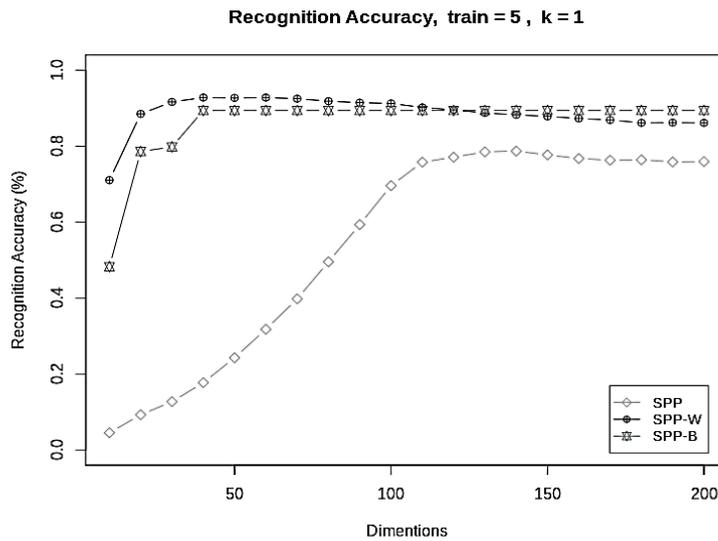


Figure 3. Recognition accuracy (%) vs. number of projected vectors of the SPP's algorithms on the ORL database with a sample of 5 images per subject for training.

Method/ Dim.	Avg/ Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
SPP	Avg.	4.6	9.3	12.8	17.8	24.3	31.8	39.8	49.6	59.4	69.6	75.8	77.1	78.5	78.7	77.7
	Std.	1.3	1.3	2.7	3.5	3.7	4.8	5.4	7.1	6.4	4.8	3.6	3.2	2.6	3.0	3.5
SPP-W	Avg.	71.1	88.5	91.7	92.8	92.7	92.8	92.5	91.9	91.5	91.3	90.2	89.6	88.8	88.3	87.9
	Std.	6.4	3.2	2.1	2.4	2.3	2.4	3.1	2.9	2.8	3.0	2.8	2.3	2.7	2.1	2.5
SPP-B	Avg.	48.3	78.6	79.8	89.4	89.4	89.4	89.4	89.4	89.4	89.4	89.4	89.4	89.4	89.4	89.4
	Std.	4.2	3.8	4.3	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7

Table 2. Recognition accuracy (%) vs. number of projected vectors of the SPP's algorithms on the ORL database with a sample of 5 images per subject for training.

4.1.1 Results

In this experimental phase we observe a clear advantage of the supervised modifications of SPP over the classical SPP. This is in line with the expectations and demonstrates explicitly the additional discriminant information contribution of the SR when it is embedded as supervised in compiled algorithms. In regard to the SPP-W versus SPP-B comparison, SPP-W reaches higher absolute recognition accuracy in the 50%-training-samples case. On the other side, SPP-B stays more robust when the training sample decreases. That confirms the considerations that the out-of-the-class samples approach could provide a stable preservation of discriminant information.

4.2 Experiments on the introduced ISDPP

To evaluate the performance of the proposed ISDPP algorithm we conduct further experiments on the public ORL and AR databases and compare it with the effectiveness of other well-known algorithms like DSNPE, SLPDP and SDPP.

4.2.1 Experiments with the ORL database, experimental design.

This experimental round involves the ORL database, described above. In order to simultaneously examine the robustness of the proposed ISDPP we decide to occlude the 112×92 pixels training samples with a random 24×24 pixels square, in which each pixel of the square takes a random value from the interval $[0, 1]$, whereas a value of zero represents pure black and a value of one pure white color, respectively – **Fig. 4** shows an example of three original and randomly “occluded” images from the database. Further, the images are cropped to 32×32 pixels and a PCA with 98% energy retention is performed. Similarly to the first experimental round above, we define training sets with three and five random images per subject, respectively, and do 10 runs for each case. The new proposed ISDPP is executed together with the other three algorithms for a subspace learning over the so defined training subsets. The parameter ε in the sparse representations has the value of 10 and for the parameter γ in SPNPE we take the value of 1 (as in [15]) and the value of 0.1 in SLPDP (as in [16]), respectively. The test samples are mapped to the respective projections and evaluated with 1-NN classifier. The obtained classification results are presented in **Tables 3-4** and visualised on **Fig. 5, 6**.



Figure 4. Example of three original and randomly “occluded” images from the ORL database

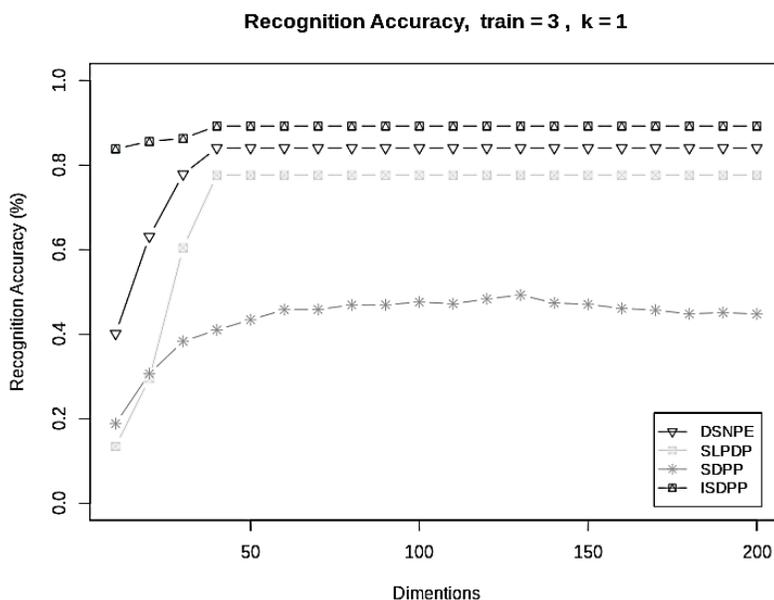


Figure 5. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the ORL database with a sample of 3 images per subject for training.

Method/ Dim.	Avg/ Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
DSNPE	Avg.	40.1	63.1	77.9	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1
	Std.	4.1	5.6	2.8	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
SLPDP	Avg.	13.5	29.5	60.4	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6
	Std.	2.5	5.4	4.0	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
SDPP	Avg.	18.9	30.7	38.3	41.0	43.4	45.8	45.9	46.9	47.0	47.6	47.2	48.4	49.3	47.4	47.1
	Std.	8.5	11.1	13.6	14.6	14.7	14.3	16.2	15.6	16.6	17.6	18.8	19.1	19.1	19.1	19.5
ISDPP	Avg.	83.9	85.6	86.3	89.3											
	Std.	2.9	3.0	2.7	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5

Table 3. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the ORL database with a sample of 3 images per subject for training.

4.2.2 Experiments with the AR database, experimental design.

The AR face database contains over 4000 color face images of 126 individuals (70 men and 56 women). In our experiment we use a subset of the AR face database provided and preprocessed by Martinez et al. [18]. This subset contains 1400 face images, with a resolution of 165 x 120 pixels, corresponding to 100 person (50 men and 50 women), where each person has 14 different images with illumination change and expressions – **Fig. 7** shows the images of one person from the database.

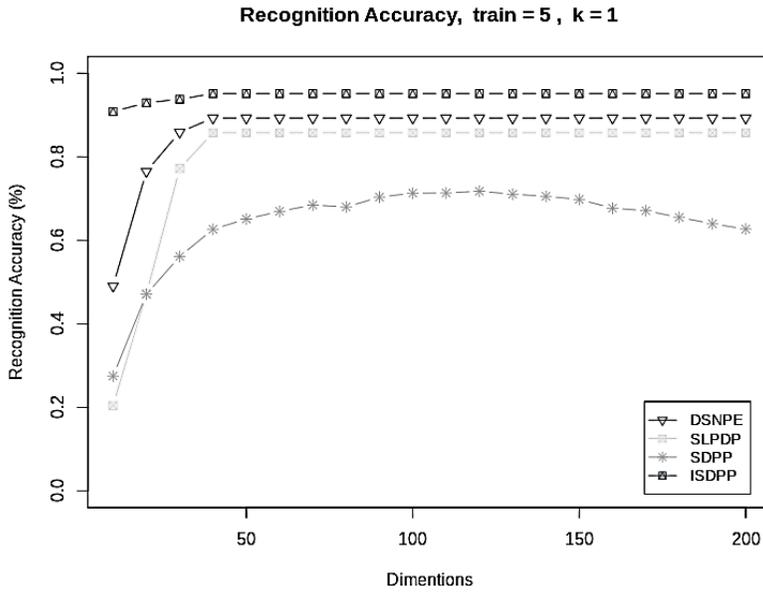


Figure 6. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the ORL database with a sample of 5 images per subject for training.

Method/ Dim.	Avg/Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
DSNPE	Avg.	49.1	76.5	85.9	89.3	89.3	89.3	89.3	89.3	89.3	89.3	89.3	89.3	89.3	89.3	89.3
	Std.	5.6	5.4	3.4	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6
SLPDP	Avg.	20.5	47.2	77.2	85.8	85.8	85.8	85.8	85.8	85.8	85.8	85.8	85.8	85.8	85.8	85.8
	Std.	4.6	6.3	5.1	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3
SDPP	Avg.	27.5	47.1	56.2	62.7	65.1	67.0	68.5	68.0	70.4	71.3	71.4	71.8	71.1	70.6	69.8
	Std.	7.5	11.0	9.4	10.5	10.5	11.3	10.5	11.3	11.6	10.3	9.5	9.9	10.8	11.1	12.5
ISDPP	Avg.	90.9	93.0	93.8	95.2	95.2	95.2	95.2	95.2	95.2	95.2	95.2	95.2	95.2	95.2	95.2
	Std.	2.3	1.4	1.8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5

Table 4. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the ORL database with a sample of 5 images per subject for training.

We design two scenarios with four and seven images per subject for training. Similarly, we occlude the training images with a random 32×32 pixels square – **Fig. 8** shows an example of three original and randomly “occluded” images from the database – and perform a PCA over the training set with a 98% energy retention. For the sparse representations we use the second modified sparse representation, Eq. (16) from [14] and the parameter γ in SPNPE takes the value of 1 and 0.1 in SLPDP, respectively. The obtained recognition accuracies of the methods are presented on **Tables 5-6** and **Fig. 9, 10**.



Figure 7. The images of one person from the AR database.



Figure 8. Example of three original and randomly “occluded” images from the AR database

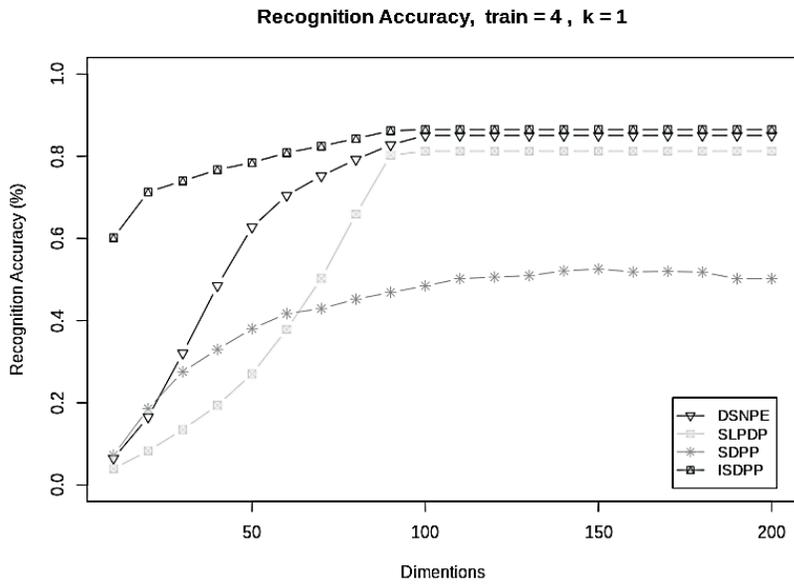


Figure 9. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the AR database with a sample of 4 images per subject for training.

Method/ Dim.	Avg/ Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140
DSNPE	Avg.	6.5	16.5	32.1	48.5	62.8	70.5	75.2	79.2	82.8	85.1	85.1	85.1	85.1	85.1
	Std.	1.3	2.6	3.5	3.9	2.4	2.0	2.1	1.7	1.7	1.1	1.1	1.1	1.1	1.1
SLPDP	Avg.	4.0	8.3	13.5	19.4	27.0	37.9	50.3	65.9	80.3	81.2	81.2	81.2	81.2	81.2
	Std.	0.8	1.0	2.2	3.0	2.9	4.7	6.8	6.2	1.5	1.4	1.4	1.4	1.4	1.4
SDPP	Avg.	7.3	18.6	27.6	33.0	38.0	41.7	42.9	45.2	46.9	48.5	50.2	50.6	51.0	52.1
	Std.	1.7	1.6	2.9	3.4	2.7	2.5	3.1	4.2	4.1	4.2	4.2	3.8	3.8	4.6
ISDPP	Avg.	60.2	71.3	74.0	76.7	78.5	80.9	82.5	84.3	86.2	86.5	86.5	86.5	86.5	86.5
	Std.	1.9	2.5	2.7	2.7	3.0	2.5	2.5	2.4	1.7	1.8	1.8	1.8	1.8	1.8

Table 5. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the AR database with a sample of 4 images per subject for training.

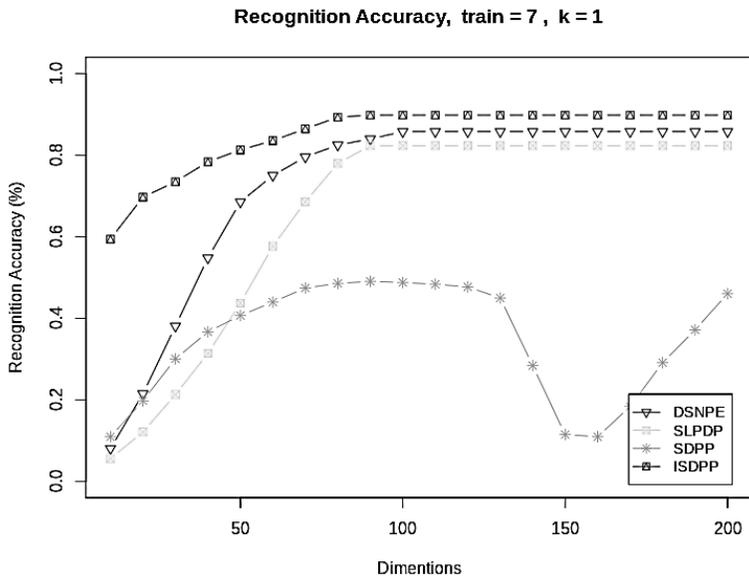


Figure 10. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the AR database with a sample of 7 images per subject for training.

4.3 Experiments on the ISDPP - results

In these experimental runs we register a clear advantage of the ISDPP algorithm over the comparison ones. In addition, the maximal recognition accuracy of the ISDPP is reached faster or, in the worst case, together with the maximal RA of the other models and with the minimal average standard deviation in the half of the cases. The significance of the results is further increased by considering the circumstance of the sizable artificial occlusion of the training data.

Method/ Dim.	Avg/ Std.	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
DSNPE	Avg.	8.1	21.5	38.1	54.8	68.5	75.0	79.6	82.5	84.0	85.8	85.8	85.8	85.8	85.8	85.8
	Std.	0.7	1.9	1.4	1.9	2.7	2.4	2.0	1.9	1.0	0.9	0.9	0.9	0.9	0.9	0.9
SLPDP	Avg.	5.6	12.1	21.3	31.4	43.7	57.7	68.6	78.0	82.3	82.3	82.3	82.3	82.3	82.3	82.3
	Std.	1.0	1.0	1.9	1.9	3.3	0.6	1.3	1.5	0.8	0.8	0.8	0.8	0.8	0.8	0.8
SDPP	Avg.	10.9	19.8	30.1	36.7	40.7	44.0	47.4	48.5	49.1	48.8	48.3	47.7	45.0	28.4	11.5
	Std.	1.5	3.0	4.2	4.1	2.4	3.1	2.9	2.7	3.1	2.6	2.9	2.9	1.7	12.0	6.0
ISDPP	Avg.	59.4	69.7	73.5	78.4	81.3	83.6	86.5	89.3	89.8						
	Std.	1.4	1.8	2.0	2.0	2.1	1.7	1.1	1.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 6. Recognition accuracy (%) vs. number of projected vectors of the proposed ISDPP together with another algorithms on the AR database with a sample of 7 images per subject for training.

5. Conclusion and further work

In this paper we propose a new subspace learning algorithm, called Inverted Sparse Discriminant Preserving Projection (ISDPP). ISDPP borrows the conceptual idea of SDPP and likewise combines a supervised sparse representation (SR) with the Fisher’s criterion. Prior to the ISDPP construction, we argue the way a supervised SR may be constructed and assume that the contribution of a discriminant information of the between-the-classes SR could exceed the intra-class one. As it is difficult for one to define the effect of a component of a model in the productivity effectiveness of the assembled algorithm it is part of, we decided to compare separately the classification performance of the SR under the three scenarios when the sparse weight matrix remains unsupervised or is built within- and between-the-classes, respectively. The results on the public images ORL database in this preliminary experimental phase support the construction idea of ISDPP. In the following experiments on the ORL and AR database we observe a stable and robust superior recognition accuracy of the new proposed algorithm compared to other well-known successful algorithms like DSNPE, SLPDP and SDPP, that is achieved under the conditions of randomly occluded training data. As our experiments are conducted on well-structured, taken in controlled environment, face images data, the recognition accuracy we observe may have a big gap from a real face recognition application. However, we argue that the proposed method could serve as a useful pattern recognition tool in more complex face recognition strategies applied on images with higher variations and/or such taken in the wild. As a further work, we intend to conduct experiments on face images with higher variations or images from another, various, types.

Acknowledgements

This work was supported by a research project 80-10-93/25.03.2021 of Sofia University “St. Kliment Ohridski”.

References

- [1] S. Wold, K. Esbensen and P. Geladi, "Principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 2, Issues 1–3, pp. 37-52, 1987.
- [2] M. Turk and A. Pentland, "Face recognition using eigen faces", In *Proceeding of IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586-591, 1991.
- [3] P. Belhumeur, J. Hefanpha and D. Kriegman, „Eigenfaces vs. Fisherfaces: recognition using class specific linear projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 711-720, 1997.
- [4] A.M. Martinez and A.C. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), pp. 228-233, 2001.
- [5] X. He and P. Niyogi, "Locality Preserving Projections", *IEEE Transactions on Reliability – TR*, 16, 2002.
- [6] H. Li, T. Jiang, K. Zhang, "Efficient and robust feature extraction by maximum margin criterion", *IEEE Transactions on Neural Networks* 17:1157–1165, 2006.
- [7] S.T.Roweis, L.K.Saul, "Nonlinear dimensionality reduction by locally linear Embedding", *Science* 290:2323–2326, 2000.
- [8] J.B. Tenenbaum, V. de Silva, J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science* 290:2319–2324, 2000.
- [9] M. Belkin, P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural Computation*, 15:1373–1396, 2003.
- [10] X. He, D. Cai, S. Yan, H. Zhang, "Neighborhood preserving embedding", *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV)*, 1208–1213, 2005.
- [11] W. Yu, R. Wang, F. Nie, F. Wang, Q. Yu and X. Yang, "An improved locality preserving projection with ℓ_1 -norm minimization for dimensionality reduction", *Neurocomputing*, 316:322-331, 2018.
- [12] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections", *Image Vision Computing*, 24, pp. 239–248, 2006.
- [13] G.F Lu, Z. Lin and Z. Jin, "Face recognition using discriminant locality preserving projections based on maximum margin criterion", *Pattern Recognition*, 43-10:3572-3579, 2010.
- [14] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition", *Pattern Recognition*, 43-1:331-341, 2010.

- [15] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei and S. Ji, “Discriminant sparse neighborhood preserving embedding for face recognition”, *Pattern Recognition*, 45-8:2884-2893, 2012.
- [16] J. Zhang, J. Wang and X. Cai, “Sparse locality preserving discriminative projections for face recognition”, *Neurocomputing*, 260:321-330, 2017.
- [17] J. Zhang, J. Wang and K. Zhan, “Sparse Discriminant Preserving Projections for Face Recognition”, at 3rd International Conference on Information Technology and Industrial Automation (ICITIA 2018), ISBN: 978-1-60595-607-7, 2018.
- [18] A.M. Martinez, R. Benavente, “The AR Face Database”, CVC Technical Report, 1998.