Statistical binary pattern and post-competitive representation for pattern recognition

Mohamed Anouar Borgi[1], Thanh Phuong Nguyen[2,3], Demetrio Labate[4], Chokri Ben Amar[1]

[1]Research Groups on Intelligent Machines, University of Sfax, BP 1173, Sfax 3038, Tunisia
[2] Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France
[3]Université de Toulon, CNRS, LSIS, UMR 7296, 83957 La Garde, France
[4]Department of Mathematics, University of Houston, Houston, TX 77204, USA

{anoir.borgi@ieee.org; nthanhphuong@gmail.com;dlabate@math.uh.edu; chokri.benamar@ieee.org}

Abstract

During the last decade, sparse representations have been successfully applied to design high-performing classification algorithms such as the classical sparse representation based classification (SRC) algorithm. More recently, collaborative representation based classification (CRC) has emerged as a very powerful approach, especially for face recognition. CRC takes advantage of sparse representation based classification through the notion of collaborative representation, relying on the observation that the collaborative property is more crucial for classification than the l_1-norm sparsity constraint on coding coefficients used in SRC. This paper follows the same general philosophy of CRC and its main novelty is the application of a virtual collaborative projection (VCP) routine designed to train images of every class against the other classes to improve fidelity before the projection of the query image. We combine this routine with a method of local feature extraction based on high-order statistical moments to further improve the representation. We demonstrate using extensive experiments of face recognition and classification that our approach performs very competitively with respect to state-of-the-art classification methods. For instance, using the AR face dataset, our method reaches 100% of accuracy for dimensionality 300.

Keywords Statistical binary pattern · virtual projection · twin collaborative representation · face recognition · image categorization · action recognition

1 Introduction

One of the main challenges of current research in pattern recognition (PR) is to improve the robustness of exiting algorithms with respect to confounding factors including noise, rigid transformations, changes in viewpoint, illumination, etc. Recent advances from statistical learning [1] have brought attention to the notion of sparsity to extract the salient image features in such a way to obtain more accurate and robust classification. Wright et al. [18], in particular, introduced a very influential framework called Sparse Representation based Classification (SRC) for face recognition (FR) and successfully applied this method to identify human faces with varying illumination changes, occlusion and real disguise. In their method, a test sample image is coded as a sparse linear combination of the training images and classification is achieved by identifying which class yields the least residual. Several

other methods were inspired by SRC including: the FR method based on sparse representation of facial image patches by Theodorakopoulos et al. [4]; Kernel Sparse Representation for image classification and FR, which applies a sparse coding technique in a high dimensional feature space via some implicit feature mapping [39]; the Gabor occlusion dictionary for SRC by Yang and Zhang which reduces the computation cost by using Gabor feature [5]; a robust regularized coding model to enhance the robustness of face recognition to confounding factors [6] [7]; the method based on maximum correntropy criterion for robust face recognition by He et al. [8]. An alternative point of view was proposed by Zhang et al. [9] who argued that rather than sparsity "the collaborative representation mechanism used in SRC is much more crucial to its success of face classification". Based on this observation, they introduced a method called Collaborative Representation based Classification with regularized least square (CRC) [9] which was shown to perform very competitively against SRC with a lower computational cost. As a further refinement of CRC, some of the authors proposed a method called Relaxed Collaborative Representation (RCR) which is designed better capture the similarity and distinctiveness of different features for the classification [10]. An alternative approach is the two-phase test sample representation method [54] and relies on detecting first the training samples located away from the test sample (assuming they have negligible effect on classification); next the test sample is represented as a linear combination of the M nearest neighbors and the representation result is used for classification. Another method proposed in [55] consists in partitioning face images into blocks and then creating an indicator to remove the contaminated blocks and choose the k nearest subspaces; SRC is finally used to classify the occluded test sample in the new feature space.

We also recall the Fisher Discrimination Dictionary Learning (FDDL) algorithm by Yang et al. [11] which embeds the Fisher criterion in the objective function design. The FDDL scheme has two remarkable properties. First, dictionary atoms are learnt to associate the class labels so that the reconstruction residual from each class can be used in classification; second, the Fisher criterion is imposed on the coding coefficients so that they carry discriminative information for classification. To improve this method, Feng et al. [12] propose to learn jointly the projection matrix for dimensionality reduction and the discriminative dictionary for face representation JDDLDR. The joint learning combines more effectively the learned projection and the dictionary with the result of improving FR performance. Within the general framework of the discriminative dictionary learning (DDL), the Projective Dictionary Pair Learning (DPL) algorithm [56] learns a synthesis dictionary and an analysis dictionary jointly to achieve the goal of signal representation and discrimination. The vector guided dictionary learning (SVGDL) method is proposed in [57] as a special case of the Fisher discrimination dictionary learning (FDDL) method; here the weights are determined by the numbers of samples of each class and a parameterization method is used to adaptively determine the weight of each coding vector pair. Compared with FDDL, SVGDL can adaptively assign different weights to different pairs of coding vectors. Yet another DDL approach recently proposed is the Locality Constrained and Label Embedding Dictionary Learning (LCLE-DL) algorithm [58], where locality information is preserved using the graph Laplacian matrix of the learned dictionary rather than the conventional one derived from the training samples; next, the label embedding term is constructed using the label information of atoms instead of the classification error term; the coding coefficients derived by combinig locality-based and label-based reconstruction are shown to be very effective for image classification. Very recently, it was proposed a probabilistic interpretation of the collaborative

classification mechanism to explain the classification mechanism of CRC and following this analysis it was introduced a method called probabilistic collaborative representation based classifier (ProCRC) which jointly maximizes the likelihood that a test sample belongs to each of the multiple classes [48].

On other hand, a class of algorithms described as Local Feature based methods [13], [14], [15], [16], [17], [19], [20], [21], [22], [23] also demonstrated very promising results in problems of object recognition and texture classification. For instance, some of these methods use Gabor filters to extract local directional features on multiple scales and have been successfully applied in FR [14], [15]. Compared to more conventional methods such as Eigenface [2] and FisherFace [3], Gabor filtering is less sensitive to image variations. Another type of local feature widely used in FR is Statistical Local Feature (SLF), such as histogram of Local Binary Pattern (LBP) [16], whose main principle is to model a face image as a composition of micro-patterns [23]. By partitioning the face image into several blocks, the statistical feature (e.g., histogram of LBP) of these blocks is extracted, and finally the description of the image is formed by concatenating the extracted features in all blocks. For example, Zhang et al. [19], [20] proposed to use Gabor magnitude or phase map instead of the intensity map to generate LBP features. New coding techniques on Gabor features have also been proposed, e.g., Zhang et al. [21] extracted and encoded the global and local variations of the real and imaginary parts of the data using a multi-scale Gabor representation. Borgi et al. [24] [49] [1] proposed two algorithms that apply a sparse multiscale representation based on shearlets to extract the essential geometric content of facial features, one called Regularized Shearlet Network (RSN) and another one Sparse Multi-Regularized Shearlet Network (SMRSN). Finally, we recall that Meng et al. [25] proposed a kernel based representation model to fully exploit the discrimination information embedded in the statistical local features (SLF_RKR) and applied a robust regression method handle occlusions in face images.

In this paper, we adopt the same general philosophy of CRC and our main novel contribution is to integrate this method with a virtual collaborative projection (VCP) routine designed to train images of every class against the others classes with the goal to improve fidelity before projecting the query image. Additionally, inspired by the remarkable results obtained from the recent literature in Local Feature based method, our algorithm includes a routine to compute high-order statistical moments (SM) in order to extract highly discriminative local features and improve data representation. To validate our algorithm, which is called Statistical Binary Pattern with Virtual Competitive Representation (SBP_VCP), we have tested it on multiple datasets for problems of face recognition, gender classification, handwritten digit recognition, object categorization and action recognition. Experimental results show that our method consistently achieves very competitive results as compared to classical and state-of-the-art algorithms.

The rest of this paper is organized as follows. Section 2 introduces the main idea of statistical binary pattern and high order moments for feature extraction. Section 3 describes the proposed virtual collaborative projection applied to trained faces. Section 4 reports extensive numerical experiments to validate the proposed method and compare it against state-of-the-art methods on problems of face recognition under different confounding factors as well as image categorization, handwritten digit and action recognition. Finally, Section 5 concludes this paper.

2 Statistical binary pattern and high order moments

The Statistical Binary Patterns (SBP) representation is an extension of Local Binary Patterns (LBP) and it aims at enhancing the expressiveness and discrimination power of LBP for image modelling (especially texture) and recognition, while reducing sensitivity to small perturbations, e.g., noise. The main idea of this method, which was introduced by one of the authors and their collaborator in [26], consists in applying a rotation invariant uniform LBP to a set of images corresponding to the local statistical moments associated to a given spatial support. The resulting code forms the SBP and an image is then represented by joint or marginal distributions of SBPs.

2.1 Moment images

A real valued 2d discrete image f is modelled as a mapping from $\mathbb{Z}^2$ to $\mathbb{R}$. The spatial support used to calculate the local statistics is modelled as $B \subset \mathbb{Z}^2$, such that $O \in B$, where O is the origin of $\mathbb{Z}^2$. The r-order moment image associated to f and B is also a mapping from $\mathbb{Z}^2$ to $\mathbb{R}$, defined as:

$$m_{(f,B)}^r(z) = \frac{1}{|B|} \sum_{b \in B} \left(f(z+b) \right)^r \tag{1}$$

Where z is a pixel from $\mathbb{Z}^2$, and $|B|$ is the cardinality of the structuring element B. Accordingly, the r-order centered moment image (r > 1) is defined as:

$$\mu_{(f,B)}^r(z) = \frac{1}{|B|} \sum_{b \in B} \left(f(z+b) - m_{(f,B)}^1(z) \right)^r \tag{2}$$

Where $m_{(f,B)}^1(z)$ is the average value (1-order moment) calculated around z. Finally the r-order normalized centered moment image (r > 2) is defined as:

$$\beta_{(f,B)}^r(z) = \frac{1}{|B|} \sum_{b \in B} \left(\frac{f(z+b) - m_{(f,B)}^1(z)}{\sqrt{\mu_{(f,B)}^2(z)}} \right)^r \tag{3}$$

where $\mu_{(f,B)}^2(z)$ is the variance (2-order centered moment) calculated around z.

2.2 Statistical Binary Patterns

Let R and P denote the radius of the neighborhood circle and the number of values sampled on the circle, respectively. For each moment image M, one statistical binary pattern is formed as follows:
- one (P +2)-valued pattern corresponding to the rotation invariant uniform LBP coding of M:

$$SBP_{P,R}(M)(z) = LBP_{P,R}^{riu2}(M)(z) \qquad (4)$$

- one binary value corresponding to the comparison of the centre value with the mean value of M:

$$SBP_C(M)(z) = s\left(M(z) - \tilde{M}\right) \qquad (5)$$

where s denotes the pre-defined sign function, and $\tilde{M}$ the mean value of the moment M on the whole image. Hence $SBP_{P,R}(M)$ represents the structure of the moment M with respect to a local reference (the center pixel), and $SBP_C(M)$ complements the information with the relative value of the center pixel with respect to a global reference ($\tilde{M}$). As a result of this first step, a $2(P+2)$ - valued scalar descriptor is then computed for every pixel of each moment image.

2.3 Image Descriptors

Let $\{M_i\}_{1 \le i \le n_M}$ be the set of n_M computed moment images. $SBP^{\{M_i\}}$ is defined as a vector valued image, with n_M components such that for every $z \in \mathbb{Z}^2$, and for every i , $SBP^{\{M_i\}}(z)_i$ is a value between 0 and 2(P + 2). If the image f contains texture, the descriptor associated to f is made by the histogram of the values of $SBP^{\{M_i\}}$. We consider two kinds of histograms. First we consider the joint histogram H defined as follows:

$$H : \left[0, 2(P+2)\right[^{n_M} \to \mathbb{N}$$
$$H(v) = \left|\left\{z; SBP^{\{M_i\}}(z) = v\right\}\right| \qquad (6)$$

Depending on the size of the texture images, the joint distribution may become too sparse when the dimension (i.e., the number of moments) increases.
Next, we consider the marginal histograms $\{h_i\}_{i \le n_M}$ defined as:

$$H : \left[0, 2(P+2)\right[\to \mathbb{N}$$
$$h_i(n) = \left|\left\{z; SBP^{\{M_i\}}(z)_i = n\right\}\right| \qquad (7)$$

An image descriptor can then be defined using the joint histogram H or the concatenation of the n_M marginal histograms $\{h_i\}$. The length of the descriptor vector is $[2(P+2)]^{n_M}$ in the first case and $2n_M(P+2)$ in the second case.

2.4 Higher order moments

The SBP model on higher order moments is evaluated next. The objective of the SBP framework is to extend the LBP texture image descriptors from the local level, represented by the pixel z , to the regional distribution level of $z + B$ by approximating the distribution to a

set of statistical moments. It is known that the mean and variance describe faithfully a statistical distribution only in special cases, e.g., when it is a normal distribution. This assumption may fail for natural texture images. Therefore, higher order moments are needed to obtain an accurate description of a general distribution and capture the relevant information.

Regarding the size of the image descriptor, it clearly increases as the number of moments increase. When we use joint histograms, the descriptor size is $(2(P+2))^n$ where P is the number of neighbours used in LBP and n is the number of moment images. When we use marginal histograms, the size is only $2n(P+2)$ but this comes at the price of a significant loss of information. Hence we propose a trade-off between descriptor size and information loss based on the concatenation of joint histograms corresponding to pairs of moment images. Formally, we can recursively define the higher order SBP hybrid image descriptor as follows.

Let M_1 and M_2 be moments or combinations of moments by their joint or concatenated histogram. We shall denote as $SBP^{M_1 M_2}$ (resp. $SBP^{M_1 - M_2}$) the image descriptor made by the joint (resp. concatenated) histograms constructed from SBP^{M_1} and SBP^{M_2}. In our experiments for higher order moments below, we have only considered pairs of moments for joint histograms. The algorithm below summarizes the high order binary statistical moment SBP:

The SBP Algorithm

Input: f - a 2D image, $B \subset \mathbb{Z}^2$ - the spatial support used to calculate the local moments, P – the number of neighbours, R – the radius neighbouring circle.

Output: $SBP_{P,R}^{m_1 \mu_2}$ – texture descriptor of f.

Calculate moment images:

1. Calculate the first order moment image m_1 (or $m_{(f,B)}^1$) associated to f and B using the formula (1).

2. Calculate the second order centred moment image μ_2 (or $\mu_{(f,B)}^2$) associated to f and B using the formula (2).

Statistical Binary Patterns:

1. Calculate statistical binary patterns $SBP_{P,R}(m_1)$ and $SBP_C(m_1)$ from the first order moment images m_1, using the formulas (5) and (6).

2. Calculate statistical binary patterns $SBP_{P,R}(\mu_2)$ and $SBP_C(\mu_2)$ from the second order moment images μ_2, using the formulas (5) and (6).

3. Calculate $SBP_{P,R}^{m_1 \mu_2}$ as joint histogram of $SBP_{P,R}(m_1)$, $SBP_C(m_1)$, $SBP_{P,R}(\mu_2)$ and $SBP_C(\mu_2)$.

Figures 1 and 2 compare the recognition rate of the algorithms LBP, CLBP [53] and SBP. For this comparison, we used the Outex database [52], a large and comprehensive texture database which includes 24 classes of textures collected under three illuminations and at nine angles. To measure the dissimilarity between the two histograms, we used the nearest

neighborhood classifier with the chi-square distance. We considered different configurations of SBP: in Figure 1 we set the (P,R) value equal to (24,3); in Figure 2 we used values (8,1), (16,2) and (24,3).
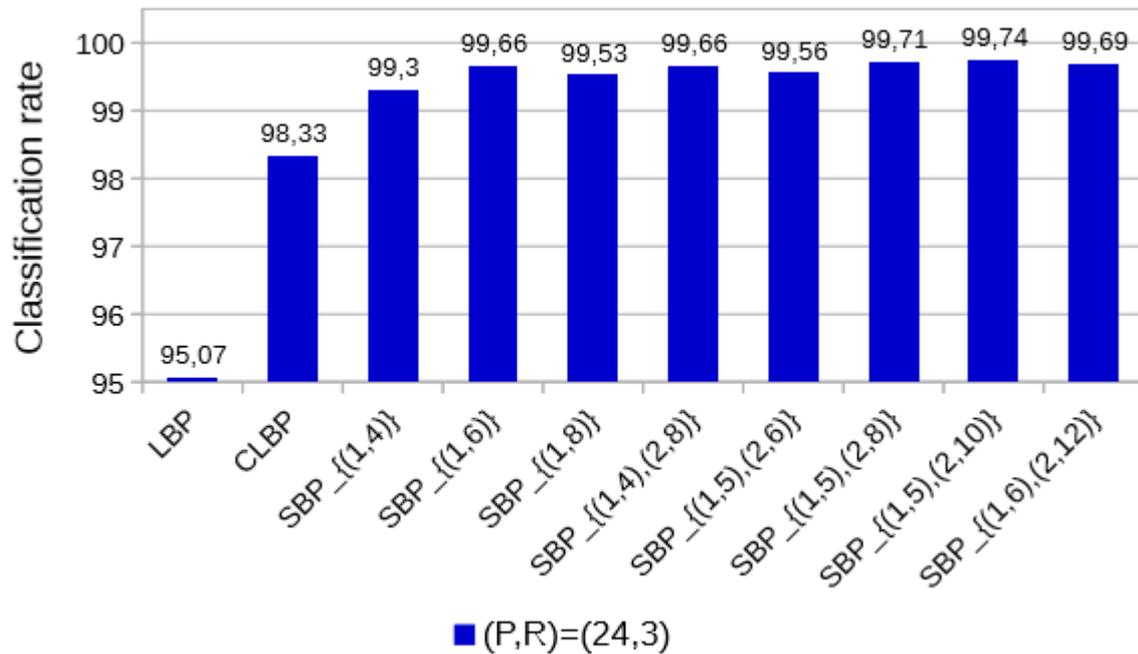
Fig. 1 Classification rate (%) of LBP, CLBP and SBP with the value (P,R) = (24,3) using the Outex texture database.

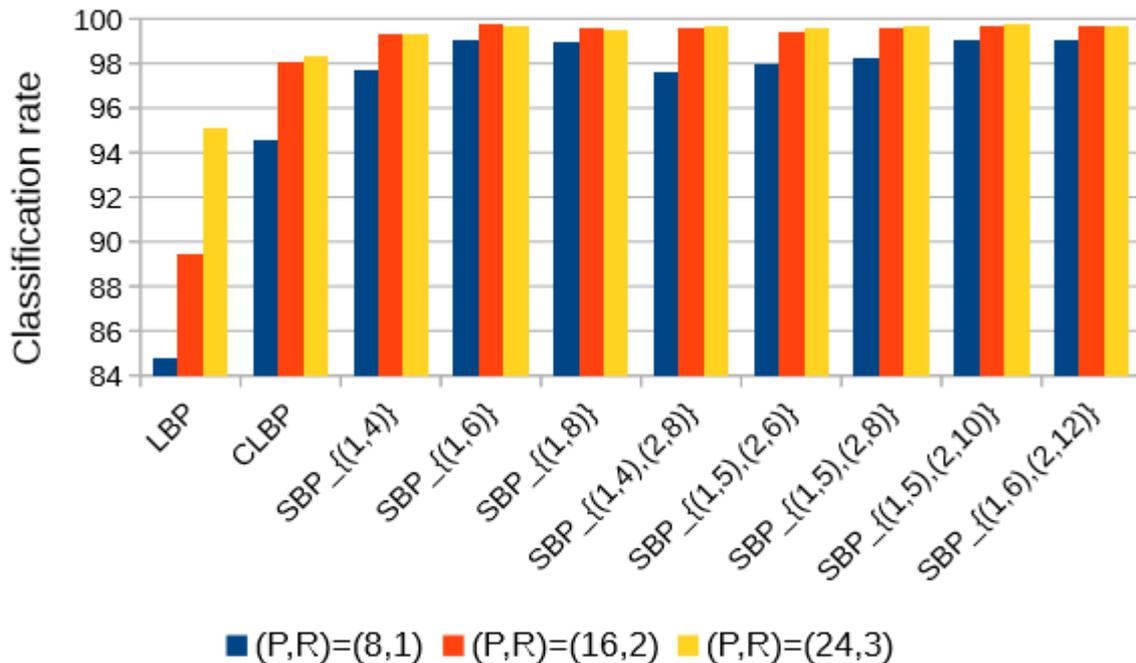**Fig. 2** Classification rate (%) of LBP, CLBP and SBP with the values (P,R) = (8,1), (P,R) = (16,2) and (P,R) = (24,3) using the Outex texture database.

3   Virtual collaborative projection

Zhang et al. [9] investigated the role of collaboration between classes in representing the query sample. In order to collaboratively represent the query sample $y \in \mathbb{R}^m$ using X (all the

gallery images where each column is a training sample) with low computational cost, they introduced a method called Collaborative Representation based Classification with Regularized Least Square method (CRC_RLS). A general model of collaborative representation is:

$$\tilde{\alpha} = \arg\min_{\alpha} \left\{ \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_2^2 \right\}$$ (8)

where α is the coding vector ($\alpha = [\alpha_1,...,\alpha_i,...]$ and $y \approx X\alpha$) and λ is the regularization parameter.

The algorithm is described below:

The CRC-RLS Algorithm

    1. Normalize the columns of X to have unit l_2-norm.
    2. Code y over X by

$$\tilde{\alpha} = Py$$

    where $P = \left(X^T X + \lambda I \right)^{-1} X^T$.

    3. Compute the regularized residuals

$$r_i = \left\| y - X_i \tilde{\alpha}_i \right\| \Big/ \left\| \tilde{\alpha}_i \right\|_2$$

    4. Output the identity of y as

$$\text{identity}(y) = \text{argmin}_i \{ r_i \}$$

---

where $\tilde{\alpha}_i$ is the coding vector associated with class i.

The method proposed in this paper improves this algorithm by increasing the fidelity of the training images and enhancing the collaboration between classes by representing not only the query sample y but also all gallery images x_i of every class i based on the idea of virtual collaborative projection (VCP).

Using this idea, we can compute the average images C_i from every class i over X, defined as:

$$C_i = \sum_1^{tr} x_i \Big/ N_{tr}$$ (9)

where N_{tr} represents the number of training images of a class i.

Next by computing P as:

$$P = \left(X^T X + \lambda I \right)^{-1} X^T$$ (10)

then the resulting virtual coefficient $\tilde{\alpha}_{virtual}$ is calculated as follows:

$$\tilde{\alpha}_{virtual} = PC_i$$ (11)

This virtual coefficient is used as a weight for every class i and reconstruct a new gallery images d_{c_i} :

$$d_{c_i} = \left\| \tilde{\alpha}_{virtual} \right\|_2 C_i$$ (12)

A new dictionary D (the update of X) is then obtained by combining all images d_{c_i} ($D = \left[d_{c_1}, ..., d_{c_i}, ... \right]$).

Next, when a query sample y is presented to be classified, we follow the same procedure as CRC_RLS by computing the regularized residuals r_i but we utilize the new dictionary D:

$$r_i = \left\| y - D_i \tilde{\alpha}_{virtual} \right\| \Big/ \left\| \tilde{\alpha}_{virtual} \right\|_2 \qquad (13)$$

where D_i represents the images of a class i. The identity of a query sample y is computing by:

$$identity(y) \ = \ \arg\min_i \left\{ r_i \right\} \qquad (14)$$

Below we present our virtual collaborative projection (VCP) algorithm when a query image y is presented to be classified:

The VCP Algorithm
1. Normalize the columns of X to have unit l_2-norm.
2. Compute the average images C_i of every class i using the formula (9).
3. Compute the virtual coefficient $\tilde{\alpha}_{virtual}$ using the formulas (10) and (11).
4. Compute d_{c_i} using the formula (12).
5. Combining all the d_{c_i} in a dictionary D.
6. Compute the regularized residuals r_i using the formula (13).
7. Return the identity of y using the formula (14).

In order to investigate the efficiency of VCP versus CRC, we conducted some experiments using the AR face dataset [27] with different dimensionality. Note that PCA is used to reduce the dimensionality of original face images, and the Eigenface features are used for this first experiment with three dimensions 54, 120 and 300.

For this comparison, we selected a subset from AR dataset that contains 50 male subjects and 50 female subjects with only illumination and expression changes. For each subject, the seven images from Session 1 were used for training and the other seven images from Session 2 were used for testing. The images were cropped and resized to 60×43. Table 1 shows that VCP performs slightly better than CRC_RLS [9]:

Table 1 Comparison VCP vs. CRC using AR data set with different dimensionality.

Dimension	54	120	300
CRC_RLS [9]	80.5%	90.0%	93.7%
VCP	**80.8%**	**91.1%**	**94.3%**

Additional experiments are conduct in Section 4 with object categorization and action recognition where we use features provide by state-of-the-art methods and not the high order statistical moments.

We conclude this section by presenting our algorithm of high order Statistical Binary Pattern with Virtual Collaborative Projection (SBP_VCP) obtained by adding the step of high order statistical moments features extraction (cf. Section 2) to the VCP algorithm. This additional step is performed for the training images X resulting in a new training set and for every query sample y.

The SBP_VCP Algorithm

1. Extract the statistical binary patterns $SBP_{P,R}^{m_1 \mu_2}$ of X using the SBP Algorithm.
2. Extract the statistical binary patterns $SBP_{P,R}^{m_1 \mu_2}$ of y using the SBP Algorithm.
3. Call VCP algorithm.

In the next section we illustrate the performance of the SBP_VCP approach.

4 Experiments

To demonstrate the performance of our SBP_VCP algorithm, we conducted extensive experiments on multiple benchmark databases for face recognition, handwritten digit recognition, gender classification, image categorization and action recognition.

4.1 Parameter settings

We first describe how we set the parameters in the SBP_VCP algorithm. A part from the choice of moments and their combinations, two additional parameters need to be set in the calculation of the SBP:
- The spatial support B for calculating local moments.
- The spatial support {P;R} for calculating the LBP.

Although those two parameters are relatively independent, it must be noticed that B has to be sufficiently large to be statistically relevant. Regarding {P;R}, this quantity is supposed to be relatively small in order to represent local micro-structures of the (moment) images.

In the following, due to space constraints, we only show experiments using structuring element B ={(1;5); (2;8)} which provides very satisfactory results on the different datasets. Regarding {P;R}, the spatial support of the LBP, we have considered the three settings commonly found in the literature: {8;1}, {16;2}, and {24;3}.

Regarding the parameters associated with the virtual collaborative projection and the collaborative classification, we used a regularization parameter λ which is initialized as follows, for:
- Face recognition (FR) without occlusion: $\lambda = 0.001$
- Face recognition (FR) with occlusion: $\lambda = 0.1$
- Gender classification (GC): $\lambda = 0.001$
- Digit handwritten recognition: $\lambda = 0.1$
- Image categorization: $\lambda = 0.001$
- Action recognition: $\lambda = 0.1$

4.2 Face recognition (FR)

4.2.1 Extended Yale B database

The Extended Yale B [28], [29] database contains 2,414 frontal face images of 38 individuals; some samples are presented in figure 1. We used the cropped and normalized face images of size 54×48, which were taken under varying illumination conditions. Three tests are considered for this dataset.

Fig. 3 Selected samples from the Extended Yale B database.

Test 1

We randomly split the database into two halves. One half, which contains 32 images for each person, was used as the dictionary, and the other half was used for testing. Table 2 shows the recognition rates versus feature dimension by nearest neighbours NN, nearest feature line NFL [30], support vector machine SVM, sparse representation based classification SRC [18], linear regression based classification LRC [31], locality-constrained linear coding LLC [32], regularized robust coding RRC [7] methods. SBP_VCP achieves the best recognition rate for all dimensions except dimension 300 where it performs slightly worse than RRC_l_1 [7] but it is still superior to all other methods considered.

Table 2 Face recognition results test 1 of different methods on the Extended Yale B database.

Dimension	84	150	300
NN	85.5%	90.0%	91.6%
SVM	94.9%	96.4%	97.0%
LRC[31]	94.5%	95.1%	96.0%
NFL[30]	94.1%	94.5%	94.9%
SRC[18]	95.5%	96.8%	98.3%
LLC[32]	96.4%	97.0%	97.6%
CRC[9]	95.0%	96.3%	97.9%
RRC_l_2[7]	94.4%	97.6%	98.9%
RRC_l_1[7]	98.0%	98.8%	**99.8%**
SBP_VCP	**98.5%**	**99.1%**	99.7%

Test 2

For each subject, *Ntr* samples are randomly chosen as training samples and 32 of the remaining images are randomly chosen as the testing data. Here the images are resized to size 96×84 and the experiment for each *Ntr* runs 10 times. For comparison, we used robust kernel representation with statistical local features SLF-RKR [25] and we used the same features extraction; statistical local features SLF with NN, LRC, SVM, CRC and SRC based methods.

We list in Table 3 the FR performance results, measured as mean recognition accuracy. The proposed algorithm SBP_VCP achieves the best performance when Ntr=5 or 20 and it is the second best method slightly behind SLF-RKR_l_2 when Ntr=10. It can also be noticed that methods based on collaborative representation (e.g., SLF-RKR [25], SLF+CRC, SLF+SRC and original SRC) perform better than other kinds of linear representation methods (e.g., SLF+LRC, SLF+NN).

Table 3 Face recognition results test 2, of different methods on the Extended Yale B database.

N_{tr}	5	10	20
Original SRC[18]	80.0%	91.4%	97.3%
SLF+NN	59.7%	76.8%	89.7%
SLF+LRC	59.0%	78.9%	93.3%
SLF+HISVM	72.0%	91.6%	99.0%
SLF+CRC	83.0%	95.5%	99.2%
SLF+SRC	82.8%	95.5%	99.3%
SLF-RKR_l_1[25]	85.6%	97.4%	99.5%
SLF-RKR_l_2[25]	85.8%	**97.5%**	99.5%
SBP_VCP	**86.3%**	97.0%	**99.6%**

Test 3

In the third test, we randomly selected between 2 and 7 images from each person as training set and used the remaining images as testing set. Similarly, all the samples were projected into a subspace of 550 dimensions (Samples in LDA+SRC and LDA+CRC schemes are projected into a subspace of 37 dimensions), in addition to SRC and CRC we compare our method with JDDLDR [12], FDDL [11] and PDL [56] based approach. The FR results are shown in Table 4.

Table 4 Face recognition results test 3, of different methods on the Extended Yale B database.

N_{tr}	2	3	4	5	6	7
JDDLDR [12]	**54.9%**	65.3%	67.4%	68.2%	69.6%	70.5%
DR-SRC	53.0%	63.6%	65.6%	67.1%	68.9%	69.8%
MFL-SRC	53.4%	63.1%	65.7%	66.8%	69.0%	69.2%
PCA+SRC	53.5%	64.1%	65.2%	67.0%	68.7%	69.0%
LDA+SRC	46.2%	53.2%	60.3%	66.5%	68.1%	68.1%
PCA+CRC	53.2%	64.4%	65.0%	67.1%	68.5%	69.2%
LDA+CRC	46.0%	53.5%	60.9%	66.2%	67.9%	68.2%
FDDL [11]	44.1%	53.8%	63.6%	67.5%	69.3%	70.1%
PDL [56]	49.7%	58.3%	60.2%	62.8%	66.9%	69.4%
SBP_VCP	**54.9%**	**65.8%**	**74.1%**	**80.1%**	**85.4%**	**90.5%**

Table 4 shows that SBP_VCP gives the best results for all values of N_{tr}. We remark that the improvement in performance is significant as compared to all others methods demonstrating the advantages of combining the statistical features with this twin competitive (collaborative) classification.

4.2.2 AR database

Test 1

As in [18], we selected a subset (with only illumination and expression changes) containing 50 male and 50 female subjects from the AR database [27]; some samples are shown in Figure 4. For each subject, the seven images from Session 1 were used for training and the other seven images from Session 2 were used for testing. The images were cropped to 60×43. The FR rates with baseline comparison reported in Table 5 show that the proposed approach yields the best performance among all methods considered for all dimensions, even when the dimension is 30 and competing methods perform rather poorly. As expected, all methods achieve their maximal recognition rates at dimension 300.

Fig. 4 Selected samples from the AR database.

Table 5 Face recognition results test 1, of different methods on the AR database.

Dimension	30	54	120	300
NN	62.5%	68.0%	70.1%	71.3%
SVM	66.1%	69.4%	74.5%	75.4%
LRC[31]	66.1%	70.1%	75.4%	76.0%
NFL[30]	64.5%	69.2%	72.7%	73.4%
SRC[18]	73.5%	83.3%	90.1%	93.3%
LLC[32]	70.5%	80.7%	87.4%	89.0%
CRC[9]	64.2%	80.5%	90.0%	93.7%
RRC_l_2[7]	61.5%	84.3%	94.3%	95.3%
RRC_l_1[7]	70.8%	87.6%	94.7%	96.3%
SBP_VCP	**82.4%**	**93.7%**	**98.9%**	**100%**

Test 2

For each subject, the seven images with illumination change and expressions from Session 1 were used for training, and the other seven images with only illumination change and expression from Session 2 were used for testing. The size of the original face image is 83×60. The recognition rates versus the number of training samples *Ntr* are reported in Table 6, showing that SBP_VCP achieves the highest recognition rates, followed in order by SLF-RKR [25] and SLF+ SRC.

Table 6 Face recognition results test 2, of different methods on the AR database.

N_{tr}	2	3	4	5	6	7
SRC [18]	67.0%	70.1%	77.9%	87.4%	93.7%	93.1%
SLF+NN	88.1%	88.7%	92.3%	97.0%	98.0%	98.3%

SLF+LRC	83.3%	82.7%	85.0%	90.0%	93.7%	94.3%
SLF+HISVM	86.7%	87.0%	90.6%	94.1%	96.6%	96.6%
SLF+CRC	87.9%	87.4%	88.0%	93.9%	98.3%	98.3%
SLF+SRC	87.6%	88.0%	89.9%	95.7%	98.7%	98.8%
SLF-RKR_l_1[25]	90.1%	91.0%	92.4%	97.0%	99.4%	99.4%
SLF-RKR_l_2[25]	90.6%	**91.1%**	92.0%	97.4%	99.4%	99,4%
SBP_VCP	**91.1%**	**91.1%**	**94.4%**	**8.4%**	**100%**	**100%**

4.2.3 MPIE database

The CMU Multi-PIE database [33] contains images of 337 subjects captured in four sessions with simultaneous variations in pose, expression, and illumination. Among these 337 subjects, all the 249 subjects in Session 1 were used for training. To make the FR more challenging, four subsets with both illumination and expression variations in Sessions 1, 2 and 3, were used for testing. We conducted two tests with this experimental protocol.

Test1

In the first test, for the training set, as in [18], we used the 7 frontal images with extreme illuminations {0, 1, 7, 13, 14, 16, and 18} and neutral expression (refer to Fig. 5(a) for examples). For the testing set, 4 typical frontal images with illuminations {0, 2, 7, 13} and different expressions (smile in Sessions 1 and 3, squint and surprise in Session 2) were used (refer to Fig. 5(b) for examples with surprise in Session 2, Fig. 5(c) for examples with smile in Session 1, and Fig. 5(d) for examples with smile in Session 3). Here we used Eigenface with dimensionality 300 as the face feature for sparse coding. Table 7 reports the recognition rates found in four testing sets.
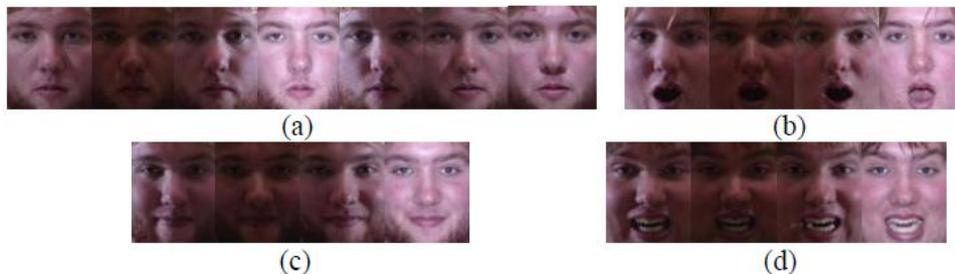

Fig. 5 A subject in the Multi-PIE database. (a) Training samples with only illumination variations. (b) Testing samples with surprise expression and illumination variations. Panels (c) and (d) show the testing samples with smile expression and illumination variations in Session 1 and Session 3, respectively.

Table 7 Face recognition results of different methods on the MPIE database.

Algorithms	Smi-S1	Smi-S3	Sur-S2	Squ-S2
NN	88.7%	47.3%	40.1%	49.6%
SVM	88.9%	46.3%	25.6%	47.7%
LRC[31]	89.6%	48.8%	39.6%	51.2%
NFL[10]	90.3%	50.0%	39.8%	52.9%
SRC[18]	93.7%	60.3%	51.4%	58.1%
LLC[32]	95.6%	62.5%	52.3%	64.0%
CRC[9]	90.3%	54.6%	41.1%	47.9%
RRC_l_2[7]	96.1%	70.2%	59.2%	58.1%

RRC_l_1[7]	97.8%	**76.0%**	**68.8%**	65.8%
SBP_VCP	**98.2%**	72.7%	62.5%	**69.7%**

Table 7 shows that SBP_VCP gives the best results using the sets smile-S1 and Squint-S2 and the second best results with the sets surprise-S2 and smile-S3. Since smile-S1 is in the same class (intra-class) as the training set, that's why we have a good result, regarding smile-S3 and surprise-S2 sets we have the second best accuracy by 72.7% and 62.5% respectively.

Test2

In the second test, we analyzed the impact of statistical binary pattern (SBP) on different state-of-the-art methods with the same experimental protocol as Test1. We considered nearest neighbours NN, linear regression LRC [31], sparse representation SRC [18], collaborative representation CRC [9] and relaxed collaborative representation RCR [10] based classification. Table 8 reports the recognition rates found on the different methods with and without SBP.

Table 8 Face recognition results of different methods with SBP on the MPIE database.

Algorithms	Smi-S1	Smi-S3	Sur-S2	Squ-S2
NN	88.7%	47.3%	40.1%	49.6%
SBP-NN	**94.5%**	**58.1%**	**51.0%**	**63.4%**
LRC[31]	89.6%	48.8%	39.6%	51.2%
SBP-LRC	**96.5%**	**69.9%**	**57.9%**	**64.1%**
SRC[18]	93.7%	60.3%	51.4%	58.1%
SBP-SRC	**98.0%**	**72.1%**	**62.2%**	**67.2%**
CRC[9]	90.3%	54.6%	41.1%	47.9%
SBP-CRC	**97.4%**	**61.7%**	**59.2%**	**64.2%**
RCR[10]	89.6%	48.5%	38.1%	40.0%
SBP-RCR	**96.2%**	**69.1%**	**64.5%**	**74.6%**

Results in Table 8 show that SBP consistently increases the performance of different approaches, especially when the classes are different from session 1. The improvement in performance is significant for collaborative classification based methods CRC and RCR; for example the recognition rate of RCR with the set square-S2 increases from 40% to 74.6%, and with the set surprise-S2 from 38.1% to 64.5%.

4.2.4 AR database, disguise

In this experiment, we considered a subset from the AR database consisting of 2,599 images from 100 subjects (26 samples per class except for a corrupted image w-027-14.bmp), 50 males and 50 females. We performed three tests: the first one follows the experimental settings in [18]; the other two, described below, are more challenging. The images were resized to 83×60 in the first and third test and to 42×30 in the second test; four representative samples of two persons are shown in figure 6.

Fig. 6 Testing samples with sunglasses and scarves from theAR database.

Test1

In the first test, 799 images (about 8 samples per subject) of non-occluded frontal views with various facial expressions in Sessions 1 and 2 were used for training, while two separate subsets (with sunglasses and scarf) of 200 images (1 sample per subject per Session, with neutral expression) were used for testing. The FR results are listed in Table 9 and show that the SBP_VCP method achieves a much higher recognition rates than CRC_RLS [9], RRC [7] (with scarf), SRC [18], Gabor Feature based Sparse Representation with Gabor Occlusion Dictionary GSRC [5] and Maximum correntropy criterion CESR [8].

Table 9 Test 1: Face recognition results using images with real disguise from the AR database.

Algorithms	Sunglass	Scarf
SRC [18]	87.0%	59.5%
GSRC[5]	93.0%	79.0%
CESR[8]	99.0%	42.0%
CRC_RLS [9]	68.5%	90.5%
RRC_l_2[7]	99.5%	96.5%
RRC_l_1[7]	**100%**	97.5%
SBP_VCP	**100%**	**99.5%**

Test 2

In the second test, we considered FR with a more complex disguise including variations of illumination and longer data acquisition interval. 400 images (4 neutral images with different illuminations per subject) of non-occluded frontal views in Session 1 were used for training, while the disguised images (3 images with various illuminations and sunglasses or scarves per subject per Session) in Sessions 1 and 2 for testing. The results, reported in Table 10, show that the SBP_VCP methods achieves better performance than CRC_RLS [9], SRC [18], GSRC [5] and CESR [8], except for sunglass-S1, where it achieve the second best result after RRC [9].

Table 10 Test 2: Face recognition results using images with real disguise from the AR database.

Algorithms	Session 1		session 2	
	Sunglass	Scarf	Sunglass	Scarf
SRC [18]	89.3%	32.3%	57.3%	12.7%
GSRC [5]	87.3%	85.0%	45.0%	66.0%
CESR [8]	95.3%	38.0%	79.0%	20,7%
CRC_RLS [9]	66.3%	62.0%	29.0%	42.0%
RRC_l_2[7]	**99.0%**	94.7%	84.0%	77.3%
RRC_l_1[7]	**99.0%**	93.3%	89.0%	76.3%

SBP_VCP	98.7%	**98.7%**	**89.7%**	**84.7%**

Test 3

In this test, a subset of 50 males and 50 females were selected from the AR database. For each subject, 7 samples without occlusion from session 1 are used for training, with all the remaining samples with disguises used for testing. These testing samples (including 3 samples with sunglass in Session1, 3 samples with sunglass in Session 2, 3 samples with scarf in Session 1 and 3 samples with scarf in Session 2 per subject) not only have disguises, but also variations of time and illumination. Table 11 reports the FR results on the four test sets with disguise.

Table 11 Test 3: Face recognition results using images with real disguise from the AR database.

Algorithms	Sunglass-S1	Scarf-S1	Sunglass-S2	Scarf-S2
Robust SRC[18]	83.3%	48.7%	49.0%	29.0%
RSC [6]	94.7%	91.0%	80.3%	72.7%
SLF+NN	98.7%	98.0%	82.3%	88.7%
SLF+LRC	96.7%	92.0%	68.7%	68.7%
SLF+HISVM	97.0%	95.7%	70.3%	78.7%
SLF+CRC	99.7%	98.7%	80.3%	86.7%
SLF+KCRC	**100%**	98.3%	82.7%	88.0%
SLF+SRC	**100%**	99.0%	85.0%	90.7%
SLF+KSRC	**100%**	98,3%	84.0%	86.7%
SLF_RKR_l_1[25]	**100%**	**100%**	93.0%	**97.6%**
SLF_RKR_l_2[25]	**100%**	**100%**	91.3%	96.0%
SBP_VCP	**100%**	99.3%	**97.0%**	97.0%

Table 11 shows that the proposed method achieves the best recognition rate with sunglasses in Session 2 and achieves 100% accuracy with Session 1 (as some others methods) and the second best accuracy in the sessions with scarf (SLF_RKR is ranked first). We remark that all methods perform better for session 1 (sunglass and scarf) than session 2, as session 2 is more challenging due to variations in illumination.

4.2.5 Georgia Tech data base with block occlusion

The Georgia Tech (GT) [51] Face Database contains 750 color images of 50 subjects (15 images per subject), as shown in Figure 7(a). These images have large variations in pose and expression and some illumination changes. Images were converted to gray scale, cropped and resized to 90×68. The first eight images of all subjects were used in the training (400 images), the remaining seven images for testing (350 images). For block occlusion, were placed a randomly located rectangle of all the testing images using an unrelated image, as illustrated in Figure 7(c).
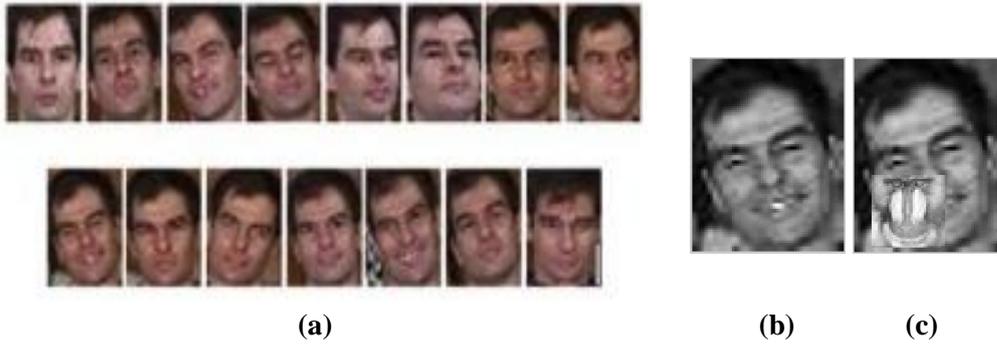
Fig. 7 (a) Original images of the same subject from Georgia Tech. (b) Original test image. (c)Test image with random block occlusion (30%).

Performance results reported in Table 12 compare the algorithms SBP_VCP, SBP-CRC, SBP-SRC, SBP-LRC, and SBP-NN in the presence of block occlusion ranging from 0% to 50% of the image. Table 12 shows that SBP_VCP achieves the best accuracy. Our interpretation is that this remarkable performance is due mostly to the VCP approach which efficiently takes advantage of the twin collaborative representation in the training and testing steps.

Table 12 Face recognition results using the GT databasewith block occlusion.

Occlusions (%)	0	10	20	30	40	50
SBP-NN	48.0%	28.9%	18.8%	10.6%	7.1%	5.1%
SBP-LRC	64.0%	62.8%	58.5%	48.6%	39.1%	26.9%
SBP-SRC	66.8%	64.3%	60.6%	55.1%	46.0%	32.2%
SBP-CRC	66.5%	63.1%	60.6%	57.3%	49.4%	34.3%
SBP_VCP	**67.1%**	**66.3%**	**61.4%**	**58.6%**	**51.1%**	**37.1%**

4.2.6 FRGC data base with block occlusion and single sample per person (SSPP)

The FRGC database [50] contains faces acquired under uncontrolled conditions as shown in Figure 8(a). Using single sample per person (SSPP) protocol as another challenging problem in FR, we randomly selected 152 images for training, 152 images for testing and replaced a randomly located block of the test image with an unrelated image, as illustrated in Figure 8(c). The images were cropped and resized to 90×68 pixels. The recognition accuracy on this dataset is reported in Table 13.

Fig. 8 (a) Original images of four different subjects from FRGC. (b) Original test image. (c)Test image with random block occlusion (30%).

The table 13 shows that also in this test with block occlusion ranging from 10% to 50% of the image our algorithm SBP_VCP achieves the best performance, as it exhibits as lightly better accuracy than all the other methods considered. Note that all methods, except SBP-NN and SBP-LRC, achieve the same recognition rates without occlusion, while their performance is different in the presence of occlusion. This shows that SBP_VCP performs remarkably well in the challenging SSPP problem.

Table 13 Face recognition results of different methods with block occlusion and SSPP using the FRGC database.

Occlusions (%)	0	10	20	30	40	50
SBP-NN	74.3%	69.1%	56.8%	42.4%	25.7%	11.2%
SBP-LRC	82.2%	80.9%	75.6%	71.1%	62.5%	45.4%
SBP-SRC	**83.5%**	80.3%	77.6%	68.4%	53.9%	38.2%
SBP-CRC	**83.5%**	80.3%	76.9%	68.4%	61.2%	45.1%
SBP_VCP	**83.5%**	**83.5%**	**78.2%**	**71.7%**	**63.8%**	**46.1%**

4.3 Gender classification (GC)

4.3.1 AR database

We selected a non-occluded subset (14 images per subject) of AR [16] consisting of 50 male and 50 female subjects. Images of the first 25 males and 25 females were used for training and the remaining images were used for testing. The images were cropped to 60×43. PCA was used to reduce the dimension of each image to 300. Table 14 reports the comparison of SBP_VCP versus the methods: Regularized Nearest Subspace (RNS) [34], Multi-Regularized features Learning (MRL) [35], CRC_RLS [9], SRC [18], SVM, LRC [31] and NN. The table 14 shows that SBP_VCP outperforms the others methods considered and illustrates that the proposed method based on statistical local features is very effective for gender classification.

Table 14 Performance results on GC using the AR database.

SBP_VCP	RNS_l_1[34]	RNS_l_2[34]	MRL [35]	CRC_RLS [9]	SRC[18]	SVM	LRC[31]	NN
97.81%	94.90%	94.90%	92.83%	93.70%	92.30%	92.40%	27.30%	90.70%

4.3.2 FEI database

There are 14 images for each of 200 individuals with a total of 2800 images [36]. The number of male and female subjects is exactly the same and equal to 100. The first nine images of all subjects are used in the training (1800 images, 900 per gender) and the remaining five images serve as testing images (1000 images, 500 per gender). Figure 9 shows all samples from one person. The images were cropped to 60×43.

Fig. 9 All samples from the same person from FEI database.

Here we compare SBP_VCP to the MRL [35] and CRC_RLS [9] algorithms on different dimensionality. Table 15 shows that SM_VCP outperforms MRL and CRC_RLS with all dimensionality except for dimension 30.

Table 15 Performance results on GC using the FEI database.

Dimension	30	54	120	300
CRC_RLS [9]	88.2%	90.3%	91.4%	93.1%
MRL [35]	**93.7%**	93.4%	94.1%	94.0%
SBP_VCP	92.6%	**93.8%**	**95.0%**	**96.9%**

4.4 Handwritten digit recognition

We next considered the problem of handwritten digit recognition on the widely used USPS database (Hull, J.J. 1994), which has 7,291 training and 2,007 test images.We used two different values of N_{tr}: 100 and 300 images. Results in the Table 16 below show that SM_VCP outperforms all competing methods considered when N_{tr} is 300 images. When $N_{tr} =$ 100, fisher discrimination dictionary learning FDDL [11] is the best performing algorithm but our approach has the second best performance.

Table 16 Handwritten digit recognition results of different methods on the USPS database.

N_{tr}	100	300
FDDL [11]	94.1%	94.1%
Simplified FDDL [37]	**94.2%**	95.0%
CRC_RLS [9]	89.8%	90.6%
SBP-CRC	90.3%	92.2%
SBP_VCP	93.4%	**95.1%**

4.5 Image categorization

We tested the proposed method on the problem of multi-class object categorization. We used one of the two Oxford flower datasets, 17 category data set, [38], some samples of which are show in Figure 10. We adopt the default experimental settings provided at the website www.robots.ox.ac.uk/˜vgg/data/flowers, including the training, validation, test splits and the multiple features. It should be noted that, in this setting, features are only extracted from those flower regions which are well cropped by segmentation. This set contains 17 species of

flowers with 80 images per class. As in [40], we directly use the χ^2 distance matrices of seven features (i.e., HSV, HOG, SIFTint, SIFTbdy, color, shape and texture vocabularies) as inputs, and perform the experiments based on the three predefined training, validation, and test splits. Performance results (in terms of accuracy) comparing VCP vs. other state-of-the-arts are presented in Table 17 and show that VCP slightly outperforms all other methods. Note that, as we follow [40], we did not use the SBP for the representation in this test.

Fig. 10 Samples from Oxford flower data sets with 17 categories.

Table 17 Categorization accuracy on the 17 category Oxford Flowers data set.

Methods	Accuracy (%)
SRC combination	85.9 ± 2.2
MKL [46]	85.2 ± 1.5
CG-Boost [47]	84.8 ± 2.2
LPBoost[47]	85.4 ± 2.4
MTJSRC-RKHS [40]	88.1 ± 2.3
MTJSRC-CG [40]	88.9 ± 2.9
RCR-DK [10]	87.6 ± 1.8
RCR-CG [10]	88.0 ± 1.6
VCP	**89.1 ± 0.9**

4.6 Action Recognition

Finally, we conducted an experiment of action recognition on the UCF sport action dataset (Rodriguez et al. [43]) and the large scale UCF50 dataset. The video clips in the UCF sport action dataset were collected from various broadcast sports channels (e.g., BBC and ESPN). There are 140 videos in total and their action bank features can be found in Sadanand et al. [41]. The videos cover 10 sport action classes: driving, golfing, kicking, lifting, horse riding, running, skateboarding, swinging-(pommel horse and floor), swinging-(high bar) and walking. The UCF50 dataset has 50 action categories such as baseball pitch, biking, driving, skiing (figure 11), and there are 6,680 realistic videos collected from YouTube.

On the UCF sport action dataset, we followed the experimental settings in Rodriguez et al. [43] and evaluated VCP via five-fold cross validation, where one fold is used for testing and the remaining four folds for training. Since we use the action bank features of [41], we do not use SBP as a local feature in this test.

Fig. 11 UCF Sports Dataset: sample frames of 10 action classes along with their bounding box annotations of the humans shown in yellow.

We compared VCP against state-of-the-art methods and reported the recognition rate in Table 18. Again, results show that VCP performs very competitively, illustrating the impact of the collaborative method.

Table 18 Recognition accuracy on the UCF Sports data set.

Methods	Accuracy
Hough forest (data A) [42]	86.6%
Hough forest (data B) [42]	81.6%
Hough forest (data C) [42]	79.0%
Rodriguez et al. [43]	69.2%
Yeffet & Wolf [44]	79.2%
Wang et al. [45]	85.6%
VCP	**88.8%**

4.7 Running time

In practical applications, training is usually an offline stage while recognition (classification) is usually an online step. Since we adopted the same classification procedure of collaborative representation based classification CRC, the speed-up we achieve is remarkable when compared to many other methods due to the significant reduction in computational complexity. In fact, after projecting a query sample y via $P = \left(X^T X + \lambda I\right)^{-1} X^T$, y is classified to the class which gives the minimal $r_i(\alpha) = \|y - X_i\alpha\|_2^2 + \lambda\|\alpha\|_n$ where $n = 1$ or 2 and α_i is the coding vector associated with class i ($\alpha = [\alpha_1,...,\alpha_i,...]$ and $y \approx X\alpha$).

All experiments were carried out using MATLAB on a 2.20 GHz with Dual-core CPU machine with 3.00 GB RAM. Table 19 lists the average computational cost of training step on Test1 and Test2 from the AR dataset with real face disguise. The comparison of the LBP [16] to SBP algorithms shows that LBP has the least computation time, but SBP is close.

Table 19 Average running time (seconds) of training step using AR dataset with real face disguise.

Algorithms	Test 1	Test 2
LBP [16]	**0.02**	**0.005**
SBP	0.03	0.014

Table 20 lists the average computational cost classification of different methods on Test1 and Test2 from the AR dataset with real face disguise. SBP_VCP has the least computation time followed by RRC while GSRC has the highest computation time.

Table 20 Average running time (seconds) of competing methods using AR dataset with real face disguise.

Algorithms	Test 1-sunglass	Test 1-scarf	Test 2-sunglass	Test 2-scarf
CESR[8]	2.50	3.61	0.45	0.47
SRC [18]	13.98	13.73	2.34	2.35
GSRC[5]	119.32	118.05	12.95	12.49
RRC[7]	2.17	2.04	0.23	0.23
CRC [9]	**0.13**	**0.17**	**0.04**	**0.04**
SBP_VCP	**0.13**	**0.17**	**0.04**	**0.04**

5 Conclusion

In this paper, we have introduced a novel approach for pattern recognition combining high order statistical binary pattern and collaborative projection for robust local representation and classification. We have demonstrated that the extraction of statistical features based on the high-order moments of the images is particularly effective against images outliers. When this is property is combined with our strategy for competitive or collaborative representation based on a trained virtual projection, we obtain a method we call SBP_VCP which is a powerful refinement of the collaborative representation based classification recently proposed in the literature. We have validated SBP_VCP on a wide range of problems from pattern recognition and classification which include face recognition, gender classification, object categorisation and action recognition. Extensive numerical tests and detailed comparison with standard and state-of-the-art methods demonstrate that the proposed SBP_VCP approach performs very competitively even on challenging classification tests. Additionally, our method can be implemented at a relatively small computational cost as it relies on the same efficient framework used in CRC for the classification step.

References

1. Borgi MA, Labate D, El'arbi M, Amar CB (2015) Sparse multi-stage regularized feature learning for robust face recognition. Expert Syst. Appl. 42(1): 269-279

2. Turk M, Pentland A (1991) Eigenfaces for recognition. J. Cognitive Neuroscience 3(1): 71-86

3. Belhumeur PN, Hespanha JP, Kriengman DJ (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19(7): 711-720

4. Theodorakopoulos I, Rigas I, Economou G, Fotopoulos S (2011) Face recognition via local sparse coding. In: Proceedings of the ICCV: 1647–1652.

5. Yang M, Zhang L (2010) Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary. In: Proceedings of the ECCV: 448-461

6. Yang M, Zhang L, Yang J, Zhang D (2011) Robust sparse coding for face recognition. In: Proceedings of the ICCV: 625-632

7. Yang M, Zhang L, Yang J, Zhang D (2013) Regularized Robust Coding for Face Recognition. IEEE Transactions on Image Processing 22(5): 1753-1766

8. He R, Zheng WS, Hu BG (2011) Maximum correntropy criterion for robust face recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 33(8): 1561-1576

9. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: Which helps face recognition? In: Proceedings of the ICCV: 471-478

10. Yang M, Zhang L, Zhang D, Wang S (2012) Relaxed collaborative representation for pattern classification. In: Proceedings of the ICCV: 2224-2231

11. Yang M, Zhang L, Feng X, Zhang D (2011) Fisher discrimination dictionary learning for sparse representation. In: Proceedings of the ICCV: 543-550

12. Feng Z, Yang M, Zhang L, Liu Y, Zhang D (2013) Joint discriminative dimensionality reduction and dictionary learning for face recognition. Pattern Recognition 46(8): 2134-2143

13. Lades M, Vorbrüggen JC, Buhmann J et al (1993) Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers 42(3): 300-311

14. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Processing 11(4): 467-476

15. Shen L, Bai L (2006) A review on Gabor wavelets for face recognition. Pattern Analysis and Application 9(10): 273-292

16. Timo A, Abdenour H, Matti P (2004) Face recognition with local binary patterns. In: Proceedings of the ECCV: 469-481

17. Ojala T, Pietikäinen M, Mäenpää T (2002) Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns.IEEE Trans. Pattern Anal. Mach. Intell. 24(7): 971-987

18. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation.IEEE Trans. Pattern Analysis and Machine Intelligence 31(2): 210–227

19. Zhang W, Shan S, Gao W et al (2005) Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Proceedings of the ICCV: 786-791

20. Zhang W, Shan S, Chen X, Gao W (2009) Are gabor phases really useless for face recognition?. Pattern Analysis and Application 12(3): 301-307

21. Zhang B, Shan S, Chen X, Gao W (2007) Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Trans. Image Processing 16(1): 57-68

22. Xie SF, Shan SG, Chen XL, Chen J (2010) Fusing local patterns of gabor magnitude and phase for face recognition. IEEE Trans. Image Processing 19(5): 1349-1361

23. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12): 2037-2041

24. Borgi MA, El'arbi M, Labate D, Amar CB (2015) Regularized directional feature learning for face recognition. Multimedia Tools Appl. 74(24): 11281-11295

25. Yang M, Zhang L, Shiu SC, Zhang D (2013) Robust Kernel Representation With Statistical Local Features for Face Recognition. IEEE Trans. Neural Netw. Learning Syst. 24(6): 900-912

26. Nguyen TP, Vu NS, Manzanera A (2016) Statistical binary patterns for rotational invariant texture classification. Neurocomputing 173: 1565–1577

27. Martinez A, and Benavente R (1998) The AR face database. CVC Tech. Report 24

28. Georghiades A, Belhumeur P, Kriegman D (2001) From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE PAMI 23(6): 643–660

29. Lee K, Ho J, Kriegman D (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE PAMI 27(5): 684–698

30. Li SZ, Lu J (1999) Face recognition using nearest feature line method. IEEE Trans. Neural Network 10(2): 439-443

31. Naseem I, Togneri R, Bennamoun M (2010) Linear regression for face recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 32(11): 2106-2112

32. Wang JJ, Yang JC et al (2010) Locality-constrained Linear Coding for Image Classification. In: Proceedings of the CVPR: 3360-3371

33. Gross R, Matthews I et al (2010) Multi-PIE. Image and Vision Computing 28(5): 807–813

34. Zhang L, Yang M et al (2011) Collaborative Representation based Classification for Face Recognition. Technical report. arXiv: 1204.2358

35. Borgi MA, El'arbi M, Labate D, Amar CB (2014) Face, gender and race classification using multi-regularized features learning. In: Proceedings of the ICIP: 5277-5281

36. Thomaz E, Giraldi GA (2010) A new ranking method for Principal Components Analysis and its application to face image analysis. Image and Vision Computing 28(6): 902-913

37. Yang M, Zhang L, Feng X, Zhang D (2014) Sparse Representation Based Fisher Discrimination Dictionary Learning for Image Classification. International Journal of Computer Vision 109(3): 209-232

38. Nilsback M, Zisserman A (2006) A visual vocabulary for flower classification. In: Proceedings of the CVPR: 1447-1454

39. Gao S, Tsang I, Chia L (2010) Kernel sparse representation for image classification and face recognition. In: Proceedings of the ECCV (4): 1-14

40. Yuan XT, Yan SC (2010) Visual classification with multitask joint sparse representation. In: Proceedings of the CVPR: 3493-3500

41. Sadanand S, Corso JJ (2012) Action bank: A high-level representation of activity in video. In: Proceedings of the CVPR: 1234-1241

42. Yao A, Gall J, Van Gool LJ (2010) A Hough transform-based voting framework for action recognition. In: Proceedings of the CVPR: 2061-2068

43. Rodriguez MD, Ahmed J, Shah M (2008) Action MACH a spatio-temporal maximum average correlation height filter for action recognition. In: Proceedings of the CVPR

44. Yeffet L, Wolf L (2009) Local trinary patterns for human action recognition. In: Proceedings of the ICCV: 492-497

45. Wang H, Ullah MM et al (2009) Evaluation of local spatio-temporal features for action recognition. In: Proceedings of the BMVC: 1-11

46. Nilsback M, Zisserman A (2008) Automated flower classification over a large number of classes. In: Proceedings of the ICCVGIP: 722-729

47. Gehler P, Nowozin S (2009) On feature combination for multiclass object classification. In: Proceedings of the ICCV: 221-228

48. Cai S, Zhang L, Zuo W, Feng X (2016) A probabilistic collaborative representation based approach for pattern classification.In CVPR: accepted

49. Borgi MA, Labate D, El'Arbi M, Amar CB (2014) Regularized Shearlet network for face recognition using single sample per person. In: Proceedings of the ICASSP: 514-518

50. Phillips PJ, Flynn PJ et al (2005) Overview of the face recognition grand challenge. In: Proceedings of the CVPR: 947-954

51. Georgia Tech Face Database (2007). http://www.anefian.com/face_reco.htm

52. Ojala T, Maenpaa T, Pietikainen M et al (2002) Outex—new framework for empirical evaluation of texture analysis algorithms. In: Proceedings of the ICPR: 701–706

53. Guo ZH, Zhang L, Zhang D (2010) A completed modeling of local binary pattern operator for texture classification. IEEE Trans. Image Process. 19 (6): 1657–1663

54. Xu Y, Zhang D, Yang J, Yang J-Y (2011) A two-phase test sample sparse representation method for use with face recognition. IEEE Transactions on Circuits and Systems for Video Technology 21(9): 1255-1262

55. Mi J-X, Liu J-X (2013) Face Recognition Using Sparse Representation-Based Classification on K-Nearest Subspace. PLoS ONE 8(3): e59430. doi:10.1371/journal.pone.0059430

56. Gu S, Zhang L, Zuo W et al (2014) Projective Dictionary Pair Learning for Pattern Classification. In: Proceeding of advances in Neural Information Processing Systems: 793-801

57. Cai S, Zuo W, Zhang L (2014) Support Vector Guided Dictionary Learning. In: Proceedings of the European Conference on Computer Vision (4): 624-639

58. Li Z, Lai Z, Xu Y et al (2015) A Locality-Constrained and Label Embedding Dictionary Learning Algorithm for Image Classification. IEEE Transactions on Neural Networks and Learning Systems. doi: 10.1109/TNNLS.2015.2508025