A new Sparse Coding Approach for Human Face and Action Recognition

Mohsen Nikpour*
Department of Electrical and Computer Engineering, Babol Noushirvani University of Technology, Babol, Iran
Nikpour@mit.ac.ir
Mohammad Reza Karami Molaei
Department of Electrical and Computer Engineering, Babol Noushirvani University of Technology, Babol, Iran
Mkarami@nit.ac.ir
Reza Ghaderi
Department of nuclear Engineering, Shahid Beheshti University of Tehran, Tehran, Iran
R_ghaderi@sbu.ac.ir

Received: 27/Jul/2016 Revised: 07/Jan/2017 Accepted: 14/Jan/2017

Abstract

Sparse coding is an unsupervised method which learns a set of over-complete bases to represent data such as image, video and etc. In the cases where we have some similar images from the different classes, using the sparse coding method the images may be classified into the same class and devalue classification performance. In this paper, we propose an Affine Graph Regularized Sparse Coding approach for resolving this problem. We apply the sparse coding and graph regularized sparse coding approaches by adding the affinity constraint to the objective function to improve the recognition rate. Several experiments has been done on well-known face datasets such as ORL and YALE. The first experiment has been done on ORL dataset for face recognition and the second one has been done on YALE dataset for face expression detection. Both experiments have been compared with the basic approaches for evaluating the proposed method. The simulation results show that the proposed method can significantly outperform previous methods in face classification. In addition, the proposed method is applied to KTH action dataset and the results show that the proposed sparse coding approach could be applied for action recognition applications too.

Keywords: Sparse Coding; Manifold Learning; Graph Regularization; Affinity; Image Representation; Image Classification.

1. Introduction

Image classification is a significant task in image processing and computer vision studies. Sparse coding method can represent images using a few active coefficients [1]. Accordingly, interpreting and applying the sparse representations are easy and simplify efficient content-based image indexing and retrieval [2]. The authors in [5] have been proposed an improved CRC-RLS method for solving the poor robustness of Collaborative Representation based Classification with Regularized Least Square algorithm.

The range of sparse coding methods have been widened every day in many fields such as pattern recognition, machine learning and signal processing [3], face recognition [4,5], image classification [7,8] and action recognition [10] in recent years. Most important targets of sparse coding is the maximum signal fidelity preservation and also improving the quality of the sparse representation. To achieve these targets, many works have been done to modify the sparsity constraint. In [7] the authors to improve the sparse coding method, have added a nonnegative constraint to the objective function of basis sparse coding method. The authors in [8] have proposed a face recognition method based on the discriminative locality preserving vectors. This method is based on the discriminant analysis on the locality preserving vectors. A robust sparse coding to improve the signal fidelity has been proposed in [9]; however, in the case of the similar images, they may be transformed into identical visual words of the codebook and encoded with the same representations. The dictionary learned from the images cannot effectively encode manifold structure of the images face in this case, and the similar images from different classes may be classified in the same class accordingly. This similarity will highly challenge the robustness of existing sparse coding algorithms for image classification problems such as face images. Similar face images are lying on a manifold structure and the face images from different classes are lying on different manifold structures [10]. It has been shown that if the geometrical structure is used and the local invariance is considered, the learning performance can be significantly improved. Recently, many literatures have focused on manifold learning problems, which represent the samples from the different manifold structures. To preserve the geometrical information of the data, the authors in [11] proposed to extract a good feature representation through which the manifold structure of data is spotted.

Regarding the recent progress in sparse coding and manifold learning, we propose a novel Affine regularized sparse coding algorithm to construct robust sparse representations for classifying similar face images

* Corresponding Author

accurately. Specifically, the objective function of sparse coding has incorporated this criterion to make the new representations of the similar face images far from each other. Moreover, to improve the objective function with more discriminating power in data representation, we also incorporate the graph Laplacian term of coefficients [8] in our objective function. For more consideration, the proposed method is applied on a well-known human action recognition dataset. The experimental results verify the effectiveness of our sparse coding approach.

This paper is continued as follows: In Section 2, some related works are introduced. The sparse coding and graph regularized sparse coding is then described in Section 3. Section 4 contains the proposed method. The experiment setup and results on face and action datasets are indicated in Section 5 and consequently, some conclusions and future work are presented in Section 6.

2. Related Work

In this section, we discuss some prior papers in sparse coding and manifold learning area. In recent years, sparse coding has been widely used in many fields in computer vision. The authors in [3] proposed a feature sign search method. This method reduces the non-differentiable problem to an unconstrained quadratic programming (QP). This problem can be solved rapidly by the optimization process. Our work also uses their method to solve the proposed optimization problem. For adapting the dictionary to achieve sparse representation, the authors in [12] proposed a K-SVD method to learn the dictionary using orthogonal matching pursuit or basis pursuit. Adding nonnegative term to the sparse constraint is a method to improve the quality of sparse representations [7]. The other methods such as graph regularization [5,8] and using weighted ℓ2-norm constraint are also introduced for improving the sparse representation. In the machine learning literature, manifold learning has also attracted extensive research interest. The authors in [8] proposed a graph based algorithm, called Graph regularized Sparse Coding (GraphSC), to give sparse representations that well consider the local manifold structure of the data. By using graph laplacian as a smooth operator, the obtained sparse representations vary smoothly along the geodesics of the data manifold. Our work in addition to the affinity constraint, incorporates the graph laplacian term of coefficients [8] in the objective function, and can discover more discriminating representations for image classification.

3. Preliminaries

This section introduces sparse coding and affine graph regularized sparse coding.

3.1 Sparse Coding

Assuming a data matrix $Y = [y_1, ..., y_n] \in R_{m\times n}$ where n is the number of samples in the m-dimensional feature space. Let $\Phi = [\varphi_1, ..., \varphi_k] \in R_{m\times k}$ be the dictionary matrix where each column φ_i represents a basis vector in the dictionary, and $X = [x_1, ..., x_n] \in R_{k\times n}$ be the coding matrix where each column $\boldsymbol{x_i}$ is a sparse representation for a data point $\boldsymbol{y_i}$. Assuming the reconstruction error for a data point follows a zero-mean Gaussian distribution with isotropic covariance, while taking a Laplace prior for the coding coefficients and a uniform prior for the basis vectors, then the maximum posterior estimate of Φ and X given Y is reduced to:

$$\min_{\Phi,X} \|Y-\Phi X\|_F^2 + \alpha \sum_{i=1}^{n} |x_i| \quad st. \|\varphi_j\|^2 \le c, \forall i=1,2,...,k \tag{1}$$

In the above equation α is a parameter for regularizing the level of sparsity of the obtained codes and the approximation of initial data. The objective function in (1) is not convex in Φ and X, therefore solving the above equation is not easy in this case. But it is convex in either Φ or X. Therefore, solving this problem is done by alternatively optimizing Φ while fixing X and vice versa. As a result, the above mentioned problem can be split into two reduced least squares problems: an $\ell 1$-regularized and an $\ell 2$-constrained, both of which can be solved efficiently by existing optimization software [3,4].

3.2 Graph Regularized Sparse Coding

The authors in [8] have proposed a method called Graph Regularized Sparse Coding (GraphSC) method, which considers the manifold assumption to make the basis vectors with respect to the intrinsic geometric structure underlying the input data. This method assumes that if two data points y_i and y_j are close in the intrinsic geometry of data distribution, then their codes φ_i and φ_j are also close. Consider a set of n-dimensional data points $\{y_1, ..., y_n\}$, GraphSC constructs a p-nearest neighbor graph G with n vertices each representing a data point. Let W be the weight matrix of G, if y_i is among the p-nearest neighbor of y_j, $W_{i,j}$ =1; otherwise, $W_{i,j}$ =0. $d_i = \sum_{j=1}^{n} W_{i,j}$, $D = diag(d_1, \dots, d_n)$ and graph Laplacian $L = D-W$. A reasonable criterion for preserving the geometric structure in graph G is to minimize:

$$\frac{1}{2}\sum_{i,j=1}^{n} \|x_i - x_j\|^2 W_{ij} = \mathrm{Tr}(XLX^T) \tag{2}$$

By replacing the result into (1) the GraphSC [7] is obtained:

$$\min_{\Phi,X} \|Y-\Phi X\|_F^2 + \gamma \mathrm{Tr}(XLX^T) + \alpha \sum_{i=1}^{n} |x_i| \quad st. \|\varphi_i\|^2 \le c, \; i = 1, \dots, k \tag{3}$$

In (3) γ is a parameter for regularizing the weight between sparsity of the obtained codes and preserving the geometrical structure.

4. The Proposed Method: Affine Graph Regularized Sparse Coding

In this section, we present the Affine graph regularized sparse coding algorithm for robust image representation, which extends GraphSC by taking into account the affinity constraints on the samples.

4.1 Problem Definition

In linear sparse coding, a collection of k atoms $\varphi_1, \varphi_2, \dots, \varphi_k$ is given that forms the columns of the overcomplete dictionary matrix Φ. With a *l0*-minimization problem, the sparse codes of a feature vector $y \in R^m$ can be determined:

$$\min_{w \in R^m} \|w\|_0, \quad \text{s.t. } x = G_\Phi(w) \tag{4}$$

Where the function G_Φ is defined as $G_\Phi(w) = \Phi w$. In the proposed method the main technical difficulty is the proper interpretation of the function $G_\Phi(w)$ in the manifold setting, where the atoms $\varphi_1, \varphi_2, \dots, \varphi_k$ are now points in M and Φ now denotes the set of atoms, and because of the nonlinearity property in this case, it is no longer possible to create a matrix with atoms. Moving to the more general manifold setting, we have forsaken the vector space structure in R^m. In the linear sparse coding, each point is considered as a vector whose definition requires a reference point. However, in the affine graph regularized sparse coding approach, each point cannot be considered as a vector and therefore, must be considered as a point. This particular viewpoint is the main source of differences between linear and the proposed sparse coding.

In this paper, a new method is proposed to modify the usual notion of sparsity by adding an affine constraint to reduce the feature vectors dimension on a manifold. A vector y is defined as an affine sparse vector if it can be written as follows [13]:

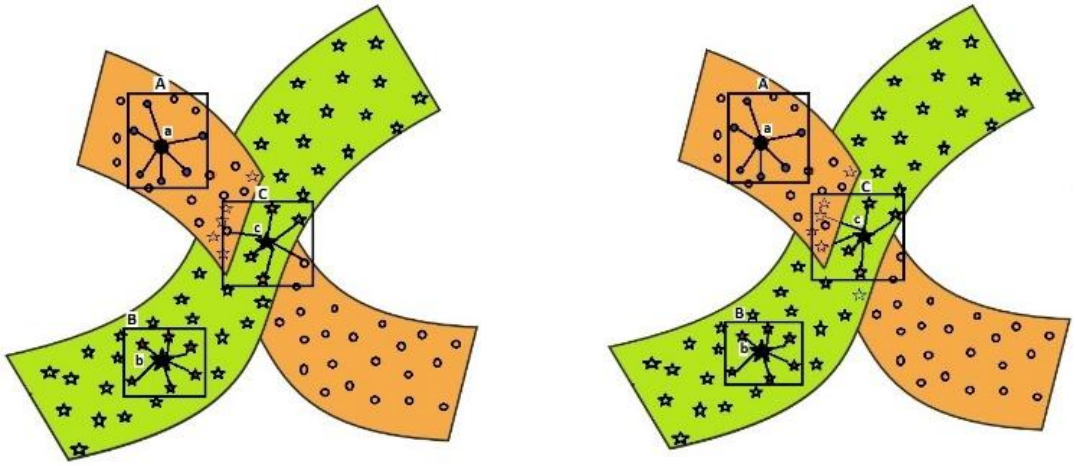
$$y = w_1\varphi_1 + w_2\varphi_2 + \cdots + w_n\varphi_n; \qquad w_1 + w_2 + \cdots + w_n = 1 \tag{5}$$

According to the definition, if the vector is constructed with combination of the affine samples, can be mapped on the space with the lower dimension. As we know, nevertheless the vectors are in the space with high dimension manifold, but they are locally on the low dimension manifold. Representing a vector in places where the manifolds are interferences is very challenging. In these cases, for representing the vectors, if the samples selection are done based on only the nearest neighbors and the sparsity term, maybe some selected samples are from the irrelevant manifold, however if the selected samples have the affinity constraint in addition, because one can consider the samples on the manifold with locally low dimension, only the samples on the relevant manifold could be selected.

For better perception of the proposed method, see Figure 1.

Two overlapped manifolds are shown in this Figure. Figure 1.a indicates a representation of the samples a,b and c regarding only the sparsity term and Figure 1.b indicates the representation of the same points regarding the manifold constraints in addition to the sparsity constraint. The samples a and b in the both Figures 1.a and 1.b are represented by the atoms from the corresponding manifolds correctly. These two samples hasn't any conflict with the other manifold.

The sample c is under different conditions. As indicated before, this sample is located on the green manifold. If you represent this sample with its adjacent atoms and only consider the sparsity term, we should consider the other manifold samples for representation same as Figure 1.a. However, if we consider $tr(XLX^T)$ and Affinity terms for its representation in addition, we will reach a better conclusion. As previously pointed out, the term $tr(XLX^T)$ emphasizes on the problem that if the samples of a manifold are closed to each other, their codes will be closed to each other as well.



a) Representation of samples a, b, c without the affinity constraint b) Representation of samples a, b, c by adding the affinity constraint

Fig. 1. The effectiveness of affinity constraint in representation of samples from the overlapped manifolds

Also, the Affinity constraint emphasizes on the problem that a collection of the closest neighbors of the concerned sample to represent every sample and then chooses a collection of weights for every sample in a way that every point is represented by the linear combination of its neighbors. The former samples are located on a manifold with high dimensions and the objective of the Affinity term is to reduce its dimensions. It is to be considered that despite the fact that the samples are on manifolds with many dimensions but they are locally located on manifolds with low dimensions. The characteristic of this new term causes the sample c to be represented with utilization of the concerned manifold data (Figure 1.b).

According to the above mentioned descriptions, we can add an affinity term to (1):

$$\min_{\Phi,X}\|Y-\Phi X\|_F^2+\gamma Tr(XLX^T)+\alpha\sum_{i=1}^{n}|x_i| \quad st.\sum_{i=1}^{n}x_i=1 \tag{6}$$

The constraint term $\sum_{i=1}^{n} x_i = 1$ is added to the main criterion as a lagrangian coefficient leading to:

$$\min_{\Phi,X}\|Y-\Phi X\|_F^2+\gamma Tr(XLX^T)+\alpha\sum_{i=1}^{n}|x_i|+\beta(1-\sum_{i=1}^{n}x_i)^2 \tag{7}$$

where β is a parameter for tuning the affinity constraint. For tuning α, β and γ parameters some experiments have been done that can be seen in the next section. In Figure 2, one can see the steps of the proposed method. After preprocessing the input data the samples are clustered using k-means algorithm to make the initial dictionary. Then the KSVD is applied for optimizing the dictionary. Finally the proposed sparse coding method is applied to extract the optimal coefficients and then classifying the test data based on the minimum error.

4.2 Solution of the proposed method

We apply the feature-sign search algorithm [3] to solve the optimization problem (7). For solving non-differentiable problems in non-smooth optimization methods, a necessary condition for a parameter vector to be a local minimum is that the zero-vector is an element of the sub-differential set containing all sub-gradients in the parameter vector [14].

Following [6,14], the optimization of the proposed sparse coding has been divided into two steps: 1) $\ell 1$-regularized least squares problem; the affine graph regularized sparse codes X are learned with dictionary Φ fixed and 2) $\ell 2$-constrained least squares problem; the dictionary Φ has been learned with affine graph regularized sparse codes X fixed. The above two steps are repeated respectively until a stop criterion is indulged.

The optimization problem in the first step can be solved by optimizing over each x_i individually.

Since (7) with *l1*-regularization is non-differentiable when x_i contains values of 0, for solving this problem, the standard unconstrained optimization methods cannot be applied.

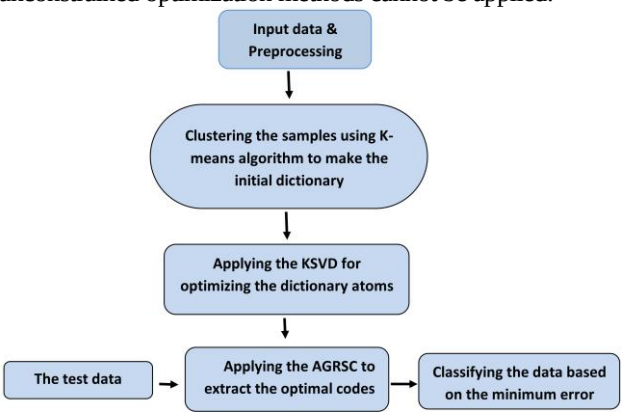


Fig. 2. The diagram of Proposed method

Several approaches have been proposed to solve the problem of this form [15,16]. In the following, we introduce an optimization method based upon coordinate descent to solve this problem [17]. It can easily be seen that (7) is convex, thus the global minimum can be achieved.

We update each vector individually by holding all the other vectors constant. In order to solve the problem by optimizing over each x_i, we should rewrite the (7) in a vector form. The reconstruction error $\|Y-\Phi X\|_F^2$ can be rewritten as:

$$\sum_{i=1}^{m}\|Y-\Phi X\|_F^2 \tag{8}$$

The Laplacian regularizer $tr(XLX^T)$ can be rewritten as:

$$Tr(XLX^T)=Tr\left(\sum_{i,j=1}^{n}L_{i,j}x_i x_j^T\right)=\sum_{i,j=1}^{n}L_{ij}x_i x_j^T=\sum_{i,j=1}^{n}L_{ij}x_i^T x_j. \tag{9}$$

□

Combining (7) , (8) , (9) the problem can be written as:

$$min\sum_{i=1}^{n}\|y_i-\Phi x_i\|_F^2+\gamma\sum_{i,j=1}^{n}L_{ij}x_i^T x_j+\alpha\sum_{i=1}^{n}|x_i|+\beta(1-\sum_{i=1}^{n}x_i)^2, \tag{10}$$

When updating x_i, the other vectors $\{x_j\}_{i\neq j}$ are fixed. Thus, we get the following optimization problem:

$$\min_{x_i}G(x_i)=\|y_i-\Phi x_i\|^2+\gamma L_{ii}x_i^T x_i+x_i^T H_i+\alpha\sum_{j=1}^{k}\left|x_i^{(j)}\right|+\beta(1-\sum_{i=1}^{n}x_i)^2 \tag{11}$$

Where $H_i=2\gamma\left(\sum_{j\neq i}L_{ij}s_j\right)$ and $x_i^{(j)}$ is the j'th coefficient of x_i.

Following the feature-sign search algorithm proposed in [18], the (11) can be solved as follows. In order to solve the non-differentiable problem, we adopt a sub-gradient strategy, which uses sub-gradients of $G(x_i)$ at non-differentiable points. Primarily we define:

$$p(x_i) = \|y_i - \Phi x_i\|^2 + \gamma L_{ii} x_i^T x_j + x_i^T H_i + \beta(1 - \sum_{i=1}^{n} x_i)^2 \quad (12)$$

Then,

$$G(x_i) = p(x_i) + \alpha \sum_{j=1}^{k} |x_i^{(j)}| \quad (13)$$

Recall that a necessary condition for a parameter vector to be a local minimum in nonsmooth optimizations is that the zero-vector is an element of the subdifferential, the set containing all subgradients at this parameter vector [14]. We define $\nabla_i^{(j)}|x_i|$ as the subdifferentiable value of the jth coefficient of x_i. If $|x_i^{(j)}| > 0$ then the absolute value function $|x_i^{(j)}|$ is differentiable, therefore $\nabla_i^{(j)}|x_i|$ is given by sign($x_i^{(j)}$). If $x_i^{(j)} = 0$ then the subdifferentiable value $\nabla_i^{(j)}|x_i|$ is set [-1,1]. So, the optimality conditions for achieving the optimal value of $G(x_i)$ is:

$$\begin{cases} \nabla_i^{(j)} p(x_i) + \alpha sign(x_i^{(j)}) & if \quad |x_i^{(j)}| > 0 \\ \left|\nabla_i^{(j)} p(x_i)\right| \le \alpha & if \quad x_i^{(j)} = 0 \end{cases} \quad (14)$$

Then, we consider how to select the optimal sub-gradient $\nabla_i^{(j)} G(x_i)$ when the optimality conditions are violated, i.e., in the case that $\left|\nabla_i^{(j)} p(x_i)\right| > \alpha$ if $x_i^{(j)} = 0$. When $x_i^{(j)} = 0$ we consider the first term in the previous expression $\nabla_i^{(j)} p(x_i)$. Suppose that $\nabla_i^{(j)} p(x_i) > \alpha$, this means that $\nabla_i^{(j)} G(x_i) > 0$ regardless the sign of $x_i^{(j)}$. In this case, in order to decrease $G(x_i)$, we will want to decrease $x_i^{(j)}$. Since $x_i^{(j)}$ starts at zero, the very first infinitesimal adjustment to $x_i^{(j)}$ will make it negative. Therefore, we can let $sign\left(x_i^{(j)}\right) = -1$. Similarly if $\nabla_i^{(j)} p(x_i) < -\alpha$ then we let $sign\left(x_i^{(j)}\right) = 1$. To update x_i suppose that we have known the signs of the $x_i^{(j)}{}'s$ at the optimal value, then we can remove the $l1$-norm on $x_i^{(j)}$ by replacing each term $|x_i^{(j)}|$ with either $x_i^{(j)}$ (if $x_i^{(j)} > 0$) or $-x_i^{(j)}$(if $x_i^{(j)} < 0$) or 0 (if $x_i^{(j)} = 0$). Thus, (13) is converted to a standard unconstrained quadratic optimization problem (QP). In this case, the problem can be solved by a linear system. The algorithmic procedure of learning affine graph regularized sparse codes is described in the following:

- for each x_i, search for signs of sign $\left(x_i^{(j)}\right)$ i = 1, … , k
- solve the reduced QP problem to get the optimal x_i^* which minimizes the objective function
- return the optimal coefficients matrix $X^* = [x_1^*, x_2^*, \dots, x_n^*]$

In the algorithm, we maintain an active set $A = \left\{j | x_i^{(j)} = 0, \left|\nabla_i^{(j)} p(x_i)\right| > \alpha\right\}$ for potentially nonzero coefficients and their corresponding signs $\theta = [\theta_1, \dots, \theta_k]$ while updating each x_i. Then, it systematically searches for the optimal active set and coefficient signs that minimize the objective function (9). In each activating step, the algorithm uses the zero-value whose violation of the optimality condition $\nabla_i^{(j)} p(x_i) > \alpha$ is the largest.

The detailed algorithmic procedure of learning affine graph regularized sparse codes is stated in Algorithm 1.

Algorithm 1: Learning Affine Graph Regularized Sparse codes

Input: Data set of n data points $Y = [y_1, \dots, y_n]$, the dictionary Φ, the graph laplacian matrix L, the parameters α, β, γ.

1- For all i such that $1 \le i \le n$ do

2- **Initializing**: $x_i = \vec{0}, \theta = \vec{0}$, and active set $A = \emptyset$,where $\theta_j \epsilon \{-1,0,1\}$denotes sign($x_i^{(j)}$).

3- **Activating:** from zero coefficient of x_i, select j=arg max$_j \left|\nabla_i^{(j)} p(x_i)\right|$. Activate $x_i^{(j)}$ (add j to the active set) only if it locally improves the objective function:

$if \quad \nabla_i^{(j)} p(x_i) > \alpha$, then set $\theta_j = -1$, $A = \{j\} \cup A$

$if \quad \nabla_i^{(j)} p(x_i) < -\alpha$, then set $\theta_j = 1$, $A = \{j\} \cup A$

4- Feature sign: let us separate Φ as some submatrix that contains only columns corresponding to the active set as $\hat{\Phi}$. Let $\hat{x}_i$ and $\hat{p}_i$ be subvectors of x_i and p.

The resulting unconstrained QP is as follows:

$$\min u(\hat{x}_i) = \left\|y_i - \hat{\Phi}\hat{x}_i\right\|^2 + \gamma L_{ii} \hat{x}_i^T \hat{x}_i + \hat{x}_i^T \hat{H}_i + \beta(1 - \sum_{i=1}^{n} \hat{x}_i)^2 + \alpha \hat{\theta}^T \hat{x}_i^T$$

Let $(\partial u(\hat{x}_i)/\partial \hat{x}_i) = 0$, the optimal value of x_i under the current active set is obtained as follows:

$$-2\hat{\Phi}^T(y_i - \hat{\Phi}\hat{x}_i) + 2\boldsymbol{\gamma} \boldsymbol{L_{ii}} \hat{x}_i + 2\boldsymbol{\gamma}\left(\sum_{\boldsymbol{j \neq i}} \boldsymbol{L_{ij}} \hat{x}_i\right) + 2\boldsymbol{\beta}(\mathbf{1} - \mathbf{1^T}\hat{\boldsymbol{x}}_i)\mathbf{1} + \boldsymbol{\alpha}\hat{\boldsymbol{\theta}} = \mathbf{0}$$

$$\hat{x}_i^{new} = \left(\hat{\Phi}^T\hat{\Phi} + \boldsymbol{\gamma}\, \boldsymbol{L_{ii}} \boldsymbol{I} + \boldsymbol{\beta}\mathbf{1}\mathbf{1}^{\boldsymbol{T}}\right)^{-1}(\hat{\Phi}^T \boldsymbol{y_i} + \boldsymbol{\beta}\mathbf{1} - \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\alpha}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{H}}_{\boldsymbol{i}}))$$

Where I is the identity matrix.

In the next step a discrete line search is performed on the line segment from $\hat{x}_i$ to $\hat{x}_i^{new}$ and checks the objective value at $\hat{x}_i^{new}$ and all points where the sign of any coefficient changes. Then the point with lowest objective value is replaced with $\hat{x}_i$. At last the zero coefficients of $\hat{x}_i$ are removed from the active set and update $\theta = sign(x_i)$.

5- **The optimality conditions**:

Condition (1): nonzero coefficients have the optimality condition as:

$$\nabla_i^{(j)} p(x_i) + \alpha sign\left(x_i^{(j)}\right) = 0, \forall \; x_i^{(j)} \neq 0$$

If condition(1) is not established go back to step 4
Else
Check condition(2).
Condition(2): zero coefficients have the optimality condition as:
$\left|\nabla_i^{(j)} p(x_i)\right| \leq \alpha, \forall x_i^{(j)} = 0$
If condition (2) is not established go back to step 3
Else
Return x_i as the solution.
6- End

4.3 Learning Dictionary

The learning dictionary Φ with the sparse codes X fixed is transformed to the following least square problem with quadratic constraints:

$$\min_{\Phi} \|Y - \Phi X\|_F^2, \quad s.t. \quad \|\varphi_i\|^2 \leq c \ , \forall i = 1, \dots, k \qquad (16)$$

Many methods have been proposed for solving this problem.

In this paper, we use the Lagrange dual method, which has been shown more efficient than gradient descent. The solution for this problem has been well described by prior works [6] and in this paper we do not consider it.

5. Experiments

In this section, for evaluating the proposed approach, some experiments for image classification has been performed.

5.1 Data Preparation

ORL, Yale face database are two well-known datasets widely used in computer vision and pattern recognition researches. The experiments has been done on these two datasets. In continuation we have introduced these two datasets.

- **ORL face dataset**

The ORL (Olivetti Research Laboratory) dataset contains 400 images consisting of 10 different images from 40 distinct persons [19]. The images of each person, were taken under different conditions such as times, lighting, facial expressions such as open / closed eyes, smiling / not smiling and facial details such as glasses / no glasses. The background of the whole images was homogeneous and dark. The size of each image was 92x112 pixels (Figure 3).

- **Yale face dataset**

The Yale Face Database [20] contains 165 images consisting of 11 images from 15 different persons under different conditions. The size of each image is 243×320. The conditions are consists different facial expression or configuration, center-light, with glasses, without glasses, left-light, right-light, normal, happy, sad, sleepy, surprised, and wink (Figure 4).

Fig. 3. some examples of the ORL dataset images

Fig. 4. some examples of the Yale face dataset images

5.2 Experimental Setup

For evaluation of the proposed approach, the results of this method on two defined datasets are compared with two state-of-the-art basic approaches, Sparse Coding (SC) [2] and Graph Regularized SC (GraphSC) [6] for image classification.

Each of the three methods can learn sparse representations for input data points. In particular, SC is a special case of the proposed method with $\beta = \gamma = 0$ and GraphSC is a special case with $\beta = 0$.

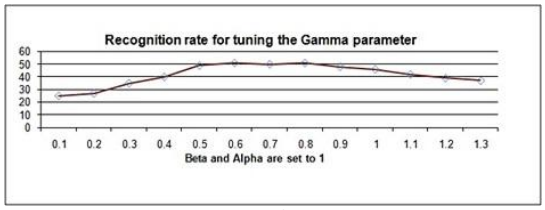
Following [6,21] the SC, GraphSC, and proposed methods are performed on data as an unsupervised dimensionality reduction procedure. For reducing a dimension of data, before applying the above algorithm, PCA is applied by keeping 98% information in the largest Eigen vectors.

Under our experimental setup, we have tuned the optimal parameters for the target classifier using leave one subject out cross validation method. Therefore, we evaluate the three baseline methods on datasets by empirically searching the parameter space for the optimal parameter settings, and report the best results of each method. For Sc and Graphsc the parameters have been set according to [11].

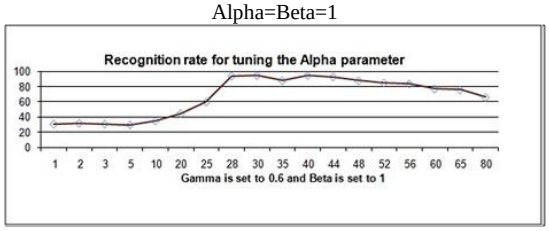
For the proposed method, we set the trade-off parameters α, β, γ through searching. In Figure 5a the plots are show the parameter value changes for ORL face dataset.

As can be seen from Figure 5, the parameters α, β, γ are set to 30, 0.1 and 0.6 respectively.

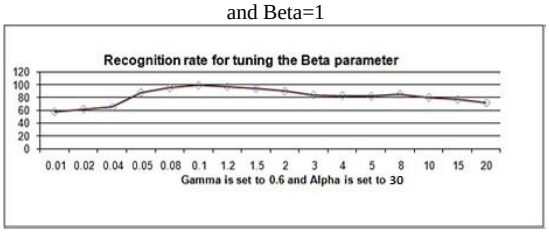
At first the value for γ parameter is achieved, for the best recognition rate assuming $\alpha, \beta = 1$. As can be seen from Figure 6a the highest recognition rate is achieved for $\gamma = 0.6$. At the next step, the value for α is achieved assuming $\gamma = 0.6$ and $\beta = 1$ for the best recognition rate. As can be seen from Figure 6b the best value for this parameter can be a number between 28 and 45. We set $\alpha = 30$ and using the same experiments the best value for β is achieved 0.1.



a) Recognition rate variations for gamma changes by setting Alpha=Beta=1



b) Recognition rate variations for alpha changes by setting Gamma=0.6 and Beta=1



c) Recognition rate variations for Beta changes by setting Gamma=0.6 and Alpha=30

Fig. 5. The parameters setting using ORL dataset

It should be noted that, the affinity constraint can be more successful when the sparsity is large enough because have the coefficients not enough sparsity, the coefficients may be selected from the hyperplane with higher dimensions than data's original dimension. In this case if the affinity constraint is added to the objective function, it may even worsen the performance with respect to the GraphSC method.

5.3 Experimental Results

For evaluating the proposed method, two experiments have been carried out. The first experiment is done on ORL face dataset for face recognition and the second one is done on Yale face dataset for face Expression recognition.

The classification accuracy of the proposed method and the two baseline methods on the two face image datasets ORL and Yale face dataset is illustrated in Figure 6 and Figure 7 respectively. As mentioned before the ORL dataset contain 40 classes of faces. Due to the lack of space in the Table only 10 classes are depicted. Among the whole dataset, classes 4 and 6, classes 8 and 10, classes 14 and 17, classes 5 and 18 are very similar to each other. Therefore, we use these classes in addition to classes 1 and 2 in the confusion matrix to show the superiority of the proposed method in classifying face datasets in Figure 6. Also in Figure 7 the results for Yale face dataset are shown. From the results we observe that the proposed method achieves much better performance than the two baseline methods.

The average classification accuracies of the proposed method on the two datasets are 91% and 63.46%, respectively for the only classes shown in the Figures. We have noticed that our proposed approach outperforms the first two baseline methods. Incorporating the graph Laplacian term of coefficients in addition to affinity constraint, leads to improve the sparse representations with more discriminating power to alleviate the classification problems.

The mean recognition rate for the selected classes are shown in Table 1. The performance improvements are 18%, 15.98% and 3.1%, 4.5% compared to baseline methods SC and GraphSC, respectively.

	Class1	Class2	Class4	Class5	Class6	Class8	Class10	Class14	Class17	Class18
Class1	92	0	2	1	0	0	2	1	2	0
Class2	0	98	0	0	0	2	0	0	0	0
Class4	1	0	65	1	22	5	3	2	0	1
Class5	2	0	2	61	0	2	0	0	13	20
Class6	0	1	28	1	68	0	0	1	0	1
Class8	0	1	0	0	0	69	14	2	8	6
Class10	1	0	0	2	0	19	69	1	7	1
Class14	2	0	3	3	2	0	3	71	15	1
Class17	1	0	0	7	3	2	7	8	69	3
Class18	1	0	0	21	4	2	1	1	2	68

a) SC recognition rate for the selected dataset from ORL

	Class1	Class2	Class4	Class5	Class6	Class8	Class10	Class14	Class17	Class18
Class1	98	0	1	0	0	0	0	0	1	0
Class2	0	98	0	0	0	1	0	0	0	1
Class4	0	0	89	1	5	1	1	1	1	1
Class5	1	0	2	91	0	0	0	0	2	4
Class6	0	0	3	0	91	0	2	2	1	1
Class8	0	2	0	0	1	87	8	0	2	0
Class10	0	0	1	0	0	8	79	4	5	3
Class14	0	0	1	0	0	0	0	83	13	3
Class17	0	0	0	1	2	1	8	11	75	2
Class18	1	0	0	6	1	0	0	0	4	88

b) GraphSC recognition rate for the selected dataset from ORL

	Class1	Class2	Class4	Class5	Class6	Class8	Class10	Class14	Class17	Class18
Class1	99	0	1	0	0	0	0	0	0	0
Class2	0	100	0	0	0	0	0	0	0	0
Class4	1	0	95	1	2	0	0	0	1	0
Class5	0	0	2	96	0	0	0	1	1	0
Class6	0	0	2	1	95	0	1	0	0	1
Class8	0	0	0	0	1	92	4	1	2	0
Class10	0	0	0	0	0	7	84	3	5	1
Class14	0	0	0	0	0	0	2	79	14	5
Class17	0	0	0	1	2	1	5	11	78	2
Class18	0	0	0	1	1	0	2	4	0	92

c) The recognition rate for the selected dataset from ORL for the proposed method

Fig. 6. The image confusion matrix for ORL data between three methods Sc, GraphSc and the proposed method

	center-light	With glasses	happy	left-light	No glasses	normal	right-light	sad	sleepy	surprised	Wink
center-light	23.6	7.8	4.2	6	21.2	21.8	4.2	4.2	0	0.6	6
w/glasses	6	76.6	3	4.2	0	4.2	3	0	1.8	1.2	0
happy	7.8	4.2	65.2	6	7.8	4.2	0	1.2	1.8	0	1.8
left-light	1.8	3	4.2	45.1	15.7	21.2	1.8	1.2	0.6	3	2.4
w/no glasses	23.6	3	3	7.2	25.1	23.6	3	4.8	4.2	1.8	0.6
normal	23.3	0	4.2	10.9	21.8	23.6	4.8	1.2	3	3	4.2
right-light	6	1.8	6	4.8	3	1.2	57.2	6	3	6.8	4.2
sad	0	0	1.8	6	0.6	0	7.2	35.6	30.8	12	6
sleepy	5.4	0	1.8	3	1.8	1.2	6.8	30	39.8	6	4.2
surprised	1.8	1.2	4.2	4.2	0.6	0	6	12	6	60.4	3.6
wink	0	1.8	3	1.8	1.2	1.2	6	9	4.8	1.2	70

a) SC recognition rate for Yale face dataset

	center-light	With glasses	happy	left-light	No glasses	normal	right-light	sad	sleepy	surprised	wink
center-light	38	7.8	4.2	6	18	18.8	1.8	3	0	0	2.4
With glasses	6	82.6	1.8	3	1.2	1.8	1.2	0	1.8	0.6	0
happy	7.8	3	74.8	3	4.2	1.2	0	1.2	0	0	4.8
left-light	1.8	1.2	3	52.5	12	17.5	1.8	2.4	0.6	3	4.2
No glasses	20	1.8	1.8	7.4	46.1	13.3	1.8	4.2	1.8	1.2	0.6
normal	18	0	1.2	6	12.6	43	6	1.2	6	3	3
right-light	6	1.2	6	4.8	1.8	1.2	68.8	4.2	1.2	3	1.8
sad	0	1.2	1.8	6	1.2	1.8	6	54.2	20	6	1.8
sleepy	6	0	0	3	1.2	1.2	4.8	29	53.6	1.2	0
surprised	2.4	0.6	3	4.8	0	1.8	4.2	12	6.8	64.4	0
wink	0	1.2	1.8	3.6	0.6	1.2	3.6	4.8	2.4	10.3	70.5

b) GraphSC recognition rate for Yale face dataset

	center-light	w/glasses	happy	left-light	w/no glasses	normal	right-light	sad	sleepy	surprised	wink
center-light	47.4	6.8	4.2	6	15.7	16.9	1.8	3	0	0	1.2
w/glasses	6	83.4	1.8	1.8	1.2	0	0.6	1.2	1.8	1.2	0
Happy	6.8	1.8	78.2	3	4.2	1.2	0	1.2	0.6	0	3
left-light	1.8	1.2	2.4	60.3	12	13.9	1.8	1.2	0.6	1.8	3
w/no glasses	13.9	1.8	1.8	7.2	42.9	12	7.2	5.4	3	3	1.8
Normal	15	0	0.6	6	9.6	49.6	6	1.2	6	1.8	4.2
right-light	4.2	1.2	4.2	4.8	2.4	0.6	72.4	4.2	1.8	3	1.2
Sad	0	1.2	1.8	6	3	1.8	4.2	57.7	13.9	5.4	3
sleepy	4.8	1.2	0	3	1.8	1.2	1.2	17.5	68.1	1.2	0
surprised	1.8	1.2	3	4.8	2.4	1.2	1.8	6	8.4	68.8	0.6
Wink	0	1.2	1.2	3	3	0	0.6	0	4.8	11.5	74.3

c) Recognition rate for Yale face dataset for the proposed method

Fig. 7. The image confusion matrix for Yale face dataset between three methods Sc, GraphSc and the proposed method

Table 1. Mean Recognition rate for SC, GraphSc and the proposed methods

	Mean Recognition rate for ORL data	Mean Recognition rate for Yale data
SC method	**73%**	**47.48%**
GraphSc method	**87.9%**	**58.96%**
DLPV method [8]	**97.6%**	**-----**
Proposed method	**98.8%**	**63.46%**

If we add the ORL absent classes, the recognition rate is raised up to 98.8%. For better evaluating the proposed method, the recognition rate for all classes of ORL dataset is compared with a recent non-sparse based method in face recognition application at 2015 [8]. The authors in this paper have proposed a face recognition method based on the discriminative locality preserving vectors (DLPV). The best result in this paper for the ORL dataset is reported 97.6%. With comparing these two methods the superiority of the proposed algorithm becomes clear.

5.4 Experiment on action dataset

For more evaluation the proposed sparse coding method, this method is applied on KTH action dataset. The KTH dataset [22], contains six actions including walking, jogging, running, boxing, hand waving and hand clapping, performed several times by 25 subjects in four different scenarios. Overall, it contains 2391 sequences. Some samples of KTH dataset are shown in Figure 8.

Fig. 8. Some samples of KTH dataset

At first the dataset has been divided into two 50% portions as train and test data randomly for each class. Then the HOG3D descriptor is extracted from data. For reducing the dimension, PCA is applied with keeping 98% of information. Then the proposed sparse coding method is applied for feature extraction from the descriptor. We use SVM method for classifying the data.

The results show that the proposed method could classify the KTH data with about 94% precision. This result show that the proposed method can be applied in action recognition applications same as to the face recognition.

6. Conclusion and future work

In this paper, a novel approach for robust face recognition namely affine graph regularized sparse coding has been proposed. In the proposed method, the well-defined graph regularized sparse coding method has been improved by adding the affinity constraint. Using this term, until the sparsity is big enough the manifold structure of features is better preserved. The results indicate that the proposed method with comparison to some other approaches has the better performance for face recognition. In addition the results show that the proposed method could be applied for human action recognition datasets as well. In future works we would apply the proposed method for action recognition artificial and real world datasets for evaluating the proposed method.

References

[1] M. Long, G. Ding, J. Wang, J. Sun, Y. Guo, and P. S. Yu, "Transfer Sparse Coding for Robust Image Representation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[2] J.Zhang, D. Zhao, W. Gao, "Group-based Sparse Representation for Image Restoration", IEEE Transactions on Image Processing, vol. 23, no. 8, pp. 3336 – 3351, 2014.

[3] H. Lee, A. Battle, R. Raina, and A. Y. Ng., "Efficient sparse coding algorithms", In Advances in Neural Information Processing Systems, 2006, pp. 801-808.

[4] Y. Cheng, Z. Jin, T. Gao, H. Chen and N. Kasabov, "An Improved Collaborative Representation based Classification with Regularized Least Square (CRC-RLS) Method for Robust Face Recognition", Neuro computing, vol. 215, no. c, pp. 250-259, 2016.

[5] Y. Censor and S. Zenios, "Parallel Optimization: Theory, Algorithms, and Applications," 1st ed., New York: Oxford Univ. Press, 1997.

[6] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation", IEEE Transactions on Image Processing, Vol. 20, No.5, pp. 1327-1336, 2011.

[7] Y. N. Liu, F. Wu, Z. H. Zhang, Y. T. Zhuang, and S. C. Yan, "Sparse representation using nonnegative curds and whey", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2010.

[8] Y. Wen, L. Zhang, K. M. von Deneen, L. He, "Face recognition using discriminative locality preserving vectors", Digital Signal Processing, Vol. 50, pp. 103-113, March 2016.

[9] M. Yang, L. Zhang, J. Yang, and D. Zhang. "Robust sparse coding for face recognition", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011.

[10] Z. Lu, Y. Peng, "Latent semantic learning with structured sparse representation for human action recognition", Pattern Recognition, vol. 46, no. 7, pp. 1799-1809, July 2013.

[11] M.Zheng, J.Bu, C.Chen, C.Wang, L.Zhang, G.Qiu and D.Cai, "Graph Regularized Sparse Coding for Image Representation", Journal of Latex Class Files, vol. 6, no. 1, Jan 2007.

[12] B. Quanz, J. Huan, and M. Mishra, "Knowledge transfer with low-quality data: A feature extraction issue", IEEE Transactions on Knowledge and Data Engineering, vol. 24, no.10, pp. 1789-1802, October 2012.

[13] Hazewinkel, Michiel, ed. (2001), "Affine transformation, Encyclopedia of Mathematics" , Springer, ISBN 978-1-55608-010-4.

[14] R. Fletcher, "Practical methods of optimization", 2nd ed., Wiley-Interscience, New York, 1987.

[15] M. Aharon, M. Elad, A. Bruckstein, and Y. Katz, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation", IEEE Transactions on Signal Processing, vol.57, no.11, pp. 4311-4322, October 2006.

[16] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering", In Advances in Neural Information Processing Systems 14, NIPS, 2001, pp. 585-591.

[17] X.Lu, Y. Yuan and P. Yan, "Alternatively Constrained Dictionary Learning for Image Super resolution", IEEE Transactions On Cybernetics, vol. 44, no. 3, March 2014.

[18] E. Cand`es and T. Tao, "Near-optimal signal recovery from random projections: niversal encoding strategies?", IEEE transactions on information theory, vol. 52, no. 12, pp. 5406–5425, 2006.

[19] F. Samaria, A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification", Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, 1994.

[20] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", IEEE Transaction on PAMI, vol. 19, no. 7, pp. 711-720, July 1997.

[21] S. J. Pan, I.W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis", IEEE Transactions on Neural Networks, vol.22, no.2, pp.199–210, Feb 2011.

[22] X. Wu, D. Xu, L. Duan, and J. Luo, "Action recognition using context and appearance distribution features", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11), 2011, pp. 489–496.

Mohsen Nikpour received the B.Sc And M.Sc degrees in electrical engineering from Mazandaran University, Iran, in 2004 and 2008 respectively and Know he is a Ph.D Student in electrical engineering in Babol noushirvani University, Islamic Republic of Iran. His current research interests include Image processing, video processing, video coding, image segmentation and image denoising.

Mohammad Reza Karami-Mollaei received the B.Sc in Electrical and Electronic Engineering in 1992, M.Sc of signal processing in 1994, and Ph.D in 1998 in Biomedical Engineering from I.N.P.L d'Nancy of France. He is now an associate professor with the Department of Electrical and Computer Engineering, Babol University of Technology. Since 1998 his research is in signal and speech processing. He published more than 100 articles in journals and conferences. His research interests include Speech, Image and signal processing.

Reza Ghaderi received B.Sc in 1989 from Ferdoosi University of Mashhad, IRAN, M.Sc in 1991 from Tarbiat Modaress University, IRAN and Ph.D in 2001 from Surrey University UK all in Electronic Engineering. Currently he is an associate prof. at Nuclear Eng. Dept. of Shahid Beheshti Univ., Tehran, Iran. His research interests are neural networks, pattern recognition, system modeling, signal processing, Fuzzy logic, artificial intelligent.