Int. J. Nonlinear Anal. Appl. 12 (2021) No. 2, 1141-1152
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.5187

Face recognition via weighted non-negative sparse representation

Hoda Khosravi[a], Aboozar Ghaffari[b,*], Javad Vahidi[c]

[a]*Department of Software Engineering, Sari Branch, Islamic Azad University, Sari, Iran*
[b]*Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran*
[c]*Department of Mathematics, Iran University of Science and Technology, Tehran, Iran*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Face recognition is one of the most important tools of identification in biometrics. Face recognition has attracted great attention in the last decades and numerous algorithms have been proposed. Different researches have shown that face recognition with Sparse Representation based Classification (SRC) has great classification performance. In some applications such as face recognition, it is appropriate to limit the search space of sparse solver because of local minima problem. In this paper, we apply this limitation via two methods. In the first, we apply the nonnegative constraint of sparse coefficients. As finding the sparse representation is a problem with very local minima, at first we use a simple classifier such as nearest subspace and then add the obtained information of this classifier to the sparse representation problem with some weights. Based on this view, we propose Weighted Non-negative Sparse Representation WNNSR for the face recognition problem. A quick and effective way to identify faces based on the sparse representation (SR) is smoothed L_0-norm (SL_0) approach. In this paper, we solve the WNNSR problem based on the SL_0 idea. This approach is called Weighted Non-Negative Smoothed L_0 norm ($WNNSL_0$). The simulation results on the Extended Yale B database demonstrate that the proposed method has high accuracy in face recognition better than the ultramodern sparse solvers approach.

Keywords: face recognition, face subspace, sparse decomposition, smoothed L_0-norm, weighted L_0-norm, non-negative smoothed L_0 norm.

*Corresponding Author
Email addresses: Broumandnia@gmail.com (Aboozar Ghaffari), j.vahidi@iust.ac.ir (Javad Vahidi)

Received: February 2021 *Accepted:* May 2021

1. Introduction

Given the increasing need for the creation and development of automated systems, the problem of detecting and identifying the faces of people in the images has been considered by the researchers. Although face recognition methods have experienced significant advances in various applications and research areas, they still face major challenges such as lightness, gesture, expression and occlusion [1, 2, 3, 4, 5, 6, 7, 8]. In recent years, the sparse representation based classification (SRC) has been of great interest to researchers [9, 10, 11, 12, 13]. In the first step, all training images are used to calculate the SR coefficients of the test image. Then, the reconstruction errors between the test images and representation of each class are obtained. Finally, the test image is assigned to the class label that has the least amount of residual in the reconstruction errors. The results of the test show that SRC will provide excellent performance especially with respect to occlusion and resistance against noisy environments. In this method, the L_1-norm optimization, Linear Programming (LP) or Basis Pursuit (BP) [14, 15] are used to find the sparse solution. To minimize the computational complexity and execution time of the program, the Smoothed L_0-norm optimization (SL_0) [16], which has a higher convergence rate than BP, is used instead of the BP algorithm. If the uniqueness conditions are correctly met, SL_0 algorithm is superior to BP algorithm in both terms of performance and speed [16].

Sparse recovery is an optimization problem with local minima. In the BP approach the sparse recovery problem has been replaced with a convex problem. SL_0 tries to minimize the zero-norm directly with the smoothed version of zero-norm. In some applications such as classification, it is need to limit the search space of sparse solutions. Here, to avoid the local minima problem, we propose two approaches for increasing the face recognition accuracy. The first approach is adding the non-negative condition to sparse recovery problem. In [9], researches showed that the condition improved the accuracy of facial expression recognition. The second approach to limit the search space is fusion of SRC and a simple classifier such as Nearest Subspace NS [17]. This fusion is accomplished via weighted sparse representation that its weights are determined via NS classifier. Based on these approaches we propose weighted non-negative sparse representation WNNSR problem for face recognition. To obtain sparse solution of WNNSR, we use the smoothed L_0 norm idea. In fact we propose a new version of SL_0 called Weighted Non-Negative Smoothed L_0 norm ($WNNSL_0$). This decomposition is used in the face recognition problem. In this application, the WNNSR is a fusion operator of NS and SRC classifiers considering the non-negative constrain.

The remainder of the paper is organized as follows. The sparse solvers are surveyed in Section 2. The Proposed Approach based on weighted sparse representation with non- negative constraint is discussed In Section 3, the experimental results for simulated in extended Yale b are reported in Section 4. Finally, the paper is summarized in Section 5.

2. Sparse Representation Based Classifier

The fundamental issue in identifying an object is to determine the test data class using labeled training samples of k distinct classes. n_i Training samples of i-th class are placed in columns of the matrix $A_i = [V_{i,1}, V_{i,2}, \cdots, V_{i,n}] \times R^{m \times n_i}$. In the face recognition, a black and white image with a dimension of $w \times h$ is represented through the vector R by sequentially placing the image columns. The columns of Ai denote the training facial images related to i-th person.

A variety of models and representations have been presented for A_i. According to one of the simplest and most effective models, every image or feature extracted from a class belongs to a linear subspace. Based on another model called as "face subspace", the same facial images under lightness changes cover the approximate subspace with a low dimension in space R^m [18, 19], i.e. each image class occupies a small part of the m-dimensional space. For simplicity, it is assumed that the training samples of each class form a subspace. Hence, each test vector sample $X \in R^m$ of i-th class can be approximated as a linear combination of the training sample vectors of i-th class with scalar values $a_{i,j} \in R$, $j = 1, 2, \cdots, n_i$:

$$X = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \cdots + a_{i,n_i}v_{i,n_i} \tag{2.1}$$

Since the test sample belonging to i-th class is not clear, we define the matrix A by putting in sequential order the matrices A_i as follows:

$$A = [A_1, A_2, \cdots, A_k] \in \mathbb{R}^{m \times n} \tag{2.2}$$

Therefore, the linear representation of the test vector sample $x \in \mathbb{R}^m$ is rewritten as follows:

$$X = As_0 \in \mathbb{R}^m \tag{2.3}$$

Where $s_0 = [0, \cdots, 0, a_{i,1}, \cdots, a_{i,n}, 0, \cdots, 0]^T$ is the coefficients vector. All entries of the coefficients vector are zero except for i-th class. Since entries of the vector s_0 specify the identity of the test sample x, the solution of the linear equation $X = As$ is important. Compared to local NS and NN methods, the methodology used here is global. It is clear that the solution of the linear equation $X = As$ depends on the ratio of the number of equations m to the number of unknowns n. If $m > n$, then the linear equation $X = As$ will be over determined and the unique solution can be found.

For face recognition, the equation $X = As$ is usually underdetermined $(m < n)$ and then the solution is not unique. To solve this problem, we can usually use the L_2-norm solution as follows:

$$(L_2) \qquad \widehat{s}_0 = argmin\|s\|_2 \quad s.t,\ X = As \tag{2.4}$$

Although the optimization of this problem can be simply solved by quasi-inverse matrix A, but because the solution is not sparse (in general), this solution does not contain specific information for the face recognition application.
Considering the nature of the problem and concept of the face subspace model, SR techniques are suggested to solve this problem. Each very sparse solution can identify the identity of the test sample X according to the stated concepts. Hence, the method for finding the sparsest solution of Equation $X = As$ by solving the following optimization problem is proposed.

$$(L_0) \qquad \widehat{s}_0 = argmin\|s\|_2 \quad s.t,\ X = As \tag{2.5}$$

$\|\cdot\|_0$ Denotes the L_0-norm and counts the number of non-zero entries. It is shown that if the matrix A is random and the equation $X = As$ has the answer that the number of its non-zero elements is less than half the number of equations $m/2$, this answer is unique, i.e. $\widehat{s}_0 = s_0$ [14].

The general theorem of uniqueness of sparse solution is expressed by the definition of the spark of the matrix A. The problem of finding the sparsest answer is a hard problem because L_0-norm is not derivable. Several approximate approaches have been presented to solve this problem. In the following, we summarize the two methods proposed to solve this problem.

2.1. Finding the sparse solution by L_1-norm minimization

Many methods of solving (2.5) using the cost function, which is a measure of response sparseness, try to find the sparsest answer by solving an optimization problem. If the selected cost function is a better criterion for being sparse the vector x, then the accuracy of the final solution is higher. On the other hand, the desirable cost function is a function that simplifies the solving of the optimization problem.

Some methods use the L_1-norm minimization metric to measure the vector sparsity, namely:

$$(L_1) \qquad \widehat{s}_0 = argmin\|s\|_1 \quad s.t,\ X = As \tag{2.6}$$

Substituting of L_0-norm with L_1-norm converts the optimization problem into a convex one, which has appropriate methods (e.g. BP). In [15], it has been proven that in majority of underdetermined linear systems, L_0-norm and L_1-norm minimization leads to one solution, which is the sparsest one. Of course, in this method, the uniqueness condition is very limited compared to that of the answer (2.5). This is one of the reasons that some researchers have focused on ideas that directly solve the L_0-norm problem.

In [10], researchers utilize the noise model (2.6) for face recognition as follows:

$$(L_1) \qquad \widehat{s}_0 = argmin\|s\|_1 \quad s.t,\ \|X - As\|_2 \leq \epsilon \tag{2.7}$$

Here we use the model $X = As_0 + z$ to represent the test sample, which $z \in \mathbb{R}^m$ includes the noise model with limited energy $\|z\|_2 \leq \epsilon$.

2.2. Finding the sparse solution by smoothed L_0-norm minimization

As shown below, the main idea of this method is to use a continuous and smooth function for the L_0-norm approximation of the vector s:

$$\|s\|_0 \approx m - \sum_i \exp\left(\frac{-s_i^2}{2\sigma^2}\right) = m - F_\sigma(s) \tag{2.8}$$

If σ tends to zero, this approximation is converted to equality. Therefore, maximizing the function $F_\sigma(s)$ for a small σ is equivalent to minimizing the L_0-norm. Therefore, the optimization problem will be as follows:

$$(SL_0) \qquad \widehat{s}_0 = argmaxF_\sigma(s) \quad s.t,\ X = As \tag{2.9}$$

To find the answer of the above problem, we can use the steepest descent maximization algorithm. For details of the optimization algorithm of problem (2.9) called as SL_0, refer to [16]. An important feature of this algorithm is the high convergence rate and its good performance compared with the BP algorithm.

2.3. Classification based on sparse representation

For the new test sample x associated with one of training set classes, firstly we obtain SR by using (2.5) or (2.6) or (2.7) or (2.9). Ideally, non-zero entries in s_0 estimation are related to i-th class columns. Therefore, we can easily assign the test sample x to that class. However, this is not the case due to model error and noise. In fact, small non-zero entries corresponding to other classes can also be found. Accordingly, many classifications can be designed for the identification area. In one of the easiest ways, the class with the largest non-zero entry is considered as the winner. Nevertheless, the method presented in [10] is a more logical and subjective one. This approach examines which linear combination of classes training samples with coefficients obtained from SR offers a better approximation of the test sample x.

For each class i, vector $\delta_i(\widehat{s}_0) \in \mathbb{R}^m$ derived from $\widehat{s}_0$ is a vector in which entries corresponding to classes other than i are set to zero. By this definition, the approximation of the test sample x using the training samples of i-th class is equal to $x_i = A\delta_i(\widehat{s}_0)$. So the classification of x is based on the best approximation, that is, the class with the least estimation error is the winner one:

$$\min_i r_i(x) = \min_i \|x - A\delta_i(\widehat{s}_0)\|_2 \tag{2.10}$$

The classification algorithm is summarized in Table 1.

Table 1: SRC algorithm's summary

1: Input: matrix of training samples $A \in \mathbb{R}^{n\times m}$ and test sample $x \in \mathbb{R}^m$.
2: normalizing the columns of matrix A.
3: obtaining SR for the test sample using (2.5) or (2.6) or (2.7) or (2.9).
4: calculating the estimation error for all classes by $r_i(x) = \|x - A\delta_i(\widehat{s}_0)\|_2$, $i = 1, \cdots, k$.
5: output: identifying x with $\min_i r_i(x)$.

3. The Proposed Approach: SRC based on Weighted Non-Negative Sparse Recovery

In the SRC classifier, finding the sparse solution has a key role. The recognition accuracy is dependent on this optimization problem. The presence of noise in the real data and the error of linear model affect this problem which has local minima. Here, to improve the SRC performance we propose two considerations of non-negative constrain and weighted sparse representation in the sparse recovery problem. In rest of this section, at first these considerations are illustrated and then the proposed approach is demonstrated.

3.1. Non-Negative Constrain

Researchers show that in the application of face recognition the non-negative constrain of sparse coefficients improves the SRC performance [9]. This optimization problem is in the following form:

$$(NN\ell_0) \qquad \widehat{s}_0 = argmin\|s\|_0 \quad s.t, \;\; x = As, \;\; \forall\, i\; s_i \geq 0 \tag{3.1}$$

Similar to the sparse recovery, this optimization problem also can be solved via some approaches such as L1-norm minimization [1] or constrain smoothed $L_0 (CSL_0)$.

3.2. Weighted Sparse Representation: Combination of SRC and a simple classifier

The goal of this method is to obtain some information about the query data via a simple classifier such as NS and add this information to the problem of sparse recovery. In traditional sparse recovery (2.5), the probability of activity of all columns of A (activity of i-th column means large $|s_i|$) is the same. This activity is dependent on the sample of the input signal and the column of the dictionary matrix A. There is not external factor controlling the activity of the columns in SR of the x signal. Here, we provide a strategy for controlling the activity probability of each column by weight. To this end, we will change the optimization problem (2.5) as follows:

$$(wL_0) \qquad \widehat{s}_0 = argmin \sum w_i s_0^i \quad s.t, \;\; x = As \tag{3.2}$$

where w_i denotes the controlling weight and is a positive number. Here, the probability of activity of all columns in A is not the same, and w_i's specify the probability of activity of the atoms. The performance of weight can be interpreted as follows: if is large in the cost function (3.2), the probability of activity of i-th column is low and if w_i is small, the probability of activity of i-th column is high. In other words, these weights limit the search space with respect to some information obtained from a simple classifier. In this way, we can control the activity of the columns in SR of the signal x by changing the w_i's. Here, the weights of all training columns of class j is same and determined based on the distance between the vector x and the class subspace of A_j. In fact we use the information of nearest subspace to determine weights. The NS classifier assigns the test sample x to j-th class, provided that distance from x to j-th class subspace created by the training samples $A_j \in \mathbb{R}^{m\times n_j}$ has the lowest value compared with all classes. This distance is defined as follows:

$$mse_j(x) = \min_{s\in\mathbb{R}^{n_j}} \|x - A_j s\|_2^2 \tag{3.3}$$

Where $mse_j(x)$ represents the degree of similarity of the test sample x to j-th class.

In summary, we extract some preliminary information about the classification of the test sample by very simple algorithms such as NS, and then use this information to optimize the sparse recovery. Problem (3.2) can be solved similar to the traditional sparse recovery [20].

3.3. SRC based on Weighted Non-Negative Sparse Recovery

In this subsection, we propose the weighted non-negative sparse recovery WNNSR for the application face recognition. In fact, here we combine two discussed considerations in the following sparse recovery problem.

$$(WNNL_0) \qquad \widehat{s}_0 = argmin \sum w_i s_i^0 \quad s.t, \quad x = As \;\; \forall\, i \;\; s_i \geq 0 \tag{3.4}$$

Where the weights w_i are determined based on the NS classifier. To solve this optimization problem, we propose two approaches based L_1 norm and smoothed L_0 norm as follows. As solving L_0 norm is NP-hard, inspired to the methods of BP, SL_0 and CSL_0 . We propose two approaches to solve $WNNL_0$ problem in the following subsections.

3.3.1. $WNNL_0$ Solver based on L_1 norm

Here, to solve the $WNNL_0$ problem we replace the L_1 norm with the L_0 norm in the following mathematical form:

$$(WNNL_1) \qquad \widehat{s}_0 = argmin \sum w_i |s_i| \quad s.t, \quad x = As \;\; \forall\, i \;\; s_i \geq 0 \tag{3.5}$$

This optimization problem is convex and can be solved via linear programming [14]. As the weight w_i is positive we have $w_i|s_i| = |w_i s_i|$, hence the $WNNL_1$ problem can be converted to the non-negative basis pursuit with changing variables of s and A.

3.3.2. $WNNL_0$ Solver based on smoothed L_0 norm

Here, the idea of smoothed L_0 norm is used to find the sparse solution of $WNNL_0$. To consider the non-negative constraint, inspired by the CSL_0 algorithm we apply a penalty for negative coefficients by adding the another weights to the sparse recovery problem as follows

$$(WNNL_0^k) \qquad \widehat{s}_0 = argmin \sum_i w_i p_k(s_i) s_i^0 \qquad s.t, \quad \mathbf{x{=}As} \quad \forall\, i \quad s_i \geq 0 \tag{3.6}$$

Where

$$p_k(s) = \left(\frac{k+1}{2} - \frac{k-1}{2} sign(s)\right) = \begin{cases} k & if\ s < 0 \\ 1 & if\ s > 0 \end{cases} \tag{3.7}$$

And k is a constant that $k \geq 1$. This weight is dependent on the sign of coefficients. This function results in the penalty of negative coefficients be higher than the positive ones. If k is infinite and $w_i > 0$ then the penalty term of negative coefficient $w_i p_k(s_i)$ is infinite. Hence the entries of sparse vector cannot be negative via the optimization of problem $WNNL_0^k$. To find the sparse solution of $WNNL_0^k$ inspired by the idea of SL_0 we use a continuous and smooth function for the L_0 norm approximation of the S_i^0 and reformulate the SL_0 for $WNNL_0^k$ as follows:

$$(WNNSL_0) \qquad \widehat{s}_0 = argmin \sum_i w_i p_k(s_i) \left(1 - \exp\left(\frac{-s_i^2}{2\sigma^2}\right)\right) \qquad s.t, \quad \mathbf{x{=}As} \quad \forall\, i \;\; s_i \geq 0 \quad (3.8)$$

The problem $WNNSL_0$ is equivalent to $WNNL_0^k$ when $\sigma \to 0$. The idea of $WNNSL_0$ is to minimize $F_\sigma^{Wk}(s) = \sum_i w_i p_k(s_i) \left(1 - \exp\left(-\frac{s_i^2}{2\sigma^2}\right)\right)$ for small σ subject to $x = As$. To escape from trapping into local minima, it uses a decreasing sequence of σ and the minimize of $F_\sigma^{Wk}(s)$ is used as a starting point to search the minimizer of $F_\sigma^{Wk}(s)$ for the next (smaller) σ. In the process of decreasing of σ, the penalty parameter k is also increased.

To minimize $F_\sigma^{Wk}(s)$ for two fixed parameters of k and σ, subject to $x = As$, it uses a steepest descent approach: each iteration composed of an unconstrained minimization step ($s \leftarrow s - \mu \bigtriangledown F_\sigma^{Wk}$), followed by projection to the feasible set of space $x = As$ with the following equation:

$$s \leftarrow s - A^T \left(AA^T\right)^{-1} (As - X) \qquad (3.9)$$

In the process of steepest descent we need the gradient of $F_\sigma^{Wk}(s)$ as follows:

$$\frac{\partial F_\sigma^{Wk}}{\partial s_i} = w_i p_k(s_i) \frac{s_i}{\sigma^2} \exp\left(\frac{-s_i}{2\sigma^2}\right) \qquad (3.10)$$

$$\bigtriangledown F_\sigma^{Wk} = \left[\frac{\partial F_\sigma^{Wk}}{\partial s_1} \cdots \frac{\partial F_\sigma^{Wk}}{\partial s_2}\right]^T \qquad (3.11)$$

Note that if the weight w_i is equal to zero, then the penalty of non-negative constraint is not considered for the coefficient s_i. In the face recognition application, this is not an important problem because $w_i = 0$ means that the distant between query data and one of class subspaces is zero. Hence the class of query data is determined. Here, to consider this problem in the general framework, we also add a projection operator on the non-negative subspace in the minimization process of function $F_\sigma^{Wk}(s)$.

Another note of the proposed approach is robustness of $WNNL_0^k$ in the presence noise. To consider this problem, we must change the constraint $x = As$ to relaxed form $\|x - As\|_2^2 \leq \epsilon$. Inspired by [21] to consider this problem in the SL_0 algorithm, we perform the projection process when the solution is outside of the feasible region $\|x - As\|_2^2 \leq \epsilon$.

The proposed algorithm is illustrated in Table 2. The following parameters are initialized in the first step: $\sigma_{\min}$ (the minimum value of σ that should be a very small number), k (the maximum penalty for non-negative coefficients), L (the number of iterations for minimizing $F_\sigma^{Wk}(s)$ in the gradient descent approach, μ (the gradient descent factor), d (the decreasing factor of σ), ϵ (the acceptable reconstruction error), $\widehat{s}$ (the initial solution based on L_2 norm minimization which is equivalent to $\sigma = \infty$), and σ (the initial value of σ that should be a large number [22]).

Table 2: The proposed sparse solver with Weighted Non-Negative Smoothed L_0 ($WNNSL_0$).

Inputs: A, x
Initialization:
(1) Choose the value $\sigma_{\min}$, K, L, μ, d, ϵ
(2) $\widehat{s} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{x}$
(3) $\sigma = 2\max|\widehat{s}|$
While $\sigma > \sigma_{\min}$
(1) **k=1+K** $\times \frac{\sigma_{\min}}{\sigma}$
(2) For $1 = 1 : L$
(a) Compute $\bigtriangledown F_\sigma^{wk}(\widehat{s})$ using (3.10) and (3.11)
(b) $\widehat{s} \leftarrow \widehat{s} - \mu/k \times \bigtriangledown F_\sigma^{wk}(\widehat{s})$
(c) If $\widehat{s}_i < \mathbf{0}$ $(i = 1, \cdots, m)$
$\widehat{\mathbf{s}}_i = \mathbf{0}$
(d) End if
(e) If $(\|\mathbf{A}\widehat{\mathbf{s}} - \mathbf{x}\|_2 > \epsilon)$
$\widehat{s} \leftarrow \widehat{s} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\widehat{s} - \mathbf{x})$
(f) End if
(3) End for
(4) $\sigma = \sigma \times \mathbf{d}$
End while
Output: $\widehat{\mathbf{s}}$

4. Simulation results

In this section, we examine the performance of SRC with features called as "down sampling" . The database used is "Extended Yale B". This database contains 2414 facial images from the front view of 38 people. All images have been taken with 192×168 dimensions in different laboratory-controlled light conditions. For each class, 54 training files and 10 test files are randomly selected. We calculate the percentage of correct identification of the algorithm for the number of different features (36, 56,132 and 504). These numbers are proportional to the down sampling ratios of 1/32, 1/24, 1/16 and 1/8, respectively. In this experiment, we use the following parameters for the $NNSL_0$ algorithm: $\sigma_{\min} = 10^{-5}$, $k = n/2 = 135$, $L = 5$, $\mu = 2$, $d = 0.75$, and $\epsilon = 10^{-2}$. Here, the performance of SRC is computed for the algorithms SL_0, $NNSL_0$, and their results are compared. The test results are presented in Figure 1. In the following performance of SRC is computed for the algorithms SL_0, $NNSL_0$, WSL_0, $WNNSL_0$, and their results are compared. The test results are presented in Figure 2. It's important to note that in the final dictionary an identity matrix is added and their coefficients can be negative. The recognition rate using this method ($NNSL_0$) on the Extended Yale B dataset in comparison to the result of the SL_0 algorithm illustrated in the figure. As shown, the recognition rate of the proposed $NNSL_0$ method is higher than SL_0 algorithm, which demonstrates the accuracy of using the non-negativity constraint in the SRC.

As shown in simulation results, the $WNNSL_0$ algorithm has low computational power and high accuracy compared to the proposed algorithm base on SR classification. When the number of features is high, has the much higher performance as other proposed algorithms, but in each case it has a lower computational volume and a much higher speed, and its ideal identification rate is approximately 100%.

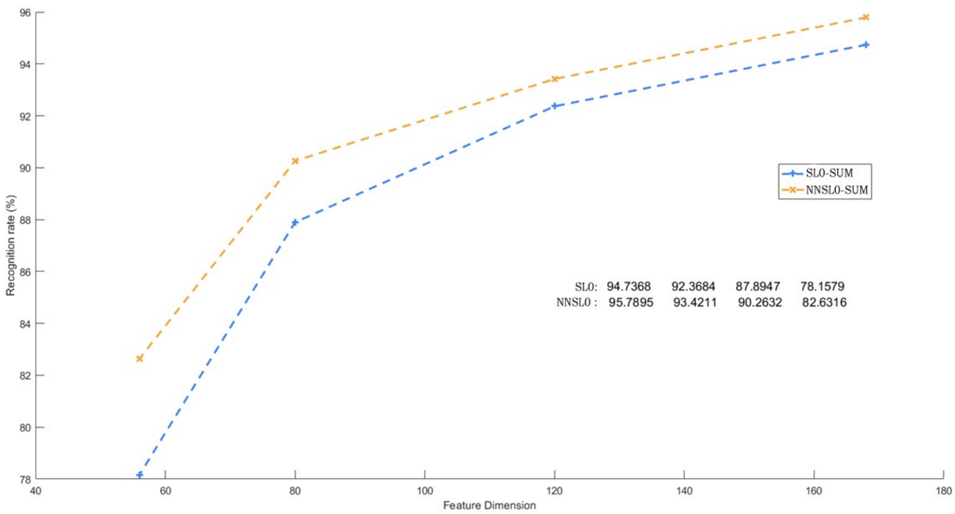

Figure 1: the face recognition rate for the data base "Extended Yale B"

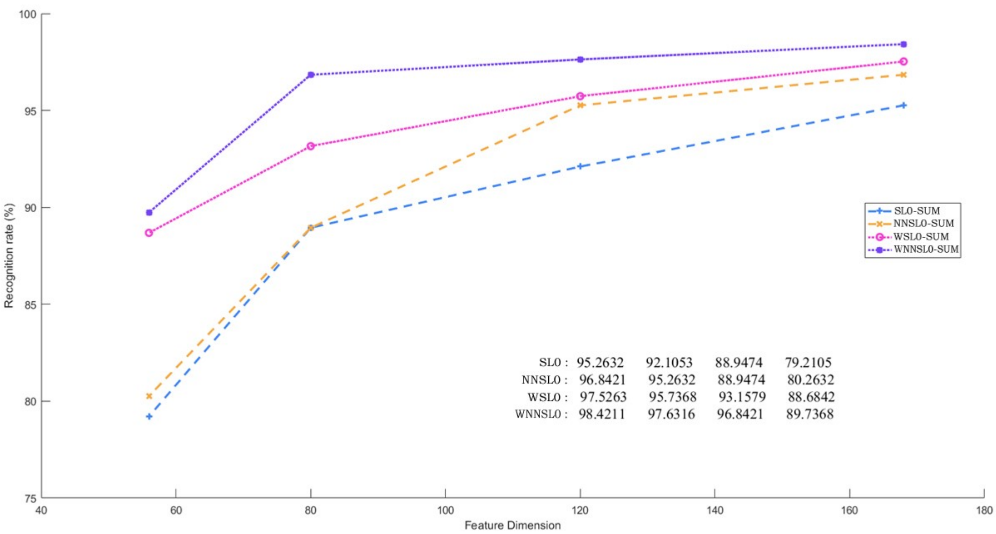

Figure 2: the face recognition rate for the data base "Extended Yale B"

5. Conclusions

In this paper, in addition to the performance of the classification methods base on SRC in face recognition, a sparse solver with non-negative constraint was suggested that minimizes a weighted L_0 norm instead of L_0 norm. In the weighted L_0 norm, the penalty of negative coefficients is larger than positive ones unlike L_0 norm where the penalty of all non-zero coeffi- cients is equal. We demonstrated that if the penalty of negative coefficients is larger than a finite value, it is sufficient to force all coefficients to be non-negative. The introduced algorithm for optimizing the weighted L_0 norm is based on non-negative Smoothed L_0 norm ($NNSL_0$) that tries to minimize the discontinuous cost function in a coarse to fine approach. Then in the proposed algorithm to achieve better performance, the combined method of NS and SR classifiers using weighted L_0-norm and SL_0 and Non-negative sparse decomposition has been investigated that We call the $WNNSL_0$ algorithm. In other words, we used the idea of the $NNSL_0$ algorithm to achieve Better performance and precision in SRC and the combination of weighted L_0-norm, SL_0, NS simple classifier and Non-negative sparse decomposition to achieve better classification performance, also $WNNSL_0$ algorithm performs better when the number of features is high. The proposed algorithm was assessed on some simulated data that confirms the higher accuracy and efficiency of the proposed algorithm compared with the other sparse algorithms. Some of the works that can be done in the future are to improve the estimation of weighted L_0-norm weights and use the different algorithms for feature extraction in terms of the challenges available to increase the performance of proposed methods.

References

[1] X. Liu, L. Lu, Z. Shen and K. Lu, *A novel face recognition algorithm via weighted kernel sparse representation*, Future Gen. Comput. Syst. 80 (2018) 653–663.

[2] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, *From few to many: Illumination cone models for face recognition under variable lighting and pose*, IEEE Trans. Pattern Anal. Machine Intel. 23(6) (2001) 643–660.

[3] Y. Adini, Y. Moses and S. Ullman, *Face recognition: The problem of compensating for changes in illumination direction*, IEEE Trans. Pattern Anal. Machine Intel. 19(7) (1997) 721–732.

[4] C. Ding, J. Choi, D. Tao and L.S. Davis, *Multi-directional multi-level dual-cross patterns for robust face recognition*, IEEE Trans. Pattern Anal. Machine Intel. 38(3) (2015) 518–531.

[5] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang and Y. Xu, *Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes*, IEEE Trans. Pattern Anal. Machine Intel. 39(1) (2016) 156–171.

[6] X. Luan, B. Fang, L. Liu, W. Yang and J. Qian, *Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion*, Pattern Recog. 47(2) (2014) 495–508.

[7] I.A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S.K. Shah and T. Theoharis, *3D-2D face recognition with pose and illumination normalization*, Comput. Vision Image Under. 154 (2017) 137–151.

[8] W. Zhang, X. Zhao, J.M. Morvan and L. Chen, *Improving shadow suppression for illumination robust face rcognition*, IEEE Trans. Pattern Anal. Machine Intel. 41(3) (2018) 611–624.

[9] M.R. Mohammadi, E. Fatemizadeh and M.H. Mahoor, *Non-negative sparse decomposition based on constrained smoothed ℓ_0 norm*, Signal Proces. 100 (2014) 42–50.

[10] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry and Yi Ma, *Robust face recognition via sparse representation*, IEEE Trans. Pattern Anal. Machine Intel. 31(2) (2008) 210–227.

[11] L. Zhang, M. Yang and X. Feng, *Sparse representation or collaborative representation: Which helps face recognition?*, Int. Conf. Comput. Vision, IEEE, (2011) 471–478.

[12] Y. Gao, J. Ma and A.L. Yuille, Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples, IEEE Trans. Image Proces. 26(5) (2017) 2545–2560.

[13] J. Wang, C. Lu, M. Wang, P. Li, S. Yan and X. Hu, *Robust face recognition via adaptive sparse representation*, IEEE Trans. Cyber. 44(12) (2014) 2368–2378.

[14] D. Donoho and M. Elad, *Optimal sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proc. Nat. Acad. Sci. 100(5) (2003) 2197–2202.

[15] D. Donoho, *For most large underdetermined systems of linear equations the minimal ℓ_1 norm solution is also the sparset solution*, Commun. Pure Appl. Math. 59(6) (2006) 797–829.

[16] H. Mohimani, M. Babaie-Zadeh and C. Jutten, *A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm*, IEEE Trans. Signal Proces. 57(1) (2008) 289–301.

[17] A. Ghaffari, *Two dimensional sparse decomposition and its application to image denoising*, MSc Thesis, Sharif University of Technology, 2009.

[18] X. He, S. Yan, Y. Hu, P. Niyogi and H. Zhang, *Face recognition using Laplacianfaces*, IEEE Trans. Pattern Anal. Machine Intel. 27(3) (2005) 328–340.

[19] R. Basri and D. Jacobs, *Lambertian reflectance and linear subspaces*, IEEE Trans. Pattern Anal. Machine Intel. 25(2) (2003) 218–233.

[20] M. Babaie-Zadeh, B. Mehrdad and G.B. Giannakis, *Weighted sparse signal decomposition*, Proc. IEEE Int. Conf. Acoustics, Speech Signal Proc. 2012 (2012) 3425–3428.

[21] A. Eftekhari, M. Babaie-Zadeh, C. Jutten and H.A. Moghaddam, *Robust-SL_0 for stable sparse representation in noisy settings*, Proc. IEEE Int. Conf. Acoustics, Speech Signal Proc. (2009) 3433—3436.

[22] B. Horn, *Determining lightness from an image*, Comput. Graph. Image Proc. 3.4 (1974) 277–299.