Sparse Representation-based Classification of Farsi Handwritten Digits Using Fisher Discrimination Criterion and Local Linear Embedding (LLE)

Amir Bahador Bayat

Tafresh University, Iran

amirbahador87@ymail.com

***Abstract*-In recent years, sparse representation-based methods have performed well in machine vision and image processing. The main challenge in designing a proper classifier to detect Farsi digits is modeling the data subspace and classification based on the presented model. In this paper, we propose a method based on sparse representation. Local Linear Embedding (LLE) to recognize the handwritten Farsi digits. The conducted approach to recognizing Farsi digits in this paper is sparse representation, based on Fisher Discrimination Criterion for linear transformation of the data. In this method, data sets belonging to the same class are grouped together, whereas data sets that belong to different classes are kept apart. LLE is used as a regulator to maintain the local neighborhood of the data. Experimental results of Hoda database test data with 80,000 samples indicate that the proposed method has a higher accuracy than previously presented methods and has achieved the accuracy of 99.36% for 60,000 training samples and 20,000 test data.**

Keywords- Fisher Discrimination Criterion; Local Linear Embedding (LLE); Handwritten Digits; Sparse Representation

I. INTRODUCTION

To classify a set of handwritten digits means assigning each image to one of the pre-defined classes. Considering that there are usually a large number of images, it is ideal to create a machine that automatically classifies the digits. First, the machine receives images with labels indicating the class that image belongs to. Then, in test stage, a few test images are given to the machine to classify. Thus, the digit classification task turns into machine training. Generally, a machine consists of two basic parts; feature extraction and classifier. In feature extraction part, a feature vector is extracted for each image which represents the prominent features of the input image. Then the extracted feature vectors are employed to train and test the classifier. Several successful methods have been proposed to recognize the Latin digits and letters [1]. However, fewer efforts have been made in recognizing Farsi (Arabic) digits, and recognition rate is also far less in this field [2, 3].

Handwritten digits recognition is mostly applied in processing the data entry forms, such as bank data entry forms and entrance exam answer sheets. Auto processing the forms is not yet practiced in many organizations, and information is entered into computers by human users. Handwritten digit recognition is also applied in car number recognition and converting the scanned pages into texts. Considering the importance of Farsi/Arabic language in Iran and neighboring countries, several researches have been conducted to solve some of the problems with handwritten Farsi/Arabic digits recognition.

Despite the conducted studies for recognition of handwritten Farsi digits, this case still faces several challenges. One of the main challenges in recognition of Farsi digits is that some of them can be written in two different shapes [4, 5]. Fig. 1 demonstrates the double-writing of the mentioned digits.

0	1	2	3	4	5	6	7	8	9
٠,٥	١	٢,٢	٣	٤,۴	٥,۵	٦,۶	٧	٨	٩

Fig. 1 Samples of different handwritings in the handwritten digits set

In Farsi language, due to the similarity between the digits and difference in their writings, creating a recognition system accurate enough for scientific use remains to be a problem. Thus, it is necessary to develop accuracy improving methods for them.

Digit recognition systems generally consist of three main sections: 1- Pre-process, 2- feature extract and 3- classification. Each of these sections consists of different Subsections: For example, in pre-process section, there are noise canceling and data normalizing subsections. Several methods have been presented for feature extraction section. The most common methods include moment contours [6], shadow code [7], and wavelet transform [9]. For digits classification, some of the employed methods in literature are divided and conquer method [10, 11], SVM [8] and neural network [7, 12].

One of the methods that have performed well in different fields is sparse representation. The use of sparse representation in image processing has considerably increased the efficiency of algorithms in machine vision and other applications [13]. One of the primary approaches to creating a sparse representation-based classifier is using the Fisher Discrimination Criterion. Fisher's linear classification transfers the data by linear transformation to a place where Fisher Discrimination Criterion is maximum. This criterion is obtained by dividing the inter-class scatter by intra-class scatter. Thus data belonging to the same class will get closer to each other and farther from other classes [13]. The main disadvantage of this method is ignoring the local neighborhood of the data. More precisely, this method obtains the sparse representation of data of each class without considering which data sets are in local neighborhood of other data sets. In fact, it is possible that two neighboring data sets belonging to the same class become non-neighbors after being transferred to sparse representation space. This increases the probability of classifier overfitting.

In this paper, in order to overcome the challenges in recognition of Farsi digits, a novel sparse representation-based system is introduced using Fisher Criterion [13], which maintains the local pattern of the data using Local Linear Embedding [14].

This paper is organized as follows: In the first section, feature extracting methods are introduced. In second section, sparse representation is introduced, and in the third section, classification by sparse representation is presented. In the fourth section, we will introduce the proposed method, and in the fifth section, experimental results of the proposed method are presented.

II. FEATURE

Feature Extraction is a vital step in pattern recognition. In this chapter, six sets of features are extracted. These feature sets and the methods of extracting them are summarized below:

A. PCA Features

One of the most important methods to reduce the redundancy in order to facilitate the data processing and image classification is Principal Component Analysis (PCA) [15]. This transform is widely used in data analysis to reduce dimensions. PCA transmission contains the majority of the basic information of the data besides reduction of the dimensions.

Assume M vectors of $T_1,T_2,\ldots,T_M$, are vectors of training image set. A is called the mean vector, and X_m is the difference between the mean vector and feature vectors. Which are determined by Eqs. (1) and (2).

$$A = {}^{1}/_{M} \sum_{m=1}^{M} T_m \tag{1}$$

$$X_m = T_m - A \qquad ,1 < m < M \tag{2}$$

If we define matrix Y by Eq. (3), then covariance matrix will be calculated by Eq. (4).

$$Y = [X_1\ X_2\ \ldots\ X_M] \tag{3}$$

$$C = {}^{1}/_{M} \sum_{m=1}^{M} X_m X_m^T \tag{4}$$

Covariance matrix is an N×N matrix, and using PCA method we can obtain K, the important eigenvector of matrix C.

Eigenvectors obtained in previous stage are sorted in order of importance, from the most to the least, based on data components. In this stage we create a feature vector which is actually a matrix of vectors. The matrix contains the feature vectors that will be used in the future. After the PCA transformation, the data sets will become 200 dimensional feature vectors.

B. MAT-Based Gradient Features

The next approach to extracting the features is to use MAT-Based Gradient Features method. MAT is a method of highlighting the center skeleton of the character strokes with maximum grayscale values and keeping the stroke information and local information, which has richer information for image processing and pattern recognition [16]. Our iterative MAT algorithm is implemented as follows:

1. Design the structure matrix of erosion as a 3×3 matrix E with all elements being set to 1 and set the initial iteration number as 1;
2. Erode a 32 × 32 character image Im by the morphological erosion operator E, and the value of the eroded pixel in the character image is set equal to the current iteration number;
3. Increase the iteration number by 1, then repeat step 2 until no more new eroded pixels are created.

After applying MAT to the image, pixels of the normalized image are scaled from 0 to 1. Convolution process is performed by the Sobel operator on the normalized image, and finally the number of non-zero gradients of each pixel of the convoluted image is considered as a gradient feature [9]. Figs. 2 (a-b) show two binary images of Character "9", selected from Hoda. Figs. 2 (c-d) are the MAT transformed character images of Figs. 2 (a-b), respectively.

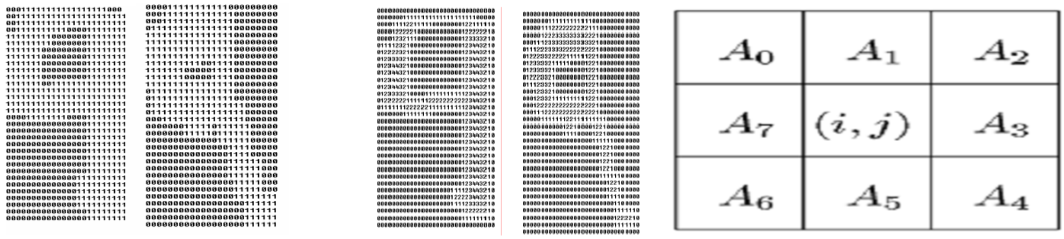

Fig. 2 Two binary character images (a, b) and their pseudo-grayscale character images after MAT (c, d), defining 8 neighbors for pixel (i, j)

C. Oriented Binary Gradient Features

The method for extracting the oriented binary gradient feature is similar to that of MAT, only MAT process is not applied to it. Gradient features are extracted directly from each binary character image.

The X-gradient character image can be calculated by:

$$I_x = I_z * S_x$$

and the Y-gradient character image is calculated by:

$$I_y = I_z * S_y$$

The gradient magnitude and phase are then obtained by:

$$r(i,j) = \sqrt{{I_x}^2(i,j) + {I_y}^2(i,j)}$$

$$\theta(i,j) = \tan^{-1} \frac{{I_y}^2(i,j)}{{I_x}^2(i,j)}$$

S_x , S_y are Sobel operator templates. Each normalized gradient image is divided into 16 sub-images. The number in each direction of each sub-image is counted as a feature.

Finally, a binary gradient feature vector for each digit image is produced.

D. Oriented Wavelet Transform Features

In this method, we use Kirsch non-linear edge enhancement algorithm in order to extract the statistic features of digit images. Kirsch operator has the advantage of directly calculating the gradient intensity in different directions. Each pixel (i, j) has an 8-pixel neighborhood around itself. A schematic arrangement of the pixels is demonstrated in Fig. (2).

Eq. (5) is used to extract features in 4 directions.

$$\boldsymbol{G(i,j)} = \mathbf{max}\{\mathbf{1}, \boldsymbol{max}_{\boldsymbol{k=0}}^{7}[\mathbf{5}\boldsymbol{S_k} - \mathbf{3}\boldsymbol{T_k}]\} \tag{5}$$

$$\boldsymbol{S_k} = A_k + A_{k+1} + A_{k+2}$$

$$\boldsymbol{T_k} = A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + \mathrm{A}_{k+7}$$

In order to extract four directional features from horizontal (H), Vertical (V), Right-diagonal (R) and Left-diagonal (L) directions, we can use the following templates:

$$G(i,j)_H = \max(|5S_0 - 3T_0|, |5S_4 - 3T_4|)$$

$$G(i,j)_V = \max(|5S_2 - 3T_2, |5S_6 - 3T_6|)$$

$$G(i,j)_R = \max(|5S_1 - 3T_1|, |5S_5 - 3T_5|)$$

$$G(i,j)_L = \max\left(|5S_3 - 3T_3|, |5S_7 - 3T_7|\right)$$

Then wavelet transform is applied on these features to obtain the main feature (GWF). Details of this algorithm are given in [16].

E. Median Filter Gradient Features

In order to extract the median filter gradient feature (MFGF), first, each image passes through a median filter, factors of which are given in Fig. 3.

1	2	1
2	4	2
1	2	1

Fig. 3 Template of 2D median filter

Then Robert operator is applied on obtained images to produce amplitude and phase. Finally, gradient feature extraction method is applied on it to yield the ultimate feature vector [9].

F. Complex Wavelet Transform Features

In complex wavelet transform (CWT) feature extracting method, in addition to the fact that transform features are maintained, we also exploit the lack of sensitivity to small displacements in image. In this method, first, low-pass and high-pass filters are applied to the image. Then each image is divided into four subbands. This procedure consists of three stages. Wavelet factors used in this method are extracted from level-3 LL subband which finally will produce the ultimate feature vector [17].

Fig. 4 demonstrates a schematic algorithm of complex wavelet transform.

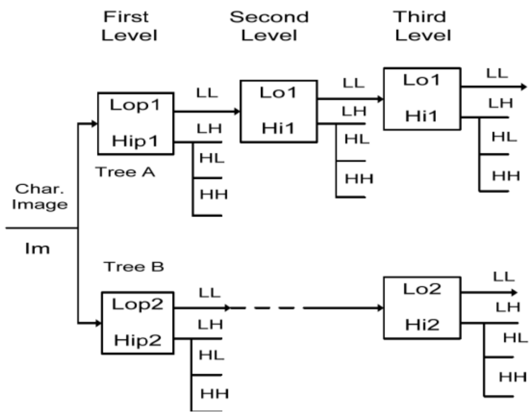

Fig. 4 schematic complex wavelet transform algorithm

III. SPARSE REPRESENTATION

Recently, sparse representation has been used widely in applications such as image sampling, compressing, representation, recovering and classification [12, 18]. The success of sparse representation is due to the fact that by considering some specific basics, most of the natural signals such as image and voice possess the sparse representation. This method was first presented by Olshausen [19].

AssumeX $= [x1, \ldots, xn]^T \in R^n$. This signal can represent an image. Sparsity implies that this information can be stated as a linear combination of a few bases. If k represents the number of bases which are important in representing the data, then (k<<n).

Bases used to represent the data form a Dictionary matrix. Column of this matrix is used to form the whole or part of the data. Each column of the dictionary matrix is called an atom.

If atoms of the dictionary increase, then it would become an over-complete dictionary. If D represents the dictionary, then the data x in x=Dα is stated by linear combination of atoms, and α represents the linear combination. Because the dictionary is over-complete, the linear system of equations is underdetermined for α and it will have infinite number of solutions. Since we are looking for sparsest representation, we use an additional condition. More precisely, we have to add a limitation to the problem to limit the number of non-zero components of sparse representation. The solution of optimization Eq. (6) yields the sparse representation of signal x.

$$\min_{\alpha} \|\alpha\|_0 \text{, subject to } x = D\alpha \tag{6}$$

In this equation, α is a vector that contains the sparse representation coefficients of the signal. Zero norm $\|\ \|_0$ counts the number of non-zero components of the matrix. This number indicates the Sparsity of representation. This problem is NP-Hard. One the best methods to approximately solve this problem is OMP method [20-23]. An approach could use zero-norm instead of 1-norm to solve this problem, which will change Eq. (6) as follows:

$$\min_{\alpha} \|x - D\alpha\|_2 \text{, s.t } \|\alpha\|_1 = L \tag{7}$$

It has been shown that in most applications, the solutions to above problem and Eq. (7) are equal, and that the above problem is an acceptable approximation to sparse representation of signal.

IV. CLASSIFICATION BY SPARSE REPRESENTATION

Most of the papers presented in sparse representation in machine vision and image processing focus on image recovery. Sparse representation of signals is also used in classification applications that aim to intensify the discrimination against different classes. Thus, the proposed methods can be divided into two groups; recovery aimed methods and discrimination aimed methods. In signal recovery problems, all of the signals are in the same class, so it is easy to define a dictionary. However, in classification problems, it is a more complex task to learn the dictionary, because the data representation by this dictionary must improve the discrimination against different classes.

A. Dictionary Learning by Fisher Criterion

One of the primary approaches to creating a classification model uses Fisher Discrimination Criterion. Fisher linear classification transfers the data by linear transformation to a space in which the Fisher Discrimination Criterion is maximum. This criterion is calculated by dividing inter-class scatter by intra-class scatter. Thus data sets belonging to the same class will get closer to each other and farther from other classes. In 2011, a method called “discrimination dictionary learning” was presented based on Fisher Discrimination criterion [13]. In this method, a structure has been considered for dictionary, which is based on data labels. We place the problem data in a column in matrix $X = [X1, X2, \dots, Xc]$, Xi represents the data belonging to class i. In this method, a similar structure has been considered for dictionary atoms. When D represents the dictionary, we can say: $D = [D_1, D_2, \dots, D_c]$, where matrix D_i consists of atoms which are suitable for recovering the data related to class i. Based on the mentioned structured dictionary, an optimization equation is defined for signal recovery and also separating the maximum according to Fisher discrimination criterion. Eq. (8) presents the optimization of the mentioned method.

$$f(D, A) = min_{D,A} r(D, A_i, X_i) + \textstyle\sum_i \|A_i\|_1 + f(A) \tag{8}$$

where A is sparse representation and f is a function defined based on sparse representation, which optimizes the Fisher discrimination criterion. Also in this equation $\|A_i\|_1$ is 1-norm summation for each matrix column. Function r (D, A, X) applies the signal recovery error on optimization, considering the structured dictionary as shown in Eq. (9):

$$r(D, A_i, X_i) = \| X_i - DA_i\|^2 + \|X_i DA_i^i\|^2 + \sum_{j \neq i} \| D_j A_i^j \|^2 \tag{9}$$

When the new dictionary representation of the data is saved in matrix X, Eq. (10) and (11) show the intra-class and inter-class scatters, respectively. Where mi represents the class i average and m represents the total average.

$$S_W(X) = \sum_{i=1}^{c} \sum_{x_k \in x_i} (x_k - m_i)\,(x_k - m_i)^T \tag{10}$$

$$S_B(X) = \sum_{i=1}^{c} n_i (x_k - m_i)\,(x_k - m_i)^T \tag{11}$$

Hence the function f is defined as Eq. (12). The first two terms are from Fisher criterion and the third term is to make the problem convex and solve the optimization.

$$f(A) = tr\left(S_W(A)\right) - tr\left(S_B(A)\right) + \eta \|A\|_F^2 \tag{12}$$

A complete description of Fisher dictionary learning algorithm is illustrated in Fig. 5.

Fisher Discrimination Dictionary Learning

1. **Initialization D**. We initialize all the p_i atoms of each D_i as random vectors with unit l_2-norm
2. **Update the sparse coding coefficients A**. Fix D and solve X_i, i=1,2,…,c, one by one by solving Eq. (7).
3. **Updating dictionary D**. Fix A and update each Di, i=1,2,…,c, by solving Eq. (8) with the method presented in [22].
4. **Output.**

Return to step 2 until the values of f(D,A)in adjacent iterations are close enough, or the maximum number of iterations is reached. Output A and D.

Fig. 5 Fisher criterion dictionary learning algorithm

V. LOCAL LINEAR EMBEDDING LLE

One of the common problems in all sparse representation-based classification methods is ignoring the local neighborhood of the data. More precisely, these methods obtain the sparse representation of data of each class without considering which data sets are in local neighborhood of other data sets. In other words, it is possible that two neighboring data sets belonging to the same class, (X neighbors Y, if and only if Euclidean distance between X and Y is shorter than the Euclidean distance between Y and any other Data) become non-neighbors after being transferred to sparse representation space. This increases the probability of classifier overfitting. In order to overcome the local neighborhood problem in sparse representation-based classifications, we use the local linear embedding concept. The main idea of the algorithm [6], which is mostly used in dimension reduction, is as follows:

Assume we have N data sets $[x_1,\ldots,x_N]$, each of which is D-dimensional. This algorithm transfers the data by a transform matrix (reduces the dimensions of the data) so that the local distance between data sets is maintained. The method conducted by this algorithm to maintain the distance is that it finds k-nearest neighbor of each data set, then it estimates the level of maintenance of local distance using Eq. (13).

$$W^* = \arg min_{W=[w_{ij}]} \sum_{i=1}^{N} \| x_i - \sum_{x_{j\in N_k(x_i)}} w_{ij}\, x_j \|^2 \quad s.t. \forall i, \sum_{x_{j\in N_k(x_i)}} w_{ij}\, x_j = 1 \tag{13}$$

where W is a $N \times N$ matrix which represents the local neighborhood level. In [14], it is used to solve the above optimization problem.

VI. THE PROPOSED METHOD

A. *Combining LLE Algorithm and Sparse Representation-based Classification Methods*

Generally, in sparse representation-based classification methods, sparse representation of the data ($A = [\alpha_1, \alpha_2, \ldots, \alpha_N]$), dictionary (D), and classifier (f) are learnt from the data. Learning these three concepts is expressed as an optimization problem. Each of the proposed methods optimizes a different objective function which is represented by F (A, D, F). For example, the objective function for discrimination dictionary learning method using Fisher criterion is as Eq. (8). By applying LLE algorithm, the objective function will change as Eq. (14).

$$F'(A,D,f) = F(A,D,f) + \gamma \sum_{i-1}^{N} \| \alpha_i - \sum_{j} W_{ij}^* \alpha_j \|^2 \tag{14}$$

where γ is the regularization coefficient learnt through cross validation.

By adding the second right term in the above equation, we add a limitation to Sparsity codes that not only they must be able to recover the principal signals properly, but also the sparse codes correspondent to neighboring signals must be similar. Since the term added to F is a quadratic function with respect to sparse coefficients, if F was convex, then the convex form of the objective function would be maintained, hence optimization of the above equation would be performed easily.

Since the method presented in [13] is one of the best methods in sparse representation-based classification, we presume the value of F (A, D, f) in the above equation to be equal to the proposed objective function in the paper. (Notice that the proposed method can be combined with any function F) So the proposed ultimate objective function (FDLLE) is expressed as Eq. (15):

$$min_{\alpha,D} \sum_i (\|X_i - D\alpha_i\|^2) + \|X_i - D_i\alpha_i^i\|^2 + \sum_{i\neq j} \|D_j\alpha_i^j\|^2 + \sum_i \|x_i\|_1) + tr\,(S_W(\alpha) - tr\,(S_B(\alpha) + \lambda_2 \|\alpha\|_F^2 + \gamma \sum_{i=1}^{N} \|\alpha_i - \sum_j W_{ij}^* \alpha_j\|^2 \tag{15}$$

Objective function is presented for optimization of sparse and dictionary coefficients values. The method presented in [13] is used to solve the above equation. We use [22] to obtain the primary dictionary coefficients and [21] to obtain the primary sparse coefficients. The obtained coefficient values in each stage are put in the objective function and are reduced as long as the algorithm allows us to. Finally, the ultimate coefficients are obtained. Algorithm performance of a test data Y is as follows:

$$\widehat{\alpha} = \arg min_\alpha \| y - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \tag{16}$$

$$label = \arg min_i \|y - D_i\widehat{\alpha}\|_2^2 + w\|\hat{\alpha} - m_i\|_2^2 \tag{17}$$

First, sparse coefficients of the test data are obtained by Eq. (16), and then, the test data label is obtained by Eq. (17).

The number of correct labels is counted and considered as our proposed system performance.

General description of the proposed algorithm is demonstrated in Fig. 6.

proposed method's algorithm Dictionary Learning

1. **Initialization D**. We initialize all the p_i atoms of each D_i as random vectors with unit l_2-norm
2. **Update the sparse coding coefficients A**. Fix D and solve X_i, i=1.2.…,c, one by one by solving Eq. (15).
3. **Updating dictionary D**. Fix A and update each Di, i=1,2,…,c, by solving Eq. (8) with the method presented in [22].
4. **Output.**

Return to step 2 until the values of f(D,A)in adjacent iterations are close enough, or the maximum number of iterations is reached. Output A and D.

Fig. 6 Proposed method's algorithm

VII. EXPERIMENTAL RESULTS

In this paper, we used Hoda database of Farsi handwritten digits [20]. 6000 samples of each class were used for training and 2000 samples for testing.

Fig. 7 shows samples of different handwritings in the database. Database images are not similar in pattern, which leads to a better evaluation of digit recognition systems.

Fig. 7 Samples with different writing styles in the dataset

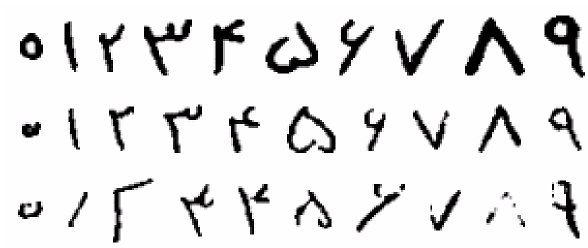

Fig. 7 Samples with different qualities in the dataset

In order to evaluate the proposed method in this paper, first we call the testing data in Hoda database and then use Eq. (17). We obtain labels of each test data and compare them with real data labels.

The performance of the proposed classifier for handwritten digits classification is analyzed using various performance metric tests. The accuracy of classifying a given handwritten digits into a particular category is calculated by comparing it with the actual test class. The formula for calculating accuracy is given below:

$$Accury(\%) = \frac{Total\ Label\ of\ correct\ predictions}{Total\ Label\ of\ in\ the\ test\ data\ set} * 100 \tag{18}$$

Total Number of correct predictions is calculated by Function (18) .The number of correctly recognized labels is the performance of our proposed algorithm. Performance of the proposed algorithm for each feature presented in Section 1 is given in Table 1.

TABLE 1 PERFORMANCE OF THE FDLLE METHOD

Feature	**BGDF**	**PCA**	**GWF**	**CWT**	**MGF**	**MFGF**	**CWT+PCA**
proposed method(FDLLE)	**95.9**	**97.9**	**98.37**	**99.01**	**93**	93.44	**99.36**

Results of the proposed algorithm are compared with results of some sparse representation-based methods and standard methods, such as Bayesian and K-nearest neighborhood. Comparison between performances of the proposed methods and other methods is presented in Table 2. Hoda database of handwritten digits is used to correctly evaluate the results.

TABLE 2 COMPARISON OF THE PROPOSED METHOD AND OTHER METHODS

method	BGDF	PCA	GWF	CWT	MGF	MFGF	CWT+PCA
k-NN	90.37	92.37	91.3	92.8	74.1	90.01	95.55
Bayes	90.7	92.7	92.14	93.6	78.5	91.14	96.06
MLP	91.22	95.22	94.15	96.44	78.6	**95.14**	96.78
MSRC	92.91	96.91	96.72	97.18	79.04	91.39	97.55
FDLLE	**95.9**	**97.9**	**98.37**	**99.01**	**93**	93.44	**99.36**

MSRC algorithm is a simple method based on sparse representation. First, it calculates the sparse representation; and then, it decides based on each class' data considering the recovery error. The recovery with the lowest error is chosen as the entry data class.

To compare the performance of our method, we noted the performances of most works that were available for Persian numeral recognition. Table 3 shows the details of comparison. It may be noted from Table 3 that all the existing works were evaluated on smaller datasets. The highest dataset of size 10,000 was used by a recent work due to Ziaratban et al. [8], whereas we used 80,000 data for our experiment. The highest accuracy was obtained from the work of Soltanzadeh et al. [3] but they have experimented with only 8,918 samples and used 257 dimensional features. We considered 80,000 data for our system and we obtained 98.71% and 99.03% accuracies using only 196 dimensional features.

TABLE 3 COMPARISON OF DIFFERENT ALGORITHMS

Algorithms	Dataset size		Accuracy (%)
	Train	Test	Test
Shiralishahreza et al. [24]	2600	1300	97.80
Soltanzadeh, Rahmati [25]	4979	3939	99.57
Harifi., Aghagolzadeh [26]	230	500	97.60
Ziaratban et al. [27]	6000	4000	97.65

Mowlaei, Faez [28]	2240	1600	91.88
Mozaffari et al. [39]	2240	1600	94.44
Sadri et al. [30]	6000	2000	98.71
Proposed Algorithm	**6000**	**2000**	**99.36**

VIII. CONCLUSION

Sparse representation is a modern method to classify the data, especially images. The main advantage of sparse representation is that it creates a non-linear model to represent the data. In fact, by optimizing 1-norm, proper subspace is learnt for each data. As explained before, one of the disadvantages of these methods is ignoring the local neighborhood of the data after transferring them into sparse subspace, which is partly solved by the proposed method using local linear embedding. This method shows a better performance with finite number of data than other sparse methods. Initially, this method seems to be inconsistent considering the locality of it and its objective of maintaining the local structure. But notice that the local data are used for learning dictionary and the subspace according to the local structure, it improves the classification accuracy.

REFERENCES

[1] Cheng L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, “Handwritten digit recognition: benchmarking of state-of-the-art techniques,” *Pattern Recognition*, vol. 36, pp. 2271-2285, 2003.

[2] M. H. Shirali-Shahreza, K. Faez, and A. Khotanzad, “Recognition of handwritten Persian/Arabic numerals by shadow coding and an edited probabilistic neural network,” *Proceedings of the international conference on image processing*, vol. 3, pp. 436-439, 1995.

[3] S. Abdleazeem and E. El-Sherif, “Online Arabic handwriting recognition: a survey,” *Computer Science of springer*, vol. 16, iss. 3, pp. 209-226, Sept. 2013.

[4] S. Abdleazeem and E. El-Sherif, “Arabic hand written digit recognition,” *IJDAR*, vol. 11, iss. 3, pp. 127-141, 2008.

[5] J. Sadri, C. Y. Suen, and T. D. Bui, “Application of Support Vector Machines for Recognition of Handwritten Arabic/ Persian Digits,” *Proceedings of the 2nd Conference on Machine Vision and Image Processing Applications*, vol. 1, pp. 300-307, 2003.

[6] M. Dehghan and K. Faez, “Farsi Handwritten Character Recognition with Moment Invariants,” *Proceedings of 13th International Conference on Digital Signal Processing*, vol. 2, pp. 507-510, 1997.

[7] H. Shirali-Shahreza, K. Faez, and A. Khotanzad, “Recognition of Hand-written Persian/Arabic Numerals by Shadow Coding and an Edited Probabilistic Neural Network,” *Proceedings of International Conference on Image Processing*, vol. 3, pp. 436-439, 1995.

[8] Alaei Alireza, Pal Umapada, and Nagabhushan P, “Using Modified Contour Features and SVM Based Classifier for the Recognition of Persian/Arabic Handwritten Numerals,” *Paper presented at the meeting of the ICAPR*, 2009.

[9] Ping Zhang, Bui Tien D., and Suen Ching Y., “A novel cascade ensemble classifier system with a high recognition performance on handwritten digits,” *Pattern Recognition*, vol. 40, no. 12, pp. 3415-3429, 2007.

[10] H. Parvin, H. Alinejad-Rokny, and S. Parvin, “Divide and Conquer Classification,” *Australian Journal of Basic and Applied Sciences*, vol. 5, no. 12, pp. 2446-2452, 2011.

[11] Shen-Wei Lee, “Effective multiple-features extraction for off-line SVM-based handwritten numeral recognition,” *IEEE*, pp. 194-197, 2012.

[12] Rajashe kararadhya S.V, “Neural network based handwritten numeral recognition of Kannada and Telugu scripts,” *IEEE*, pp. 4244-2409, 2008.

[13] Yang M., Zhang L., Feng X., and Zhang D., “Fisher discrimination dictionary learning for sparse representation,” *ICCV*, 2011.

[14] Roweis S. and Saul L., “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.

[15] Rui Li and Shiqing Zhang, “Handwritten Digit Recognition Based on Principal Component Analysis and Support Vector Machines,” *Communications in Computer and Information Science of springer*, vol. 214, pp. 595-599, 2011.

[16] P. Zhang, T.D. Bui, and C.Y. Suen, “Nonlinear feature dimensionality reduction for handwritten numeral verification,” *Pattern Anal. Appl.*, vol. 7(3), pp. 296-307, 2004.

[17] P. Zhang, T.D. Bui, and C.Y. Suen, “Extraction of hybrid complex wavelet features for the verification of handwritten numerals,” in: *Proceedings of the 9th International Workshop on Frontiers of Handwriting Recognition (IWFHR’9)*, Tokyo, Japan, pp. 347-352, 2004.

[18] Wang J., Lu C., Wang M., and Li P., “Robust Face Recognition via Adaptive Sparse Representation,” *IEEE Trans Pattern Anal Mach Intell.*, vol. 2, iss. 31, pp. 210-227, 2014.

[19] Chu-Kiong Loo and Ali Memariani, “Object Recognition Using Sparse Representation of Overcomplete Dictionary,” *IEEE*, vol. 76, pp. 75-82, 2012.

[20] H. Khosravi and E. Kabir, “Introducing a very large dataset of handwritten Farsi digits and a study on the variety of handwriting styles,” *Pattern Recognition Letters*, vol. 28, iss. 10, pp. 1133-1141, 2007.

[21] José M. Bioucas-Dias and Mário A. T. Figueiredo, “A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration,” *IEEE Transactions on Image Processin*g, vol. 16, iss. 12, pp. 2992-3004, 2007.

[22] Yang, M. and Zhang, L., “METAFACE LEARNING FOR SPARSE REPRESENTATION BASED FACE RECOGNITION,” *ICIP*, pp. 351-362, 2010.

[23] S. Mallat and Z. Zhang, “Matching pursuits with timefrequency dictionaries,” I*EEE Trans. Signal Processing*, pp. 3397-3415, 1993.

[24] M. H. Shirali-Shahreza, K. Faez, and A. Khotanzad, "Recognition of Hand-written Persian/Arabic Numerals by Shadow Coding and an Edited Probabilistic Neural Network," *Proceedings of International Conference on Image Processing*, vol. 3, pp. 436-439, 1995.

[25] H. Soltanzadeh and M. Rahmati, "Recognition of Persian handwritten digits using image profiles of multiple orientations," *Pattern Recognition Letters*, vol. 25, pp. 1569-1576, 2004.

[26] A. Harifi and A. Aghagolzadeh, "A New Pattern for Handwritten Persian/Arabic Digit Recognition," *Journal of Information Technology*, vol. 3, pp. 249-252, 2004.

[27] M. Ziaratban, K. Faez, and F. Faradji, "Language-Based Feature Extraction Using Template-Matching in Farsi/Arabic Handwritten Numeral Recognition," *Proceedings of 9th International Conference on Document Analysis and Recognition*, vol. 1, pp. 297-301, 2007.

[28] A. Mowlaei and K. Faez, "Recognition Of Isolated Handwritten Persiawarabic Characters And Numerals Using Support Vector Machines," *Proceedings of XIII Workshop on Neural Networks for Signal Processing*, pp. 547-554, 2003.

[29] S. Mozaffari, K. Faez, and M. Ziaratban, "Structural Decomposition and Statistical Description of Farsi/Arabic Handwritten Numeric Characters," *Proceedings of the 8th Intl. Conference on Document Analysis and Recognition*, vol. 1, pp. 237-241, 2005.

[30] Alaei A., "Using Modified Contour Features and SVM Based Classifier for the Recognition of Persian/Arabic Handwritten Numerals," *IEEE*, pp. 391-394.2009.