Off-line Recognition of Persian Handwritten Digits using Statistical Concepts

Omid Rashnoodi
Department of
Computer,
Khuramshahr Branch, Islamic
Azad University, Khuramshahr,
Iran

Asghar Rashnoodi
Department of
Of Management Branch,
khorramshahr Marin and
Science Technology University,
Iran

Aref Rashnoodi
Department of
Science and Research Branch,
Islamic Azad University
Khuzestan,
Iran

ABSTRACT
In this paper a novel method for recognition of Persian handwritten digits is proposed. This approach uses central moments, Covariance, Median and Variance that are obtained at each the box digit image as the features set. The features set consist of 140 dimensions for each instance. In the classification phase of our proposed method the support vector machines, K-Nearest Neighbor and Sequential Minimal Optimization separately are employed. In this paper is used the principal components analysis (PCA) to reduce dimension features set. the performance of these three classifiers is observed on this application in terms of the correct classification and misclassification and how the performance of K-Nearest Neighbor classifier can be improved by varying the value of k.To evaluate our proposed scheme a database of Persian handwritten digits consist of 1699 handwritten digit images is used. In the best case proposed scheme using SMO classifiers yields a recognition rate of 92.3875% for handwritten Farsi numerals.

Keywords
Central moments, Variance, Covariance, Median, support vector machine, principle component analysis, K-Nearest Neighbor, Sequential Minimal Optimization.

1. INTRODUCTION
Handwritten recognition is one of the subcategories of image processing that has many applications include automatic reading of postal addresses, bank checks and forms old documents, license plates recognition [35].optical character recognition (OCR) is one of the most successful applications of automatic pattern recognition[2] ,and handwritten digit recognition has long been an active topic in OCR and classification research. Numerous approaches have been proposed for pre-processing, feature extraction, classification, post–processing and some standard image databases are widely used to evaluate the performance [1]. an OCR system with a good recognition performance needs to maintain a very high recognition rate, and at the same time, to obtain a very high reliability, or a very low error rate[33,34].in the last few decades, numerous methods have been proposed for machine recognition of handwritten characters, especially for languages such as English, Japanese and Chinese due to the popularity of language use .Particularly, handwritten numeral recognition has attracted much attention, and various techniques (pre-processing, feature extraction, and classification) have been proposed [1,2,3,4,5]. In contrast, for the recognition of Persian (Arabic) handwritten digits very few works are reported [6, 7, 8, 9].

And now research on Farsi(Persian) scripts and numerals is receiving increasing attention because a lot of data such as addresses written on envelopes; amount written on checks; names, addresses, identity numbers, and Rial values written on invoices and forms were written by hand and they had to be entered into the computer for processing. Combining classifiers to achieve higher accuracy is an important research topic [10, 11, 12, 3, 28]. Essentially, the idea behind combining classifiers is based on the so-called divide and conquers principle, according to which a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks [14]. There are two main strategies in combining classifiers: fusion (static structures) and selection (dynamic structures) [15]. In classifier fusion, it is supposed that each ensemble member is trained on the whole feature space [11, 6], whereas in classifier selection, each member is assigned to learn a part of the feature space [16, 17, 18]. This way, in the former strategy, the final decision is made considering the decisions of all members, while in the latter strategy, the final decision is made by aggregating the decisions of one or a few of experts [14,19]. There are many feature extraction methods causing the recognition rate's improvement. in the last ten years a few paper were published on Arabic/Persian digits recognition and they used different methods for feature extraction such as geometric moments invariants, Zernike moments[29] and shadow coding[30].Two of best feature extraction methods for Farsi word recognition domain are Zoning [20, 21] and Gradient [22]. There are various methods in classification stage such Bays-normal [23], Decision Trees [24] Neural Networks [25], Statistical Classifiers [26] and Support Vector, Classifiers [27, 28]. Typically an ordinary optical character recognition (OCR) includes three indicated phase. Pre-processor, feature extraction and classifications phase that output of each step is the input of next step.In this paper in the phase Pre-processing Operations such as slant correction, normalization, and binarization have been done, in phase feature extraction of covariance, variance, central moments and the median is used. The last phases of SVM, K-NN and SMO separately as classifiers are used.

The paper is organized as follows. We In Section 2 feature extraction methods and principle component analysis, Section 3 classification, Section 4 block diagram the proposed method, Section 5 data set experimental results and comparative analysis are described. Finally in last section we present conclusion.

2. Feature Extraction Methods

The performance of character recognition largely depends on the feature extraction approach. To select a set of appropriate numerical attributes of features from the interested objects for the purpose of classification has been among the fundamental problems in the design of an imagery pattern recognition system. One of the solutions, the utilization of moments for object characterization has received considerable attention in recent year. In this paper, the new technique derived to increase the accuracy and the efficiency in moment computing is addressed. Based on these developments, the significant improvement on image reconstruction via central moments, covariance, variance and median has been achieved.

Since all the image size is 64*64(normalization), for extracting Suitable features from each image Each image has been divided into 64 squares with a size of 8 * 8 and the variance, covariance, mean, second-order central moment and the number of pixels per square as follows calculated and so the dimensional 320 will be feature vectors.

Features set I: Variance:

In this section, we calculated the variance of variances of each column of pixels within each square which are equal to a one value.

There are two common textbook definitions for the standard deviation s of a data vector x:

$$s = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{1}{2}} \qquad (2\text{-}1)$$

$$s = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{1}{2}} \qquad (2\text{-}2)$$

Where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

And n is the number of elements in the sample. The two forms of the equation differ only in n-1 versus in the divisor. The variance is the square of the standard deviation (STD).

Features set II: Covariance

In this section, we calculated the covariance of covariance's of each column of pixels within each square which are equal to a one value.

Covariance removes the mean from each column before calculating the result. The covariance function is defined as:

$$\mathrm{cov}(x_1, x_2) = E[(x_1 - \mu_1)(x_2 - \mu_2)] \qquad (2\text{-}3)$$

Where E is the mathematical expectation and $\mu_i = Ex_i$.

Features set III: Central moment

In this section, we calculated the second-order central moment of second-order central moments of each column of pixels within each square which are equal to a one value.

Note that the central first moment is zero, and the second central moment is the variance computed using a divisor of n rather than n – 1, where n is the length of the vector x or the number of rows in the matrix x.

The central moment of order *k* of a distribution is defined as:

$$m_k = E(x - \mu)^k \qquad (2\text{-}4)$$

Where $E(x)$ is the expected value of *x*.

Features set IV: Median

In this section, we calculated the median of medians of each column of pixels within each square which are equal to a one value.

Features set V: Numbers of pixel per each square

In this section, we calculated the numbers pixels within each square which are equal to a one value.

Therefore the length of our feature vectors is 5*64=320.

2.1 Principle component analysis

Due to the high dimensionality of visual features, dimension reduction is important for modern image retrieval systems. Principal Component Analysis (PCA) [36] is one of the most widely used dimensionality reduction methods. The advantage of the PCA transformation is that it is linear and that any linear correlation presented in the data is automatically detected. In our system, PCA is used to generate compact and comprehensive feature vectors so that an efficient index can be constructed for fast retrieval in the image database. PCA analysis examines the variance structure in the dataset and determines the directions along which the data exhibits high variance. The first principal component (or dimension) accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Using PCA, most of the information in the original space is condensed into a few dimensions along which the variances in the data distribution are the largest.

In the first stage of our proposed model the Principle Component Analysis (PCA) was used, to avoid a high dimensional and redundant input space and optimally design and train the experts.PCA is a useful statistical technique that has found application in fields such as face recognition and image compression [39], and is a common technique for finding patterns in data of high dimension.

It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is the one of powerful tool for analyzing data The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, i.e. by reducing the number of dimensions, without much loss of information [40].

Suppose T₁, T2… TM that are the projection vector of

Training data set and each of these vectors has N elements. The mean vector, A, is computed with the following equation:

$$A = \tfrac{1}{M} \sum_{m=1}^{M} T_m \qquad (2\text{-}1\text{-}1)$$

Subtract the mean from each of the data dimensions.

$$X_m = T_m - A \qquad (2\text{-}1\text{-}2)$$

$$Y = [X_1, X_2, ..., X_m] \qquad (2\text{-}1\text{-}3)$$

$$c = \tfrac{1}{M} \sum_{m=1}^{M} X_m X_m^T = \tfrac{1}{M} YY^{-T} \qquad (2\text{-}1\text{-}4)$$

Since the data is N dimensional, the covariance matrix will be N*N finally the eigenvectors and eigenvalues of the covariance matrix is calculated, and then the k most

Significant component are picked out and formed feature vector. Thus the PCA projection matrix projects input patterns from an N-dimensional image space to a K-dimensional subspace. We use PCA and reduce 320 dimension features set to 120 dimension features set.

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD).

3. CLASSIFICATION

3.1 support vector machine

Generally, an SVM solves a binary (two-class) classification problem and multi-class classification is accomplished by combining multiple binary SVMs. An M-class problem can be decomposed into M binary problems with each separating one class from the others, or into $\binom{M}{2}$ binary problems with each discriminating between a pair of classes. The former scheme is adopted in our experiments. On a pattern x, the discriminate at function of a binary SVM is computed by

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i k(x, x_i) + b \qquad (3\text{-}1\text{-}1)$$

Where l is the number of learning patterns, y_i is the target value of learning pattern x_i (+1 for the first class and - 1 for the second class), b is a bias, and k(x, xi) is a kernel function which implicitly defines an expanded feature space:

$$k(x, x_i) = \phi(x).\phi(x_i) \qquad (3\text{-}1\text{-}2)$$

Where $\phi(x)$ the feature is vector in the expanded feature space and may have infinite dimensionality. Types of kernels, polynomial, RBF and linear are computed such as table 1:

Table 1. Different kernels for SVM classifiers

Kernel	Definition
Linear	$k(x, x_i) = x.x_i$
polynomial	$k(x, x_i, p) = (1 + x.x_i)^p$
RBF	$k(x, x_i, \sigma^2) = \exp(-\dfrac{\|x - x_i\|^2}{2\sigma^2})$

Respectively. In our implementation, the pattern vectors are appropriately scaled for the polynomial kernel, with the scaling factors estimated from these self-inner products of sample vectors. While for the RBF kernel, the value of σ^2

Is estimated from the variance of sample vectors.

The discriminate function (1) can be viewed as a generalized linear discriminate function with weight vector

$$w = \sum_{i=}^{l} y_i \alpha_i \phi(x_i) \qquad (3\text{-}1\text{-}3)$$

Coefficient $\alpha_i, i = 1,2,3,...,l$ are determined on the learning patterns by solving the following optimization problem.

Minimize $\tau(w) = \tfrac{1}{2}\|w\|^2 \qquad (3\text{-}1\text{-}4)$

Subject to $y_i f(x_i) \geq 1 - \xi_i, \qquad \xi_i \geq 0, i = 1,2,...,l$

This is a quadratic programming problem and can be converted into the following dual problem:

Maximize:

$$w(x) = \sum_{i=1}^{l} \alpha_i - \tfrac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \qquad (3\text{-}1\text{-}5)$$

Subject to:

$$0 \leq \alpha_i \leq C, i = 1,2,3,...,l \qquad and \qquad \sum_{i=1}^{l} \alpha_i y_i = 0$$

Where C is a parameter to control the tolerance of classification errors in learning.

The solution of the quadratic problem (2) is not trivial on large sample size. Some algorithms have been pro- posed specially to solve the quadratic programming problem for SVM learning, among which is the sequential minimal optimization (SMO) of Platt [31] and the successive over relaxation (SOR) of Mangasarian [32].

SVM is a binary classifier. In order to use it for the classification of several classes there are two approaches, one against all and one against one [38]:

3.1.1 One against all

One of the earliest and simplest implementation of multi-class SVM is One-Against-All (OAA) method. In this method, the K binary SVM classifiers are used for classification where K is the number of classes and the separator line for each class separate it from all other classes. For classification in this method, each sample is given to all SVM classifiers and the winner class is selected [54, 38].

3.1.2 One against one

Another popular approach is to train K (K -1)/2 binary SVMs, where each binary SVM is trained on data points from two classes. This approach is called One-Against-One (OAO) method [55, 56]. After training K (K - 1)/2 binary SVMs, future testing is done using "Max Wins" voting strategy. An interesting modification to OAO is Directed Acyclic Graph S VM (DAGSVM) proposed in [57].In this method, k different binary SVM classifiers are used and each SVM separates class I from class J where I and J are different class indexes [38].

3.2 K-Nearest Neighbor Classier (KNN)

The K-nearest neighbor (KNN) classifier has been both a workhorse and benchmark classifier [41, 42, 43, 44, 45]. Among the various methods of supervised statistical pattern recognition, the Nearest Neighbor rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn; it involves a training set of both positive and negative cases [19]. A new sample is classified by calculating the distance to the nearest training case; the sign of that point then determines the classification of the sample [47, 48]. The KNN classifier extends this idea by taking the k nearest points and assigning the sign of the majority. The performance of a KNN classifier is primarily determined by the choice of K as well as the distance metric applied [49]. However, it has been shown that when the points are not uniformly distributed, predetermining the value of K becomes difficult [52]. Larger k values help reduce the effects of noisy points within the training data set, and the choice of k is often performed through cross validation [46].

In parametric methods, whether for density estimation or classification, we assume a model valid for over the whole input space [46]. In classification, when we assume a normal density, we assume that all examples of the class are drawn from this same density [46]. But practically this assumption does not always hold and we may incur a large error if it does not [46]. If we cannot make such assumptions and cannot come up with a parametric model, one possibility is to use semi parametric mixture model. In nonparametric estimation, all we assume is that similar in- puts have similar outputs. Therefore, our algorithm is composed of finding the similar

past instances from the training set using a suitable distance measure and interpolating from them to find the right output [50, 51].

4. Block diagram of the proposed system

Figure 1 shows the block diagram of the proposed system for training and test steps. The algorithm starts by applying required preprocessing to input image. In preprocessing step the input image is first converted to binary image (binarization), using algorithms [53] its slant is corrected and finally the handwritten digit normalize it to a constant size of 64×64. Feature extraction is the next step of the proposed system. We used multi different methods for feature extraction including variance, covariance, central moments, median and the number of pixels each square. The extracted features are then used for training and testing the classifier algorithm. We utilized SVM and K-NN for the classification and recognition of handwritten Farsi digits.
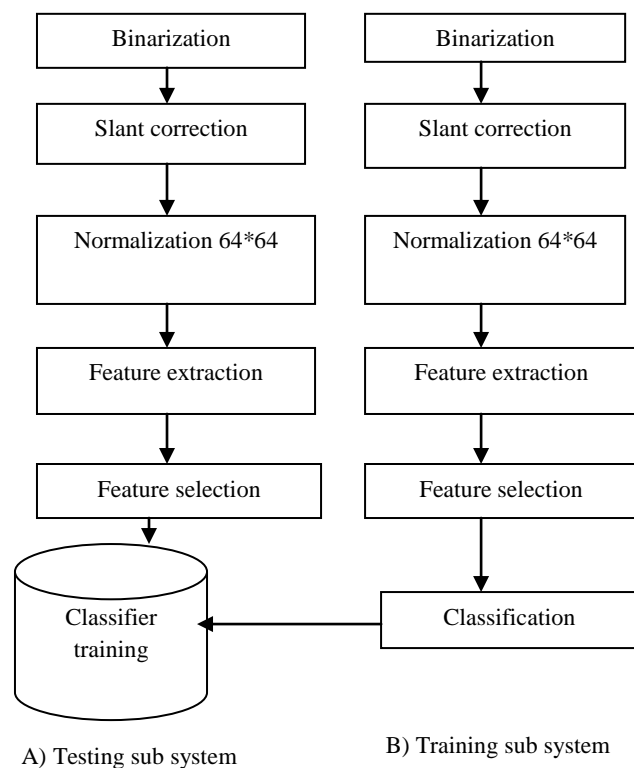

A) Testing sub system   B) Training sub system

Fig1. Farsi digit recognition system diagram

4.1 Data set

The proposed system for digit recognition implemented using a MATLAB program and evaluated by Farsi handwritten database Tabiyat Modares University (TMU). The database includes1699 samples for ten Persian handwritten numeral and number samples for each digit differenced. The table1 show number samples digits of each class. Figure 2 show samples of Farsi handwritten digits in the database. We applied 1121 samples of each digit for training and other 578 samples for test. Table 2 shows number of samples per each class in this data set.
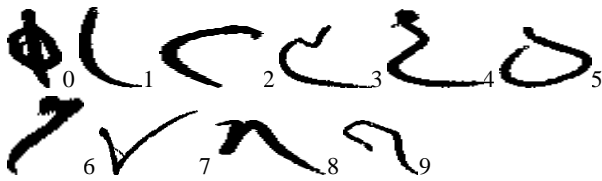
Fig 2. Samples with different writing styles in the dataset

Table2. Number of samples per each class

Digit 9	Digit 8	Digit 7	Digit 6	Digit 5	Digit 4	Digit 3	Digit 2	Digit 1	Digit 0
159	169	172	174	172	173	171	163	170	176

5. Performance of the proposed system

This section presents the results of different experiments. Our experiments were executed on TMU (Tarbiyat Modares University) data set of handwritten digits classification. The total number of 1699 digit samples is divided into two parts, 1121(66%) samples for the training step and 578(34%) samples for the test step. All datasets were normalized to the range [-1, 1]. All experiments were implemented in MATLAB software on a laptop with Intel dual core processor (2 GHz), 2GB of main memory, running windows seven. In K- Nearest Neighbor, the value of k is given like k=1 through 10. The distance metric is also given which is called Euclidean distance. Last of all for each k, the (%) or error is calculated that is shown in table5. The results obtained in the table5 showed that for this particular problem at K=4, we achieved promising accuracy rates and the minimum error which is 7.9585(%).it means that the highest accuracy value is achieved when 4 nearest neighbors contribute to the voting process. Tables 3, 4 show in order to the results achieved by the considered multiclass SVM and SMO classifications methods with the various kernels and classification accuracy on the test dataset. These experiments show that classifier SVMs with linear kernel is provided higher recognition rates than linear, polynomial or sigmoid kernel and classifier SMO with Normalized poly kernel is provided higher recognition rates than poly kernel, puk kernel and RBF kernel. The SVM classifier with liner kernel and SMO classifier with Normalized poly kernel that yielded the best result were used in the final tests.

Table 3. (%) of error for different kernels using Sequential Minimal Optimization (SMO)

Kernel	Correct (%)	Error (%)
Normalized poly kernel	**92.3875**	**7.6125**
Poly kernel	90.8304	9.1696
puk	85.2941	14.7059
RBF Kernel	89.7924	10.2076

Table 4. (%) of error for different kernels using support vector machine.

Kernel	Correct (%)	Error (%)
Linear	**91.3495**	**8.6505**
Polynomial Degree=2	25.6055	74.3945
Polynomial Degree=3	17.301	82.699
Polynomial Degree=4	8.8235	91.1765
Gaussian $\sigma = 0$	87.0242	12.9758
Gaussian $\sigma = 0.002$	85.8131	14.1869
Gaussian $\sigma = 0.005$	87.5433	12.4567
Sigmoid	82.699	17.301

Table 5. (%) of error for different K using K- Nearest Neighbor

K	Correct (%)	Error (%)
1	91.3495	8.6505
2	91.3495	8.6505
3	91.1765	8.8235
4	**92.0415**	**7.9585**
5	91.0035	8.9965
6	91.3495	8.6505
7	89.7924	10.2076
8	90.4844	9.5156
9	89.7924	10.2076
10	89.4464	10.5536

Table 6. Confusion matrix of

A	B	C	D	E	F	G	H	I	J	Classifier
49	0	0	0	0	2	0	0	0	0	A=C0
0	61	0	0	0	0	0	0	0	0	B=C1
0	2	52	3	2	0	0	0	0	0	C=C2
0	0	1	41	6	1	0	0	0	0	D=C3
0	0	1	3	55	0	0	0	0	0	E=C4
3	0	0	0	0	51	0	0	0	0	F=C5
1	0	2	0	1	1	48	0	1	4	G=C6
0	0	2	0	0	0	1	62	1	0	H=C7
0	0	0	0	0	1	0	0	56	0	I=C8
0	1	1	0	1	0	2	0	0	59	J=C9

Thus the recognition accuracy of each digit would be according to the Table 7.

Table 7.Accuracy recognition of Persian digits

9	8	7	6	5	4	3	2	1	0	digit
92.2	98.2	93.9	82.8	94.4	93.2	83.7	88.1	100	96.1	Accuracy (%)

5.1 Comparison of results

To compare the performance of the novel method, we consider most of the works that are available for Persian handwritten digits recognition. It may be noted from Table 8 that all the existing works were evaluated on smaller datasets. The highest dataset of size 80,000 was used by a recent work due to Rashnodi et al. [68, 69], Where as we used 80,000 sample for our experiment. The Highest accuracy was obtained from the work due to Soltanzadeh et al. [61] but they have experimented with Only 8,918 samples and used 257 dimensional features vector. We considered 1699 samples for the system proposed and has been acquired 92.3875% accuracies using140 dimensional Features vector.

6. Conclusion

The basic purpose of this paper is to review the three traditional machine learning techniques: a support vector machine (SVM), K- Nearest Neighbor (K-NN) and Sequential Minimal Optimization (SMO) Classifier and also investigate the performance of these three used classifiers. To do these tasks; steps are followed to implement the machine learning algorithms and application of these techniques to a dataset TMU. All the tables classification that were needed to predict the classification was shown in section 5. The process of calculating these tables basically reflects the details process in case of SVM, SMO and K-Nearest Neighbor classifiers. According to these tables, The SMO classifier with Normalized poly kernel has higher performance other than those mentioned classifiers (SVM, K-NN).In K-Nearest Neighbor method, the selection of k values application dependent. In the best case The SMO is used to the testing data set and found 92.3875 (%) correct recognition rate, SVM is found 91.3495 (%) correct recognition rate and K-NN is found 92.0415 (%) correct recognition rate.

Table8. Comparison of different algorithms

Algorithms	Data size		Accuracy	
	Train	Test	Train	Test
Shirali-shahrezaetal. [31]	2600	1300	--	97.80
Soltanzadeh,Rahmai [62]	4979	3939	--	99.57
Dehghan, Faez [64]	6000	4000	--	97.01
Harifi.,Aghagolzadh [59]	230	500	--	97.60
Ziaratban et al. [65]	6000	4000	100	97.65
Mowlaei, Faez [67]	2240	1600	100	92.44
Hosseini, Bouzerdm [60]	480	480	--	92.00
Mowlaei et al. [68]	2240	1600	99.29	91.88
Mozaffari et al. [61]	2240	1600	98.00	91.37
Mozaffari et al. [66]	2240	1600	100	94.44
Sadri et al. [63]	7390	3035	---	94.14
Omid rashnodi et al[69]	60000	20000	100	98.84
Omid rashnoodi et al[70]	60000	20000	99.6	98.945
Proposed scheme	**1121**	**578**	**---**	**92.3875**

7. REFERENCES

[1] C. L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, 2003. Handwritten digit recognition: benchmarking of state-of-the-art techniques, Pattern Recognition 36, 2271 – 2285.

[2] O.D. Trier, A.K. Jain, T. Taxt, 1996. Feature extraction methods for character recognition—a survey, Pattern Recognition Vol.29, No. 4,641- 662.

[3] Ho TK, Hull JJ, Srihari SN, 1994. Decision combination in multiple classifier systems. IEEE Trans Pattern Anal Mach Intell Vol.16, No.1, 66 – 75.

[4] Xu L, Krzyzak A, Suen CY, 1991. Associative switch for combining multiple classifiers. In: Int. Joint Conf. on Neural Networks, Vol.1. pp 43– 48.

[5] Suen CY, Nadal C, Mai TA, Legault R, Lam L, 1990. Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts. In: Proc.IWFHR, 131-143.

[6] Cheng-Lin Liu, Ching Y. Suen, 2008. A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters, Pattern Recognition.

[7] Hassan Soltanzadeh, Mohammad Rahmati, 2004. Recognition of Persian handwritten digits using image profiles of multiple orientations, Pattern Recognition Letters 25, 1569 –1576.

[8] A. Amin, 1998. Off-line Arabic character recognition: the state of the art, Pattern Recognition 31, 517–530.

[9] C.Y. Suen, S. Izadinia, J. Sadri, F. Solimanpour, 2006. Farsi script recognition: a survey, in: Proceedings of the Summit on Arabic and Chinese Handwriting Recognition, University of Maryland, College Park, MD, 101–110.

[10] Fevzi Alimoglu, Ethem Alpaydin, 2001. "Combining Multiple Representations for Pen-based Handwritten Digit Recognition," Turk J.Elec.Eng., Vol.9, No.1.

[11] L. Xu, A. Krzyzak, C.Y. Suen, 1992. Methods of combining multiple classifiers and their application to handwriting recognition, IEEE Trans. Systems Man Cybernet. Vol. 22, 418-435.

[12] A. F. R. Rahman, M. C. Fairhurst, 2003. Multiple classifier decision combination strategies for character recognition: A review, International Journal on Document Analysis and Recognition, 166–194.

[13] A. Goltsev, D.Rachkovskij, 2005. Combination of the assembly neural network with a preceptor for recognition of handwritten digits arranged in numeral strings, Pattern Recognition 38, 315 – 322.

[14] S. Haykin, 1998. Neural Networks-A Comprehensive Foundation, second ed., Prentice-Hall.

[15] K. Woods, W.P. Kegelmeyer, K. Bowyer, 1997. Combination of multiple classifiers using local accuracy estimates, IEEE Trans. Pattern Anal. Mach. Intel, Vol. 19, 405-410.

[16] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, G.E. Hinton, 1991. Adaptive mixtures of local experts, Neural Comput. Vol.3, 79-87.

[17] L.A. Rastrigin, R.H. Erenstein, 1982.Method of Collective Recognition, Energoizdat, Moscow.

[18] E. Alpaydin, M.I. Jordan, 1996. Local linear perceptrons for classification, IEEE Trans. Neural Networks ,Vol. 7, No. 3 , 788-792 .

[19] L.I. Kuncheva, 2004. "Combining Pattern Classifiers: Methods and algorithms," published by John Wiley & Sons. Inc.

[20] L. Rokach, 2010. "Ensemble-based classifiers," Artif Intelligent Rev, Vol. 33, 1–39.

[21] H. Aljuaid, Z. Muhammad and et al, 2010. "A Tool to Develop Arabic Handwriting Recognition System Using Genetic Approach," Journal of Computer Science No. 6, Vol. 6, 619-624.

[22] N. Ahmad, T. Natarjan, and K,Roa, 1974. "Discrete Cosine Transform," IEEE. Trans Compute., Vol. C-23, 90-93.

[23] H. Choi, S. J. Cho, and J. H. Kim, 2003. "Generation of Handwritten Characters with Bayesian network based On-line Handwriting Recognizers," Proceedings of the Seventh International Conference on Document Analysis and Recognition.

[24] F.ch Li and F. Guan, 2009. 'Heuristic Model Research on Decision Tree Algorithm'. Intelligent Interaction and Affective Computing, 2009. ASIA '09. International Asia Symposium on 149 – 152.

[25] DS. Lee and SN. Srihari, 1995. "A theory of classifier combination: the neural network approach." Proceedings of the third international conference on document analysis and recognition, Montreal, Canada, 42–5.

[26] G. Giacinto, F. Roli, and L, Bruzzone, 2000. "Combination of neural and statistical algorithms for supervised classification of remote-sensing images," Pattern Recognition Letters, Vol. 21, 385-397.

[27] CJC. Burges, 1998. "A tutorial on support vector machines for pattern recognition," Know Disc Data Min, Vol. 2, 1– 43.

[28] G. M. FUNG, 2005. "Multi category Proximal Support Vector Machine Classifiers," Machine Learning, Vol. 59, 77–97.

[29] M.H.Shirali-shareza, K.Faez, A.Khoanzad, 1994. "recognition of handwritten Farsi numerals by Zernike moments features and a set of class specific neural network classifiers", proceeding of the international conference of signal processing applications, and technology, Vol.2, 998-1003.

[30] M.H.Shirali-shareza, K.Faez, A.Khoanzad, 1995. "Recognition of handwritten Persian/Arabic numerals by shadow coding and an edited probabilistic neural network", proccedings of international conference of image processing, vol.3, Washington D.C.

[31] J.C.Platt, 1999.Fast training of support vector machines using Sequential minimal optimization, in: B. Scholkopf, C.J.C.Burges, A.J.Smola (Eds.), Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 185–208.

[32] O.L. Mangasarian, D.R.Musicant, 1999. Successive over relaxation for support vector machines, IEEE Trans.Neural Networks, Vol.10, No.5, 1032–1037.

[33] C.L. Liu, K. Nakashima, H. Sako, H. Fujisawa, 2004. Handwritten digit recognition: investigation of normalization and feature extraction techniques, Pattern Recognition Vol.37, No.2, 265 – 279.

[34] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai, L. Lam, 1992.Computer recognition of unconstrained handwritten numerals, Proc. IEEE,Vol. 80 ,No.7 ,1162 – 1180.

[35] A. Broumandnia and M. Fathi, 2005. "Application of pattern recognition for Farsi license plate recognition," ICGST- GVIP Journal, Vol. 5.

[36] Smith, Lindsay I, Feb 2002. "A Tutorial on Principal Components Analysis".

[37] Ronald Fagin, 1998. "Fuzzy Queries in Multimedia Database Systems", ACM PODS.

[38] S. Abe, 2005. Support vector machines for pattern recognition, Springer Verlog London limited.

[39] Turk, M. and Pent land, A.: Eigen faces for Recognition. J. Cognitive Neurosis. Vol.3, No. 1, 71-86, 1991.

[40] Martinez A, Kak A, PCA versus LDA. , 2001. IEEE Trans Pattern Anal Mach Intell Vol.23, No.2, 228-233.

[41] Athitsos, V., Alon, J., Sclaroff, and S. 2005. Efficient nearest neighbor classification using a cascade of approximate similarity measures. In: CVPR '05. IEEE Computer Society, Washington, DC, USA, 486–493.

[42] Athitsos, V., Sclaroff, S., 2005. Boosting nearest neighbor classifiers for multiclass recognition. In: CVPR '05, IEEE Computer Society, Washington, DC, USA.

[43] Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory Vol.13, No.1, 21–27.

[44] Peng, J., Heisterkamp, D.R., Dai, H.K.: LDA/SVM driven nearest neighbor classification. In: CVPR '01, p. 58. IEEE Computer Society, Los Alamitos, CA, USA (2001)

[45] Zhang, H., Berg, A.C., Maire, M., Svm-knn, J.M.: Discriminative nearest neighbor classification for visual category recognition. In: CVPR '06, pp. 2126–2136. IEEE Computer Society, Los Alamitos, CA, USA (2006)

[46] Mohammed J. Islam, Q. M. Jonathan Wu, Majid Ahmadi, Maher A. Sid-Ahmed, 2007 International Conference on Convergence Information Technology," Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers"

[47] N. Roussopoulos, S. Kelley, and F. Vincent. Nearest neighbor queries. Proceedings of the 1995 AC M SIGMOD Inter- national Conference on Management of Data, 1995.

[48] T. Cover and P. Hart. , 1967. Nearest neighbor pattern classification. IEEE Transaction on Information Theory, 21–27.

[49] Latourrette, M., 2000. Toward an explanatory similarity measure for nearest-neighbor classification. In: ECML '00: Proceedings of the 11th European Conference on Machine Learning, London, UK, 238–245.

[50] R. Duda, P. Hart, and D. Stork. Pattern Classification. Wiley Inter science, 2nd Ed.

[51] Q. J. Wu., 2007. Class Notes- Machine Learning and Computer Vision. University of Windsor, Windsor, ON, Canada.

[52] Domeniconi, C., Peng, J., Gunopulos, 2002. D.: Locally adaptive metric nearest-neighbor classification. IEEE Trans. Pattern Anal. Mach. Intell.,Vol. 24,No.9, 1281–1285

[53] M. Hanmandlu, O.V. Ramana Murthy, 2006. Fuzzy model based recognition of handwritten numerals, the journal of pattern recognition society, 1840 – 1854.

[54] Vapnik, 1998.statistical learning theory. Wiley, New York.

[55] Knerr S, Personnaz L, Dreyfus G, 1990. Single-layer learning revisited a stepwise procedure f or building and training a neural network. In: Fogelman J (Ed) Nero computing: algorithms, architectures, and applications. Springer- Verlag, New York.

[56] Hastie T, Tibshirani R, 1998. Classification by pair wise coupling. Ann Stat ,Vol.26,No.2,451–471

[57] Platt JC, C hristiani N, Shawe–Taylor, 1999. Large margin Dag's f or multiclass classification. In: Solla SA, Leen TK, Muller KR (eds) Proceedings of the neural information processing systems (NIPS'99). MIT Press, 547–553

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

[59] H. Mir Mohammad Hosseini and A. Bouzerdoum, 1996."A Combined Method for Persian and Arabic Handwritten Digit Recognition", Australian New Zealand Conference on Intelligent Information System, 80 – 83.

[60] S. Mozaffari, K. Faez & H. Rashidy Kanan, 2004. "Recognition of Isolated Handwritten Farsi/Arabic Alphanumeric Using Fractal Codes", Image Analysis and Interpretation, 6th Southwest Symposium, 104-108.

[61] H. Soltanzadeh and M. Rahmati, 2004. "Recognition of Persian handwritten digits using image profiles of multiple orientations", Pattern Recognition Letters, Vol. 25, 1569–1576.

[62] J. Sadri, C.Y. Suen and T.D. Bui, 2003. "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, 300-307.

[63] M. Dehghan and K. Faez, 1997. "Farsi Handwritten Character Recognition With Moment Invariants", Proceedings of 13th International Conference on Digital Signal Processing, Vol. 2, 507-510.

[64] M. Ziaratban, K. Faez and F. Faradji, 2007. "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, 297-301.

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

[66] A. Mowlaei and K. Faez, 2003. "Recognition Of Isolated Handwritten Persian /Arabic Characters and Numerals Using Support Vector Machines", Proceedings of XIII Workshop on Neural Networks for Signal Processing, 547-554.

[67] A. Mowlaei, K. Faez, A. Highlight, 2002. "Feature Extraction with Wavelet Transform for Recognition of Isolated Handwritten Farsi/Arabic Characters and Numerals", Digital Signal Processing, Vol. 2, 923- 926.

[68] O. rashnodi, H. sajedi,M. saniee, 2011." Persian Handwritten Digit Recognition Using Support Vector Machines",international journal of computer applications,Vol. 29, No.12.

[69] O. rashnodi, H. sajedi, M. Saniee, 2011." Using Box Approach in Persian Handwritten Digits Recognation ", international journal of computer applications, Vol. 32, No. 3.