Persian Handwritten Digit Recognition using Support Vector Machines

Omid Rashnodi
Computer Engineering Department

Science and Research Branch, Islamic Azad University Ahvaz, Iran

Hedieh Sajedi
Computer Engineering Department

Amirkabir university of Technology, Tehran, Iran

Mohammad Saniee Abadeh
Electrical and Computer Engineering Faculty

Tarbiat Modares University

Tehran, Iran

ABSTRACT
In this paper, appropriate features set based on Discrete Fourier Transform coefficients and the box approach have been proposed to achieve higher recognition accuracy, decreasing the features set dimensions and recognition time of Persian numerals. In classification phase, support vector machine (SVM) has been employed as the classifier. Feature sets consists of 154 dimensions, which are the Fourier coefficients in the contour pixels of input image, average angle and distance pixels which are equal to one in each box the box approach. The scheme has been evaluated on 80,000 handwritten samples of Persian numerals. Using 60,000 samples for training, scheme was tested on other 20,000 samples and 98.84% correct recognition rate was obtained.

General Terms

Classification, pattern recognition, handwritten digit, optical character recognition, performance

Keywords
Box approach, Discrete Fourier Transform coefficients, Support vector machines, Gaussian kernel

1. INTRODUCTION
Nowadays, recognition systems are used in many fields that have different nature. The optical character recognition (OCR) was started from the recognition of machine printed digits and characters and then it was developed to the recognition of machine printed words. Gradually, handwritten digit, character and word recognition were introduced into this domain. Most researches have been done in Latin languages. Typically an ordinary OCR includes three indicated phase. Pre-processor, feature extraction and classifications phase that output of each step is the input of next step. In the phase Pre-processing operations such as slant correction, normalization and thinning have been done, in phase feature extraction of Discrete Fourier Transform coefficients and the box approach is used. The last phase (classification) of SVM (Support Vector Machine) as classifier is used. The flowchart of a typical OCR can be shown as figure1.
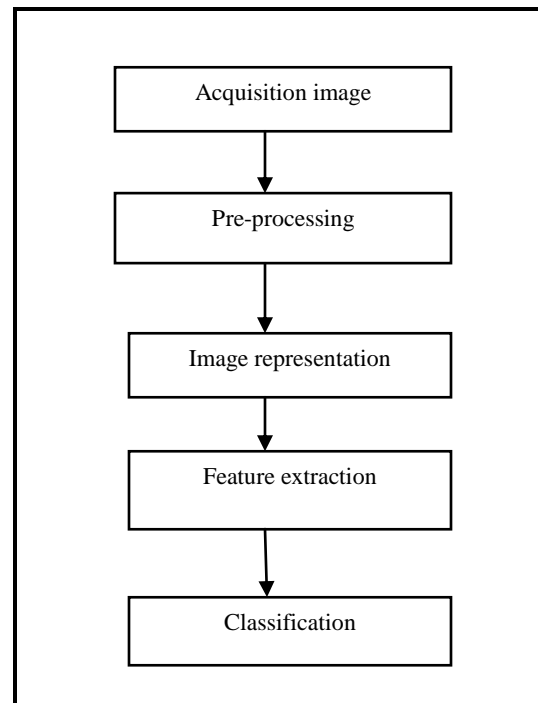

Fig1: Recognition system of handwriting digits

Recognition of handwritten characters is one of the most interesting topics in pattern recognition domain. In OCR applications, handwritten character recognition, especially digit recognition, is dealt with in postal mail sorting, bank check processing, form data entry, etc. For these applications, the performance (accuracy and speed) of digit recognition is crucial to the overall performance. While in pattern classification and machine learning communities, the problem of handwritten digit recognition is a good example to test the classification performance. Due to increasing Persian/Arabic writing usage in many day-to-day businesses in Persian countries, it has been become necessary for machines to understand handwritten materials in Persian. As a part of Persian scripts, numeral strings and isolated numerals play an enormous role. OCR for handwritten documents in some languages (English, Chinese, Japanese, etc.) has reached to a promising level [1]. The OCR for Persian has not grown up like abovementioned languages

because of the cursive-ness of handwritten in Persian and multiple forms of each character with respect to its position in words. Achieving this goal, the effect of discrete Fourier transform coefficients have been studied on the contour pixels. Also framing algorithm has been used to study of average angle and distance pixels as features which are equal to one in each box. Both of used methods keep the information of input images. In this paper, SVM with Gaussian kernel has been used as the classifier. In the literature survey particularly relevant to the Persian/Arabic languages, there are many methods for feature extraction and classification. As feature extraction methods segmentation and shadow code [2, 3, 4], fractal code [5], profiles [6, 7], moment [8], template [9], structural feature (points, primitives) [10] and wavelet [11, 12] have been used. For classification, different types of Neural Networks [2, 8, 3, 9, 4, 12, 5], SVM's [6, 11, 7] and Nearest Neighbor [10] have been applied. Investigating of previous researches on Persian/Arabic numerals recognition, it seems that more appropriate and effective feature set could have been developed to react recognition phase. To overcome this problem, supplying a more effective feature set has been proposed based on the discrete Fourier Transform coefficients in the contour pixels of input image and average angle and distance pixels which are equal to one in each box frame technique. Then SVM for classification can be used. This type of feature set expresses the physical shape of input image and extracts its local information and provides more suitable accuracy in experimental part. It is worth mentioning that in the proposed system we apply some pre-processing techniques such as slant detection/ thinning, normalization, binarization and etc. The organization of rest of the paper is as follows: In Section 2, the proposed feature extraction technique has been illustrated; Section 3 describes the classification phase. Experimental results in section 4 and comparative analysis in section 5 are described and finally in last section we present conclusion.

2. THE PROPOSED FEATURE EXTRACTION TECHNIQUE

To extract features in proposed method, which is the most effective part of OCR's systems, the following tasks have been performed:

1- Applying the normalization pre-processes, eliminating the gaps which surround image, binarization and obtaining the skeleton image.

2- For extracting features, we use the Box- approach in Refs. [13, 14]. In this approach for extracting features from the slant corrected, size normalized and thinned binary image of a character is considered. The box approach based on spatial division of character image is proposed for feature extraction. In this, the character image of

The size of 45× 45 is fitted into horizontal and vertical grid lines of 8× 8. Thus, in this case 64 boxes of size 8× 8 are devised and then these boxes are superimposed on the image, so that some of the boxes will have a portion of the image and others empty. However, all boxes are considered for analysis. By taking the bottom left corner as the absolute origin (0, 0), the vector distance fork the pixel in

b^{th} box at location (i, j) is calculated as, $d_k^{\,b} = (i^2 + j^2)^{\frac{1}{2}}$. By dividing the sum of distances of all 'l's pixels present in a box with their total number, a normalized vector distance (γ_b) for each box is computed as:

$$\gamma_b = \frac{1}{n_b} \sum_{k=1}^{n_b} d_k^{\,b}, b = 1,2,3,\ldots,64 \qquad (1)$$

Where n_b is number of pixels in bth box. The above vector distances constitute a set of features based on distance. Similarly, for each kth 'l' pixel in a box, the corresponding angle is computed as, $\theta_k = \arctan(-\frac{j}{i})$ for a pixel at (I, J).

Then the sum of all angles in a box b is normalized with the number of 'l' pixels present in that box to yield a normalized angle α_b:

$$\alpha_b = \frac{1}{n_b} \sum_{k=1}^{n_b} \theta_k^{\,b} \qquad b = 1,2,3,\ldots,64 \qquad (2)$$

These 64 pairs constitute the complete feature set of a character, which are used for recognition. The points resulting from these 64 feature pairs have been back substituted in their respective boxes store generate the pattern of original character. All the pairs of features corresponding to each box are arranged in a sequential order starting from box-1 to box-64, as shown in Table1 for further analysis:

Table1. The order of extraction of parameters from boxes

Number per box	Two features per box
Box-1	λ1 , α1
Box-2	λ2 , α2
…	…
…	…
Box-64	λ64 , α64

3- Calculating a ratio of length to width of each preprocessed image and considering it as geometrical features for image. The following formula has been utilized to achieve this ratio:

$$ratio = \frac{height}{width} = \frac{length}{width} \qquad (3)$$

4- Calculating the first 25 Fourier coefficient of each image. Fourier coefficients are used to describe two-dimensional closed shape. A closed shape is shown in two-dimensional coordinates. So to show the contour shape to k points uniformly in two-

dimensional plane, the shape is scrolling counter clockwise and its points to be obtained. Thus every digit is considered as a shape and its surrounded area's points are achieved. In this case the X axis is introduced as real axis and y axis is indicated as the imaginary axis. Sequences of discrete Fourier coefficients are obtained as follows:

$$S(m) = \frac{1}{N} \sum_{n=0}^{N-1} S(n)e^{\frac{-j2nm\pi}{N}} \qquad (4)$$

When complex coefficients s (n) is series term which are used to display the shape contour, N is the size of the input vector. S (m) is a complex number. One of the specifications of Fourier coefficients is that its low Frequencies can determine the overall shape of the object and its High Frequencies can distinguish the details shape clearly. Considering properties of Fourier coefficient, they can be used as some of shape features.

For obtaining the Fourier coefficients it is necessary to obtain the Shape contour and the rest of the shape part should be eliminated. Figure 2 shows an image with its contour.

Fig2: An image and its contour

According to above steps, 154 features are extracted for each image. In the other words, for each image a feature vector 154 dimensions is considered.

3. CLASSIFICATION

Support vector machines (SVMs) are particular classifiers which are based on the margin-maximization principle. They perform structural risk minimization, which was introduced to machine learning by Vapnik, and have produced excellent generalization performance [15, 16]. For nonlinear problems, SVMs use the kernel trick to produce nonlinear boundaries. The idea behind kernels is to map training data nonlinearly into a higher-dimensional feature space via a mapping function and to construct a separating hyper plane which maximizes the margin. The construction of the linear decision surface in this feature space only requires the evaluation of dot products $\Phi(x).\Phi(y) = k(x, y)$, where $k(x, y)$ is called the kernel function [17, 18, 19, 20].

The discriminate function of a binary SVM is computed by [21]:

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i k(x, x_i) + b \qquad (5)$$

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, x_i) is a kernel function which implicitly defines and expanded feature space:

$$k(x, x_i) = \phi(x).\phi(x_i) \qquad (6)$$

Where $\phi(x)$ the feature is vector in the expanded feature space and may have Infinite dimensionality [7]. Two types of kernels, polynomial kernel and RBF kernel are frequently used .They are computed by:

Polynomial kernel:

$$k(x, x_i, p) = (1 + xx_i)^p \qquad (7)$$

RBF kernel:

$$k(x, x_i, \sigma^2) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$$

Where p, σ are the parameters of the corresponding kernels.

The coefficients $\alpha_i (i = 1, 2, ..., l)$ in equation.(5) are determined by solving the following optimization problem:

Minimize $\quad \tau(w) = \frac{1}{2} \|w\|^2 \qquad (8)$

Subject to $\quad y_i . f(x_i) \geq 1 - \xi_i \quad , \xi_i \geq 0, i = 1, 2, 3, ..., l$

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

Minimize

$$w(x) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \qquad (9)$$

Subject to $0 \leq \alpha_i \leq c, i = 1, 2, ..., l \text{ and } \sum_{i=1}^{l} \alpha_i y_i = 0$

Where C is a parameter to control the tolerance of classification errors in learning. Various methods have been proposed to solve the optimization problem (9), especially when there is a large number of training samples, among them, we used the sequential minimal optimization method (SMO) proposed by Platt [22]. SVM was defined for two-class problem and it looked for the optimal hyper plane, which maximized the distance, the margin, between the nearest examples of both classes, named support vectors (SVs) [23]. Different techniques can be used to extend the binary support vector classifier for the multi-class problems. In this paper is used the One-rest method. In this case, N different binary classifiers are created, where N is the number of the classes. Each of these classifiers is trained to separate one of

the classes from the others. An input pattern is assigned to the class of the binary classifier, which gives the maxi mum output value for that pattern. The linear SVM can be extended to a non-linear classifier by using kernel functions like polynomial, sigmoid and Gaussian kernels. We have tested linear, Gaussian, sigmoid and polynomial kernels during our experiments and we received the best result using Gaussian kernel thus we employed SVMs with Gaussian kernel as classifier. Details of SVM can be found elsewhere [23, 24].The input feature sets were the 154-dimension. All the SVMs trained with the respective training feature sets and the results explored by using separate test data. We obtained the best results with the Gaussian kernel of gamma=0.0002.

4. EXPERIMENTAL AND COMPARATIVE RESULTS

In this paper, for experimental analysis, we considered a standard Persian numeral dataset [25] with 80000 samples.

4.1 Data set

For experimental analysis, we considered 60,000 samples for training and 20,000 samples for testing as mentioned in [25].

These samples were extracted from different registration forms of entrance examinations of universities in Iran containing Iranian Postal and National Codes. The images were scanned at 200 dpi resolution [25].

4.2 Performance of the proposed system

Using 60,000 samples for training, we tested our scheme on other 20,000 samples and obtained 98.84% accuracy. From the experiment, we got an accuracy of 99.4% when the 60,000 data were used as training and the same data set was used for testing. In another experiment, we used 5-fold cross validation scheme for recognition result calculation. We divided our database (80,000 samples) into 5 subsets and testing is done on each subset using rest of the 4 subsets for training. The recognition rates for all the five test subsets of dataset are averaged to get the accuracy. We got the average accuracy of 99.4%. Further, we considered some noisy images in our test data. The result showed the effectiveness of the proposed feature extraction technique (Table 2).

Table2. Confusion matrix of the result

A	B	C	D	E	F	G	H	I	J	◄ classified as
1958	7	0	0	0	32	0	3	0	0	A=C_0
2	1994	0	0	0	0	0	4	0	0	B=C_1
0	1	1981	6	4	0	5	1	0	2	C=C_2
0	2	42	1927	26	3	0	0	0	0	D=C_3
0	2	6	13	1979	0	0	0	0	0	E=C_4
7	0	0	0	3	1987	0	0	2	1	F=C_5
0	5	3	4	2	0	1972	0	0	14	G=C_6
0	6	1	0	0	0	0	1993	0	0	H=C_7
0	2	0	0	0	0	0	0	1997	1	I=C_8
3	6	5	0	0	0	6	0	0	1980	J=C_9

Thus the recognition accuracy of each digit would be according to the Table 3.

Table3. Accuracy recognition of Persian digits

9	8	7	6	5	4	3	2	1	0	Digits
99%	99.85%	99.65%	98.6%	99.35%	98.95%	96.35%	99.05%	99.7%	97.9%	**Accuracy**

4.2 Confusion pairs

In our experiment (with the 98.84% accuracy), we observed confusion numerals in the recognition phase between some digits. In Table I, we showed Detail of confusing results. The major confusions were amongst 2, 3 and 4. This happened because 2, 3 and 4 look likes each other. From the Table1 it may be noted that out of 2000 samples of number three (3), 42(2.1%) samples misrecognized to numeral 2 and 26(1.3%) samples Misrecognized to numeral 4. In some of the samples, little confusions were also between 0 and 1.

5. COMPARISON OF RESULTS

To compare the performance of our method, we consider most of the works that are available for Persian numeral recognition. It may be noted from Table 4 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. [9], Where as we used 80,000 data for our experiment. The Highest accuracy was obtained from the work due to Soltanzadeh et al. [6] 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 98.84% and 99.4% accuracies using only154 dimensional features.

Table4. Comparison of different algorithms

Algorithms	Data size		Accuracy	
	Train	Test	Train	Test
Shirali-shahrezaetal. [2]	2600	1300	--	97.80
Soltanzadeh,Rahmai [6]	4979	3939	--	99.57
Dehghan, Faez [8]	6000	4000	--	97.01
Harifi.,Aghagolzadh [3]	230	500	--	97.60
Ziaratban et al. [9]	6000	4000	100	97.65
Mowlaei, Faez [11]	2240	1600	100	92.44
Hosseini, Bouzerdm [4]	480	480	--	92.00
Mowlaei et al. [12]	2240	1600	99.29	91.88
Mozaffari et al. [5]	2240	1600	98.00	91.37
Mozaffari et al. [10]	2240	1600	100	94.44
Sadri et al. [7]	7390	3035	---	94.14
ProposedAlgoritm	60000	20000	100	98.84
ProposedAlgo.(5 fold)	**80000**		--	**99.4**

6. CONCLUSION

In this paper, an efficient feature extracting technique is proposed. From experimental results, it is evident that our features resulted worthy performances about (98.84%, 99.4%). We noted that most of misclassified samples were from classes of 2, 3 and 4, which have similar shapes. The recognition of such similar numerals is difficult even by human being. It is obvious that by removing confusion among few classes, we can achieve better performance. Further, to the best of our knowledge, this work is the first work, towards the recognition of Persian handwritten numerals on a huge dataset. To achieve better results (less time for testing, feature set with smaller dimensions and more accuracy recognition) combination methods which are extracting features and classifiers can be applied.

7. REFERENCES

[1] S.N. Sridhar and G. Ball, 2007. "An Assessment of Arabic Handwriting Recognition Technology", CEDAR Technical Report TR-03-07.

[2] M.H. Shirali-Shahreza, K. Faze and A. Khotanzad, 1995. "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, 436 – 439.

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

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] M. Hanmandlu, M.H.M. Yusof, M. Vamsi Krishna, 2005. "Off-line signature verification and forgery detection using fuzzy modeling", Pattern Recognition, vol. 38, No. 3, 341–356.

[14] M. Hanmandlu, K. R. Murali Mohan, S. Chakra borty, S. Goyal, D. Roy Choudhury, 2003. "Unconstrained handwritten character recognition based on fuzzy logic", Pattern Recognition, vol. 36, No. 3, 603– 623.

[15] V.N. Vapnik, 1992. "Principles of risk minimization for learning theory, Advances in Neural Information Processing Systems", vol. 4, Morgan Kaufman, San Mateo, CA, 831– 838.

[16] V.N. Vapnik, 1998. "Statistical Learning Theory", Wiley, New York.

[17] B.E. Boser, I. Guyon, V. Vapnik, 1992." A training algorithm for optimal margin classifiers", Comput. Learn. Theory, 144 –152.

[18] N. Cristianini, J. Shawe-Taylor, 2000. "An Introduction to Support Vector Machines", Cambridge University Press, Cambridge.

[19] B. Scholkopf, A.J. Smola," Learning with Kernels", 2002. MIT Press, Cambridge, MA.

[20] J. Shawe-Taylor, N. Cristianini, 2004. "Kernel Methods for Pattern Analysis", Cambridge University Press, Cambridge.

[21] Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Know. Disc. 2, 121–167.

[22] Platt, J.C., 1999. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf, B., Burges, C., Smola, A.J. (Eds.), Advances in Kernel Methods—Support Vector learning. MIT Press, Cambridge, Massa- chusetts, Chapter 12, 185–208.

[23] C. Burges, 1998. "A Tutorial on support Vector machines for pattern recognition", Data Mining & Knowledge Discovery, vol. 2, 1- 43.

[24] V. N. Vapnik, 1995. "The Nature of Statistical Learning Theory", Springer Verlang.

[25] H. Khosravi and E. Kabir, 2007."Introducing a very large dataset of handwritten Farsi digits and a study on the variety of handwriting styles", Pattern Recognition Letters, vol. 28, Issue 10, 1133-1141.