Hand Written Digit Recognition by Multiple Classifier Fusion based on Decision Templates Approach

Reza Ebrahimpour, and Samaneh Hamedi

Abstract**—**Classifier fusion may generate more accurate classification than each of the basic classifiers. Fusion is often based on fixed combination rules like the product, average etc. This paper presents decision templates as classifier fusion method for the recognition of the handwritten English and Farsi numerals (1-9).

The process involves extracting a feature vector on well-known image databases. The extracted feature vector is fed to multiple classifier fusion. A set of experiments were conducted to compare decision templates (DTs) with some combination rules. Results from decision templates conclude 97.99% and 97.28% for Farsi and English handwritten digits.

Keywords**—**Decision templates, multi-layer perceptron, characteristics Loci, principle component analysis (PCA).

I. INTRODUCTION

THE field of handwriting recognition has been a topic of insensitive research for many years. Handwritten digit recognition is still a problem for many languages like Arabic, Farsi, Chinese, English, etc [1]. Some of its important application is bank checking process, reading postal codes and reading different forms [2]. However, when the input stream of characters is not available from the keyboard, but rather through handwriting, it is useful to have a machine perform pattern recognition as an interface to be able to read and interpret the written text. A machine that do that can process more than a human being in the same time, this kind of application saves time and money and eliminates the requirement that a human perform such a repetitive task. For the recognition of English handwritten digits, various methods have been proposed [3]-[7]. But a few numbers of studies have been reported for Farsi language [8]-[10].

Learning is a process, that different procedures exceeds to different performance also feature extraction methods influence on performance. There is no standard algorithm to do best performance with a few numbers of data. Each of classification method dependent on different biases performs different from the other. For example two similar neural networks in type of multi-layer perceptron perform different because of randomly selection of first weights. In pattern recognition, the recognition rate of such systems is still low, and there is a need to improve it. The combination of multiple classifiers was shown to be suitable for improving the recognition performance in difficult classification problems [11]-[14]. It has become clear that for more datasets that are complicated various types of combining rules can improve the set of classifiers. Complexity in classification can be represented from limitation in number of data, classes overlapping and credible noise in data. Results from [15]-[17] show that when classifiers have small error rate and are independent in decision-making, combination of classifiers is useful. There are many ways to combine the results of a set of classifiers, depending on the type of the classifiers' output [18]. Classifiers are different by one of these methods.

a. Using different learning algorithms.

For example, we can use multi-layer perceptron, radial basis function network and decision trees in an ensemble to make a composite system [19].

b. Using certain algorithms with different complexity.

For example, using networks with different number of nodes and layers or nearest neighbor classifiers with different number of prototypes [20].

c. Using different learning parameters.

For example in an MLP, different initial weights, number of epochs, learning rate and momentum or cost function can affect the generalization of the network [20].

In the most general case, a classifier generates a score value for each class. In this case the sum, product, maximum, minimum, or the median of the scores of all classifiers can be calculated and the class with the highest value is regarded as the combined result. However, to actually build a multiple classifier system, one needs a number of basic classifiers.

The method developed here is based on a set of *c* matrices called decision templates (DTs). DTs are a robust classifier fusion scheme that combines classifier outputs by comparing them to a characteristic template for each class. DT fusion uses all classifier outputs to calculate the final support for each class, which is in sharp contrast to most other fusion methods. It uses only the support for that particular class to make their decision [18].

Reza Ebrahimpour is with Department of Electrical and Computer Engineering, Shahid Rajaee University, Tehran, Iran (e-mail: rebrahimpour@srttu.edu).

Samaneh Hamedi is with the Electrical and Computer Engineering Department, University of Shahid Rajaee, Tehran, Iran (e-mail: shamedi59@yahoo.com).

In this paper, the above mentioned model is investigated for recognition of both English and Farsi hand written digits. Basic classifiers are multi-layer perceptron neural networks. Section II introduces the feature extraction method and reduction large dimension. In Section III the details of combining classifier fusion schemes are discussed. Section IV contains our experimental results for two dataset (Pendigit and Hoda).

II. Feature Extraction

Transforming the input data into the set of features is called features extraction. Feature extraction is a special form of dimensionality reduction. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm, which overfits the training sample and generalizes poorly to new samples. When the input data to an algorithm is too large to be processed then the input data will be transformed into a reduced representation set of features (also named features vector). If the features extracted are carefully chosen it is expected that the features set will extract the input data in order to perform the desired task using reduced representation instead of the full size input. It is important to select type of feature extraction. Because it is the important, factor in the performance of pattern recognition systems [24]. Selection type of feature extraction is dependent on the application. Different features are purposed to recognize digits and characters. They are Furies Transform, Invariant Moments, Geometric Moments, Characteristic Loci and others [25]-[30]. In this paper Characteristic Loci is used to recognize digits. As it will be described later, Pendigit dataset is ready for classification and preprocessing has been done on pictures. Farsi dataset are 60,000 pictures that need to resize them in the same and change in to vector.

A. Characteristic Loci

It has been shown that Characteristic Loci Feature has good results in handwritten digit recognition [31]. Characteristic Loci Feature is commonly in vertical, horizontal and 45 or 135 degree orientation. A digit is related to each point of picture. This digit is dependent on the number of contacts along four directions, up, down, left and right. This feature vector has 81 components that each of them is relative amplitude of the specified digit in the picture. To normalize this, it can be divided to the number of black pixels. Fig. 1 shows how to calculate Characteristic Loci Feature vector for a pixel of a picture.

B. Dimensionality Reduction

One of the most important transform to reduce dimensionality for simple and fast data processing and picture classification is a linear Principal component (PCA). Now it is mostly used as a tool in exploratory data analysis and for making predictive models. It involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible [32]. There are three reasons why PCA is an appropriate transformation method for the handwritten digit dataset. Firstly, PCA as a linear transformation method is very simple. Secondly, the components of the transformed feature vector are statistically independent. Another reason for using PCA is that the feature components are ordered according to their importance. By removing the last *n* components, the dimensionality of the feature vector may be reduced without losing too much information. The optimal value for *n* must be empirically determined for each application. Suppose (feature vector) of train data are $T_1, T_2 \ldots, T_M$.

$$A = \frac{1}{M}\sum_{n=1}^{M} T_n \tag{1}$$

$$X_i = T_i - A \quad 1 < i < M \tag{2}$$

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

Where A is mean vector, X_i is the difference between feature vector and mean vector and Y matrix is determined by Equation (3). Covariance Matrix is N×N matrix that is represented in Equation (4); N is the dimension of feature vector

$$C = \frac{1}{M}\sum_{n=1}^{M} X_n X_n^T = \frac{1}{M} YY^T \,. \tag{4}$$

In this paper, first pictures are mapped in to center of a zero matrix with size (63×51). Then all of pictures are the same size. In this present N=81 component of characteristic Loci feature was calculated. Then a few number of PCA component was selected and these feature vector are transform matrices. These reduced vectors are given to neural network. By this method feature vector is reduced and large computational is decreased. It has been shown that 20 component of PCA concluded best results.

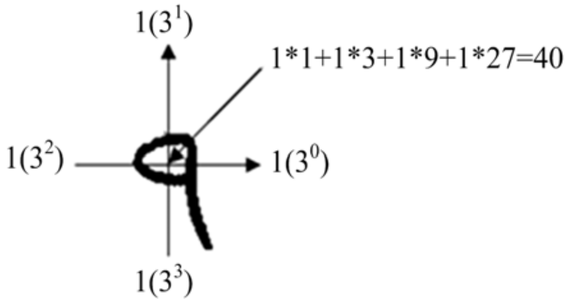

Fig. 1 Characteristic Loci Feature vector for a pixel of a picture

III. Combine Classifiers

Different algorithm for combining classifiers to achieve higher accuracy is offered. Some of them are voting pool of classifiers [20], max rules, min, average, product, Decision templates[18], mixture of experts [11-14] and divide-and-conquer classifiers [23] etc. Let x be a feature vector,

$\{\omega_1, \omega_2, ..., \omega_c\}$ be the label set of c classes and $\{D_1, ..., D_L\}$ be the set of L classifiers. The output of ith classifier is marked as $D_i(x) = [d_{i,1}(x), …, d_{i,c}(x)]^T$, where $d_{i,j}(x)$ is the degree of support given by classifier D_i to the assumption that x comes from class ω_j. D is the fused output of L firs-level classifiers as:

$$D(x) = F(D_1(x),...,D_L(x)) = [\mu_D^1(x),...,\mu_D^c(x)]^T \tag{5}$$

where F is called aggregation rule.

The classifier outputs can be formed in a decision profile (DP) as the matrix

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & ... & d_{1,j}(x) & ... & d_{1,c}(x) \\ ... & & ... & & ... \\ d_{i,1}(x) & ... & d_{i,j}(x) & ... & d_{i,c}(x) \\ ... & & ... & & ... \\ d_{L,1}(x) & ... & d_{L,j}(x) & ... & d_{L,c}(x) \end{bmatrix}$$

Some methods calculate the support for class ω_j using only the ith column of DP(x). Fusion methods that use the DP class-by-class will be called class-conscious. Examples of class-conscious fusion operators are average, minimum, maximum, product, etc.

A. *Decision Templates (DTs)*

Assuming $Z = \{z_1,..., z_N\}, z_j \in R^n$, is the labeled training data set. The decision template $DT_i(z)$ of class i is the $L \times c$ matrix $DT_i(z) = [dt_i(k,s)(Z)]$ whose (k ,s)th element is computed by:

$$dt_i(k,s)(Z) = \frac{\sum_{j=1}^N Ind(z_j,i) d_{k,s}(z_j)}{\sum_{j=1}^N Ind(z_j,i)}, k = 1,...L and s = 1,...,c, \tag{6}$$

Where $Ind(z_j, i)$ is an indicator function with value 1 if z_j has label i, and 0, otherwise [10]. The decision template DT_i for class i is the average of the decision profiles of the elements of the training set Z labeled in class i. When feature vector (x) is submitted for classification, the DT algorithm matches DP(x) to DT_i, i =1,…, c, and produces the class labels.

$$\mu_D^i(x) = \ell(DT_i, DP(x)), i = 1,..., c \tag{7}$$

Where ℓ indicates a similarity measure. The higher the similarity between the decision profile of the current x (DP(x)) and the decision template for class i (DT_i), the higher the support for that class. There are various measures of similarity that can be used [18]. Fig. 1 illustrates how the DT scheme operates. The decision templates are calculated in advance using Z in Equation (1). Measures of similarity based on the Euclidean distance between matrices DP and DT_i are used. The operation of it is shown in Equation (8).

$$N(DP, DT_i) = \mu_D^i(x) = 1 - \frac{1}{Lc}\sum_{k=1}^{L}\sum_{s=1}^{c}(dt_i(k,s) - d_{k,s}(x))^2 \tag{8}$$

In this study, two-procedure training is applied that its scheme is represented in Fig. 2. Four multi-layer perceptron with different neurons in hidden layer as basis classifiers train 15000 data that have been reduced dimensionality. There are different classifiers because of different hidden layers and different training weights. So combine of them can improve the performance of recognition. In the second procedure a decision template for each class is formed. Number of data as test are 5000 which each of them is compared with DTs of classes and final class is predicted.

IV. Experimental Results

The classic fusion and DTs method is tested on both standard Farsi and English dataset. For the English, Pendigit database is used [22]. For the Farsi dataset, experiments were performed on 15000 Farsi digit that randomly is selected from Hoda database that recently was developed [13].

A. *Farsi Digit Database*

Khosravi et al. [21] have introduced a very large corpus of Farsi handwritten digits. They extracted handwritten digits from 11,942 forms filled by diploma and bachelor students registered in the Iran's nationwide university entrance exam; 5,393 forms were filled by Diploma students and 6,549 others by BS students. All forms were scanned at 200 dpi resolution in 24 bit color format. After applying a threshold, they came into 102,352 binary images from which they chose 60,000 images for train and 20,000 for test.

B. *English Digit Database*

Fevzi et al have collect samples of ten handwritten digits from 44 writers. These writers were asked to write 25 examples from each of the digits in random order inside

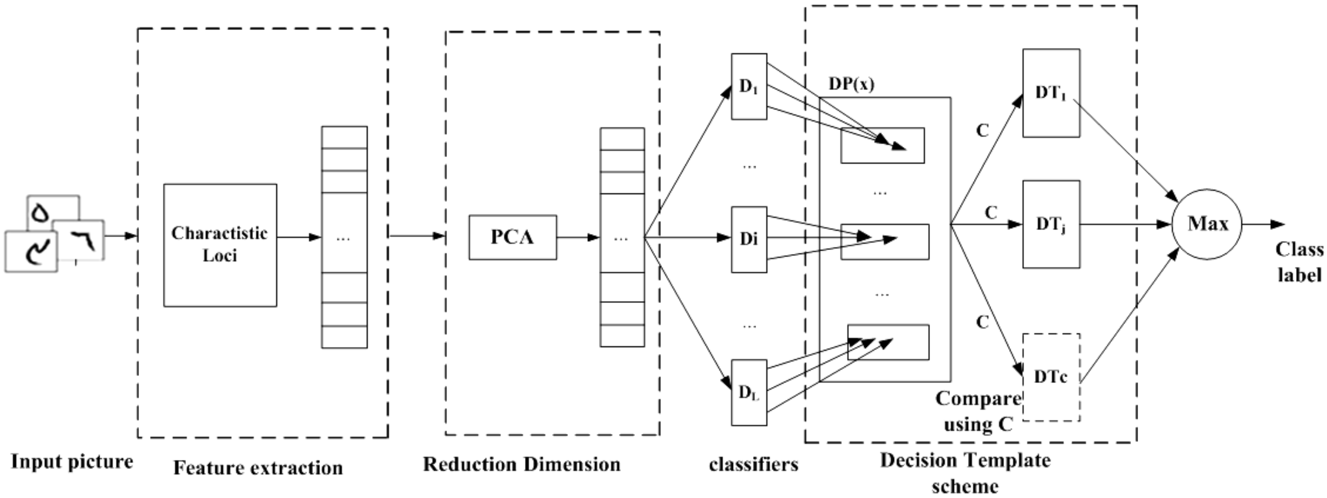

Fig. 2 Block diagram of handwritten digit recognition system, part feature extraction and reduction Dimension is just for Farsi database

TABLE I
DISTRIBUTION OF DIGITS IN TRAIN AND TEST SETS OF BOTH DIGIT DATASETS

		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**total**
Farsi	Train	6,000	6,000	6,000	6,000	6,000	6,000	6,000	6,000	6,000	6,000	60,000
	Test	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	20,000
English	Train	780	779	780	719	780	720	720	778	719	719	7494
	Test	363	364	364	336	364	335	336	364	336	336	3498

boxes of 500 by 500 tablet pixel resolution. Some preprocessing like normalize data, blur and down sample the image to get invariance to small translations and decrease dimensionality was done. After preprocessing, database is divided from 44 writers into one containing 30 writers and another containing 14 writers. The first part is for training, examples from 14 writers constitute the writer-independent test set. The database to the UCI Repository under the name Pendigits is available over the internet [22]. Distribution of digits of both database are represented in Table I. Patterns from English and Farsi datasets are illustrates in Fig. 3.

Scheme of Farsi hand written digit recognition system is represented in Fig. 2. It consists of four parts to recognition-handwritten digits. In the first part feature vector with 81 components is extracted. In dimensionality reduction, feature vector with 5, 10, 15, 20**,** 25 and 35 PCA component is selected. Tables II shows the classification accuracy of Farsi database for all PCA components. It is concluded that feature vector in length 20 has acceptable performance. The extracted vectors are fed to four individual classifiers. The classifier structures are multi-layer perceptron. The training parameters of network consist of 60 neurons in hidden layer and 400 epochs with 0.1 learning rate. After train full data, in DTs part, data that belongs to same class were simulated and matrix templates found as the class means of the classifier outputs. Finally test data is simulated and a decision profile is formed for each data. DP for each test data is compared with DTs of all classes. The nearest data to each DT is the incisive (winner) class. Decision Template part is the same for both Farsi and English hand written digit data.

Pendigit database is evaluated for different combining methods. The classifier structures are multi-layer perceptron. The training parameters of network for English database consist of 30 neurons in hidden layer and 100 epochs with 0.1 learning rate.

To evaluate the performance of DTs as fusion classifier method, experiments were conducted in handwritten digit recognition like minimum, maximum, average and product as the aggregation rule and stacked generalization method. The results are represented in Table II and III for both Farsi and English database. It can be seen that recognition rate of system for English database is 97.28% and 97.99% for Farsi database.

TABLE II
ACCURACY (IN %) FOR RECOGNITION FARSI HANDWRITTEN DIGITS BY SOME FUSION METHODS

Number of PCA components	**5**	**10**	**15**	**20**	**25**
MIN	82.62	91.66	93.78	93.98	93.76
MAX	80.64	91.24	93.60	94.20	93.92
AVG	83.94	91.72	93.72	94.5	94.46
PRO	83.3	92.12	93.92	94.42	94.5
Stacked	89.18	96.84	97.72	98.1	98.3
DT	86.24	95.78	96.80	97.99	97.82

TABLE III
ACCURACY (IN %) FOR RECOGNITION PENDIGIT DATABASE BY SOME FUSION METHODS

NN model	**MIN**	**MAX**	**AVG**	**PRO**	**Stacked**	**DT**
Accuracy	92.33	92.3	92.45	92.48	98.10	97.28

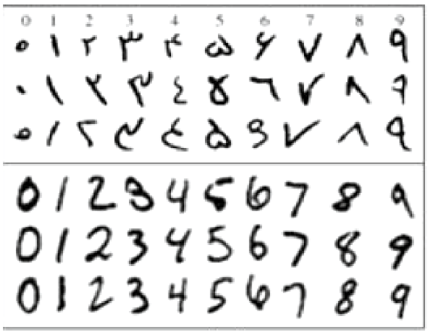

Fig. 3 Sample digits from Farsi and English datasets

V. Conclusion

In these paper decision templates (DTs) was used for recognition of Farsi and English handwritten digit. DTs are based on the similarity between the matrix of classifier outputs for an input x (DP(x)) and the c matrix templates found as the class means of the classifier outputs. Combining results of classifiers, recognition rate and generalization ability can be improved. For comparison some combining rule like minimum, maximum, average, product and stack generalization method were used to recognize handwritten digit too. It has been concluded that best recognition rate for Farsi database is 97.99% and for English database is 97.28%. Errors that limits recognition rate are because of uncorrected shape of digits. Combine classifiers have better results when basic classifiers have small error rate and they are different.

References

[1] S. Impedovo, P. Wang, and H. Bunke, editors, "Automatic Bankcheck Processing," World Scientific, Singapore, 1997.

[2] Nadal, C. Legault, R. Suen and C.Y, "Complementary Algorithms for Recognition of totally Unconstrained Handwritten Numerals," in Proc. 10th Int. Conf. Pattern Recognition, 1990, vol. 1, pp. 434-449.

[3] CL Liu, K Nakashima, H Sako and H. Fujisawa, "Benchmarking of state-of-the-art techniques," Pattern Recognition, vol. 36, no 10, pp. 2271–2285, Oct. 2003.

[4] M. Shi, Y. Fujisawa, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image," Pattern Recognition, vol. 35, no. 10, pp. 2051–2059, Oct 2002.

[5] LN. Teow and KF. Loe, "Robust vision-based features and classification schemes for off-line handwritten digit recognition," Pattern Recognition, vol. 35, no. 11, pp. 2355–2364, Nov. 2002.

[6] K. Cheung, D. Yeung and RT. Chin, "A Bayesian framework for deformable pattern recognition with application to handwritten character recognition," IEEE Trans PatternAnalMach Intell, vol. 20, no. 12, pp. 382–1388, Dec. 1998.

[7] IJ . Tsang, IR. Tsang and DV Dyck, "Handwritten character recognition based on moment features derived from image partition," in Int. Conf. image processing 1998, vol. 2, pp 939–942.

[8] H. Soltanzadeh and M. Rahmati, "Recognition of Persian handwritten digits using image profiles of multiple orientations," Pattern Recognition Lett, vol. 25, no. 14, pp. 1569–1576, Oct.2004.

[9] FN. Said, RA. Yacoub and CY Suen, "Recognition of English and Arabic numerals using a dynamic number of hidden neurons" in Proc. 5th Int Conf. document analysis and recognition, 1999, pp 237–240

[10] J. Sadri, CY. Suen, and TD. Bui, "Application of support vector machines for recognition of handwritten Arabic/Persian digits," in Proc. 2th Iranian Conf. machine vision and image procesing, 2003, vol. 1,pp 300–307.

[11] J. Kittler and F. Roli, "Multiple Classifier Systems," in Proc Int Work-shop on Multiple Classifieer Systems, 2000, Springer.

[12] C. Suen and L. Lam, "Multiple classifier combination methodologies for different output level," in Proc Workshop. Multiple Classifier Systems, 2000, pp.52-66.

[13] R. Duin and D. Tax, "Experiments with classifier combination rules," in Proc Workshop. Multiple Classifier Systems, 2000, pp.16-29.

[14] F. alimoglu and E. Alpaydin, "Combining Multiple Representation for Pen-based Hand written Digit Recognition," in Proc. 4th Int Conf. document analysis and recognition, 1997, vol. 2, pp 637–640.

[15] K. Tumer and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," Connect Sc, vol. 8, no. 8, pp. 385-404, 1996.

[16] Gunes, M. Menard, P. Loonis and S. Renaud, "Combination, Cooperation and Selection of Classifiers: a state of the art," Pattern Recognition, vol. 17, no. 8, pp. 1303-1324, 2003.

[17] LI. Kuncheva, M. Skurichina and R.P.W. Duin, "An experimental Study on diversity for bagging and boosting with linear classifiers," Information Fusion, vol. 3, no. 4, pp.245-258, Dec. 2002.

[18] LI. Kuncheva, JC. Bezdek and RP.W. Duin," Decision templates for multiple classifier fusion an experimental comparison," Pattern Recognition, vol. 34, no. 2, pp. 299-314, 2001.

[19] J. Kittler, A. Hojjatoleslami and T. Windeatt, "Weighting Factors in Multiple Expert Fusion," in Proc. Conf. British Machine Vision, 1997, pp. 42-50.

[20] T. Windeatt, R. Ghaderi, "Dynamic Weighting Factors for Decision Combining," in Proc. of IEE Int. Conf. Data Fusion, Great Malvern, UK1998, pp. 123-130.

[21] H. khosravi and E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties," Pattern Recognition letters, vol. 28, no. 10, pp. 1133-1141, July. 2007.

[22] C. J. Merz, P. M. Murphy, UCI Repository of Machine Learning Databases 1998, http://www.ics .uci.edu/mlearn/MLRepository.html

[23] LI. Kuncheva, RK. Kounchev, RZ. Zlatev, "Aggregation of multiple classification decisions by fuzzy templates", in Proc. 3th. European Congres on Intelligent Technologies and Soft Computing, Germany, August 1995, pp. 1470-1474.

[24] ID. Trier and AK. Jain, "Feature Extraction Methods for Character Recognition- A Survey," Pattern Recognition, vol. 29, no. 4, pp. 641-662, April. 1996.

[25] H. Takahashi, "A Neural Net OCR using geometrical and zonal pattern features," in Proc. 1th. Conf. Document Analysis and Recognition, 1991, pp. 821-828.

[26] L.O. Jimenez, A. Morales-Morell and A. Creus, "Classification of Hyperdimensional Data Based on Feature and Decision Fusion Approachs Using Projection Pursuit, Majority Voting, and Neural Networks," IEEE Trans. on Geoscience and Remote Sensing, vol. 37, no. 3, May 1999.

[27] Y. Li, "Reforming the theory of invariant moments for Pattern recognition," Pattern Recognition Letters, vol. 25, no. 7, pp.723-730, July. 1992.

[28] H.A. Glucksman, "Multicategory of Patterns Represented by High-Order Vectors of Multilevel Measurement," IEEE Transaction Computer, vol. C-20, no. 12, pp. 1593-1598, Dec. 1997.

[29] Sh. Shahreza, M.H. Khotanzad, A. ,"Recognition Letterpress Works Independent of Size and Displacement with Zernike Moments and Neural Networks", in Proc. 2th Iranian Conf of Electrical Engineering,1994, Trbiat modares Univ, vol. 5, pp. 417-424.

[30] K. Azmi, R. Kabir and E. Badi, "Recognition Printed Letters wit Zonong Features," Iran Computer Group, vol. 1, pp. 29-37, 2003.

[31] S.H. Nabavikahrizi, R. Ebrahimpour and E. Kabir, "Application of Combining classifiers for Recognition of Farsi handwritten digits," in Proc. 3th Iranian Conf. Machine vision and image processing, 2004, vol. 1, pp 115–119.

[32] Gorban, B. Kegl, D. Wunsch and A. Zinovyev (Eds.), Principal Manifolds for Data Visualization and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007. ISBN 978- 3 540-73749- 0

Reza Ebrahimpour was born in Mahallat, Iran, in July 1977. He received the BS degree in electronics engineering from Mazandaran University, Mazandaran, Iran and the MS degree in biomedical engineering from Tarbiat Modarres University, Tehran, Iran, in 1999 and 2001, respectively. He received his PhD degree in July 2007 from the School of Cognitive Science,

Institute for Studies on Theoretical Physics and Mathematics, where he worked on view-independent face recognition with Mixture of Experts. His research interests include human and machine vision, neural networks, pattern recognition.

Samaneh Hamedi received the B.Sc. degree in Electrical Engineering from university of Shiraz, Iran, in 2003 and she is the student of M.Sc. in Electrical Engineering from Shahid Rajaee University, Tehran, Iran. Her dissertation concerned nanometric displacement measurement systems.

She has published 3 papers in Iranian conferences and 1 paper in Honet (2009 IEEE). Mrs. Hamedi became the student member of IEEE in 2007. Currently she works in Nano-Photonics and Optoelectronics Research Laboratory and Machine Vision Laboratory Department of Electrical Engineering, Shahid Rajaee. Her research interests include nanodisplacement, optical instrumentation, optoelectronic circuits and neural networks.