International Journal of Computer Applications (0975 – 8887)
Volume 129 – No.16, November2015

The biologically inspired Hierarchical Temporal Memory Model for Farsi Handwritten Digit and Letter Recognition

Fatemeh Asgari
Department of computer engineering
Kharamzi University
Tehran, Iran

Ali Salehi
Department of computer engineering
Sharif University of technology
Tehran, Iran

ABSTRACT
It is herein proposed a handwritten digit recognition system which biologically inspired of the large-scale structure of the mammalian neocortex. Hierarchical Temporal Memory (HTM) is a memory-prediction network model that takes advantage of the Bayesian belief propagation and revision techniques. In this article a study has been conducted to train a HTM network to recognize handwritten digits and letters taken from the well-known Hoda dataset for Farsi handwritten digit. Results presented in this paper show good performance and generalization capacity of the proposed network for a real-world big dataset.

Keywords
Handwritten digit recognition; hierarchical temporal memory (HTM); Hoda handwritten digits dataset.

1. INTRODUCTION
Automatic handwritten digit recognition is used in many applications 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. Slowly, handwritten digit recognition and handwritten character recognition were introduced into this field. Recently these systems have achieved good results in recognizing of Farsi handwritten digits and characters so that their recognition rate have been reached over 98% and 97%, respectively[21,27,33,35,36,37.42].

The hierarchical temporal memory (HTM) has been used during the last decade to solve problems related to prediction, classification, recognition… etc. [6]. HTM is the hierarchical structure and temporal relations included in a training representation of the outer world's causes or, in other words, the categories of input patterns. As the state-of-the-art approach, it has principal potential for solving challenging PR problems that have not been handled so far. Thus, it is considered as the standard in Pattern Recognition.

2. HIERARCHICAL TEMPORAL MEMORY
In the following we are giving a brief overview of the essential concepts of HTM's algorithm. Hierarchical temporal memory (HTM) developed by Hawkins [7] and formalized by George [8]. Hierarchical Temporal Memory is in essence a method of modeling the mammalian neocortex. As the name implies it has a hierarchical structure and at the top of this hierarchy, an HTM network forms invariant representations of the world and the underlying learning algorithms used in it are not specific to particular sensory domains and can be applied to a wide set of problems that involve modeling complex sensory data. The basic idea in back of the HTM theory is that while causes can be seen from very different viewpoints, thus producing very different images on the retina, because the images are seen consecutively, over a very short period of time, the cortex still manages to attribute such a variation of images to the same cause when necessary.

HTM networks are made of nodes. Each node performs the same algorithm, but on different data [9]. As shown for example in Fig. 1 the number of nodes comprising level l is $2^{2\alpha-1}$, where α is the number of the levels in the network. It decreases at higher levels in the hierarchy, nodes in lower level observe causes belonging only to a restricted time and input scale while higher-level nodes are able to detect causes on a larger time scale and wider image input. Each node should group input patterns that are probable to have the similar cause. These nodes can receive many types of input vectors from spatially-specific areas, namely the receptive fields. Except the top node, all nodes through-out the hierarchy, carry out unsupervised learning. This node realizes task of a simple classifier that maps learned top-level belief patterns with known output categories [10]. Zeta1TopNode is typically established. Of course, one could potentially consider any supervised classifier to be used instead of it, so we use a HTM network using non-linear SVM classifier with RBF kernel [11] as a modified HTM.

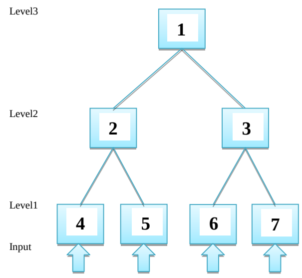


Fig 1: A simple HTM network that has 7 nodes organized in a 3 level hierarchy.

Each node uses two grouping mechanisms to form invariants, namely a spatial and a temporal pooler. These two distinct operations serve as concepts to refer to the form by which input vectors received during training are separated and stored into sets of temporally neighboring vectors. Initially spatial and temporal poolers are empty. Nodes experience two modes of function during their lifespan: the first one is the training mode and the second one is the inference mode. We describe them briefly in following section.

Unlike other soft computing approaches, time is a very important aspect of HTM; it is used by the neocortex as a supervisory signal for grouping together spatially different input vectors that tend to present themselves close together in time. The usage of temporal combination minimizes storage requirements and decreases the necessity for supervised training.

2.1 Spatial Pooler

In the training mode each frame (image) of the training sequences is presented to the sensory field (“retina”) of size m ×m pixels. At any time, only a single frame of the training dataset is presented to the level one node. This frame is reshaped into an input vector $x = (x_1 x_2 \ldots\ldots x_{m\times m})$ where characterize pixel values. Mentioned input vector are also referred to spatial coincidences in the HTM nonfictions [12]. The training data consecutively are shown to the node thus node learns the quantization space of such input vectors. In the network the first input vector frame is considered a new quantization center and is saved in the node. In the HTM network a node keeps input coming from children nodes and creates a group of different patterns seen by the node that differ by a configurable maximum distance. Commonly a threshold is used. Once the learning procedure is finished, the spatial module is learned and can produce outputs to the temporal module. Fig. 2 shows the stages an image goes through within a learning node.

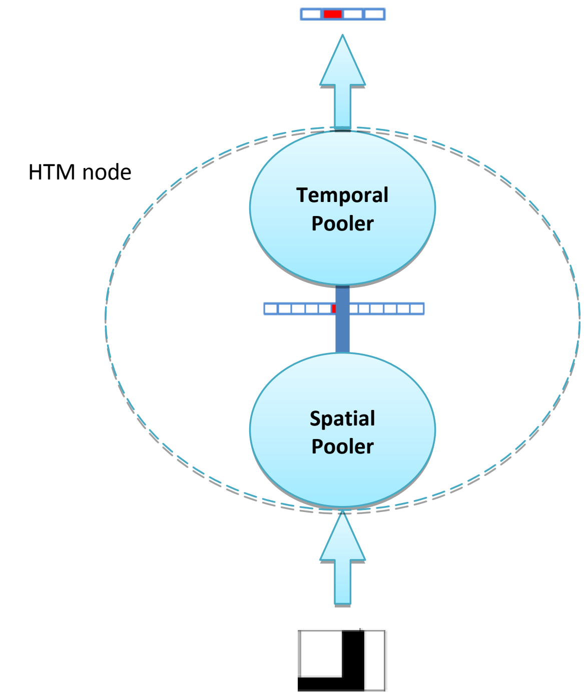


Fig 2: The stages a sample goes through within a learning node.

2.2 Temporal pooler

When training dataset is finished, the node should create a matrix called time adjacency. Each TAM’s element shows the transition between a spatial coincidence, represented by the row, and the spatial coincidence in the next time step, represented by the column. However, may be reasonable not to emphasis only on immediate steps, but also to increase transitions between current and few older steps. Consider n is the time gap between the current and past steps. Commonly it is performed by below:

$$L_n = transitionMemory - 1 + n$$

We used as a parameter to determine the number of steps that their coincidences affected, and is the increment amount [13]. Once learning of the TAM is finished and TAM matrix is stabilized adequately, a node that has completed its training phase can be switched to the inference mode.

Each node in network analyzes the transition frequencies saved in the TAM matrix. Aim of it’s to form a set of temporal groups include coincidences groups, according to how frequently they incline to follow each other in a period. In [14] could be found a sample of the learning procedure for a node. Training starts off on the lowest level of HTM layer and broadcasts throughout the hierarchy one layer at a time. When the learning procedure is finished at some level, all HTM nodes at this level are switched to the inference mode; hence the next level can remain at the learning mode.

3. INFERENCE MODE

In inference mode, the node continues to get input frames, but it does not learn of new input. Each image must be broadcasted through the hierarchy until the result vectors reaches the nodes that are currently learned. HTM nodes which are in inference mode just calculate to which temporal group the incoming input image most likely belongs. In fact first the similarity between the incoming image and saved images should be calculated. This is done by:

$$x_i = e^{-\frac{d_i^2}{\sigma^2}}$$

It is considered that the similarity between the patterns is a Gaussian distribution with zero mean. In the formula d_i is the dissimilarity between compared patterns. So $x = x_1, x_2 \ldots x_n$ shows the output vector of the first step of the inference. In this output vector, x_i indicate a probability value of ith temporal group being active. HTM nodes propagate this vector toward their parent nodes and parent nodes as their input vector aggregate all vectors incoming from children, Fig. 3. When the highest node of the HTM network is learned, training procedure is finished.

4. EXPERIMENTS AND RESULTS

To assess the performance of the HTM for handwritten digit and character recognition, in this section we apply it to two datasets: HODA Farsi handwritten digits dataset and HODA Farsi Hand Written letters dataset [15].

In all the experiments, we compared the performance of the modified HTM (HTM network using non-linear SVM classifier with RBF kernel) with other methods which applied to Farsi handwritten digits and letters. We used the Numenta platform NuPIC 1.7.1, developed by Numenta, Inc. This is a programming framework, which has a Python API that can be used to set up custom experiments. The selected library for the implementation of SVM was the LIBSVM, proposed by Chang and Lin [16].

4.1 Dataset

Khosravi et al. [15] already introduced a huge corpus of Persian handwritten digits which is termed HODA. HODA contains 102352 images digits. These images were scanned at 200 dpi with a high speed scanner. It is divided into a set of 60000 samples and 20000 samples used for the training set and testing set respectively. Some samples of 10 classes are shown in Fig. 4.

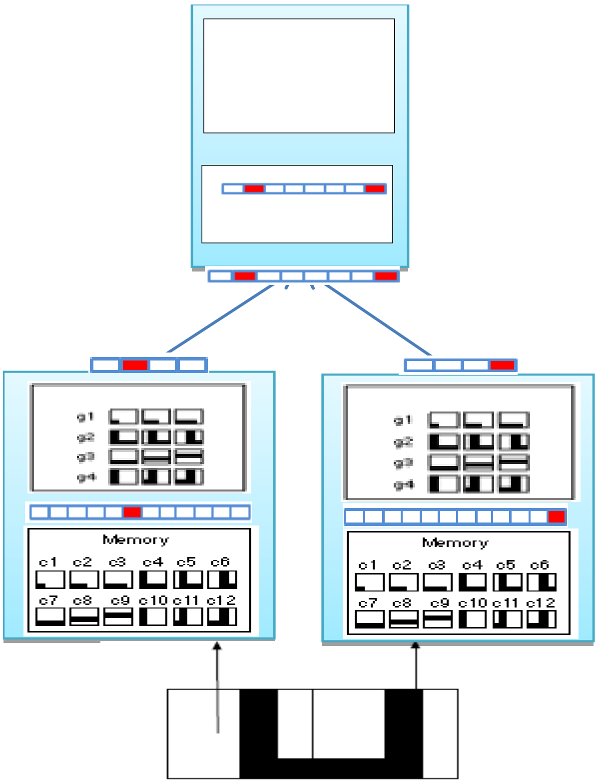


Fig 3: Each HTM nodes propagate output vector toward their parent nodes and parent nodes as their input vector aggregate all vectors incoming from children.

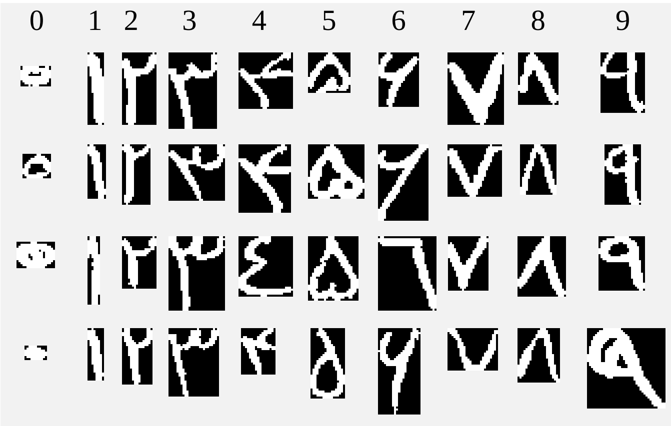
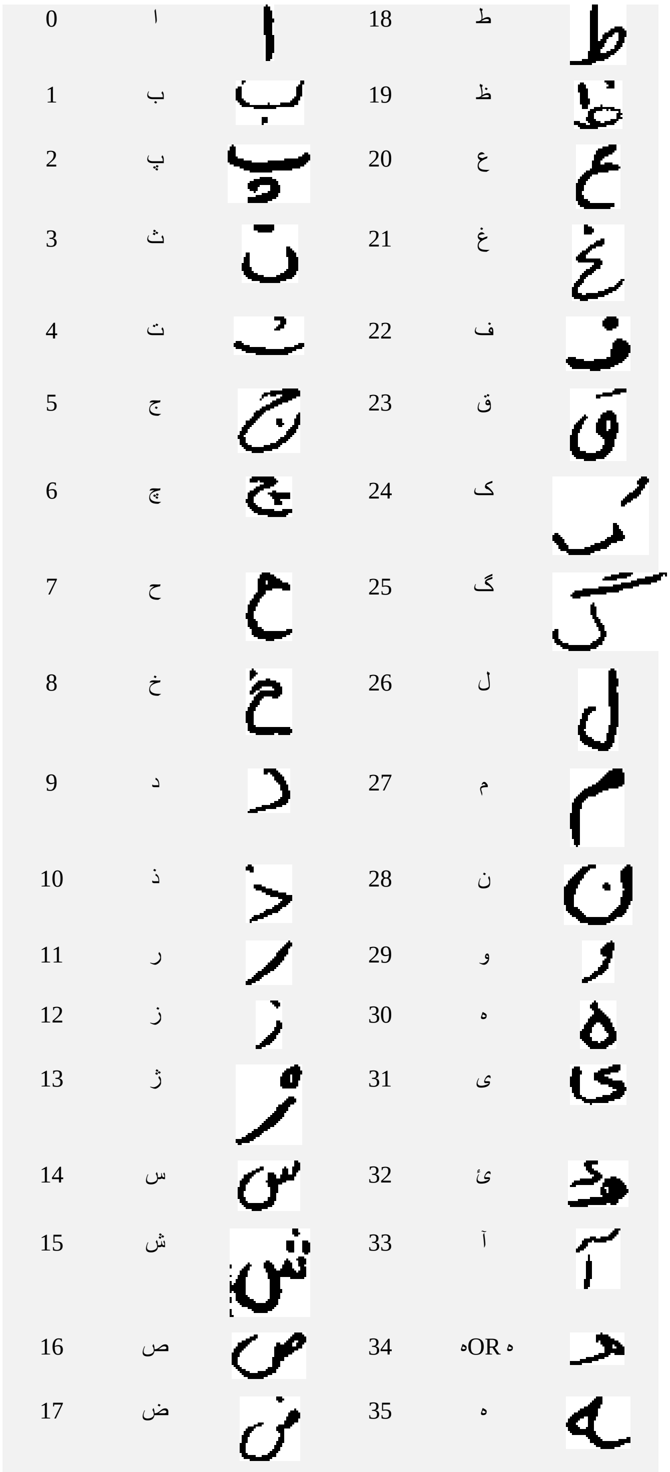


Fig 4: Example of a figure Examples of handwritten digits from Farsi digit datasets.

Also they introduced HODA Farsi Hand Written letters dataset, includes 88351 samples which 70645 samples are used for training and 17706 are used for testing phase[1]. Some samples are shown in Fig. 5.



Label	Persian Spell	One sample	Label	Persian Spell	One sample
0	ا		18	ط	
1	ب		19	ظ	
2	پ		20	ع	
3	ث		21	غ	
4	ت		22	ف	
5	ج		23	ق	
6	چ		24	ک	
7	ح		25	گ	
8	خ		26	ل	
9	د		27	م	
10	ذ		28	ن	
11	ر		29	و	
12	ز		30	ه	
13	ژ		31	ی	
14	س		32	ئ	
15	ش		33	آ	
16	ص		34	ﻫOR ﻪ	
17	ض		35	ه	

Fig 5: Sample handwritten digits from Farsi letter datasets and their labels.

5. PREPROCESSING

Frequently, the character images in handwritten character recognition systems required to be binarized. This is a restriction, but our introduced method work on both binary and gray scale data.

In Farsi handwritten digits and letters doing the preprocessing work, each image goes through a two stepped process; normalization and binarization. Character normalization is in order to standardize the character location and size in the image. All numbers are fitted into a

[1] These two datasets are available at http://www.modares.ac.ir/eng/kabir.

window of size 50 ×50 and all letters to 60 ×60. We use standard bilinear transformation, by which, every input bitmap A, of size p ×q is transformed into a normalized bitmap B, of size m ×n. Both A and B are quadrilateral regions [17]. Then by swapping all pixels they were converted to binary images in the input data.

Image binarization converts an image of up to 256 gray levels into a two-tone image. Commonly an image has two tones, i.e., 0 and 1 (1 is used to represent black and 0 is used to represent white). We convert gray scale images into binary images using Otsu algorithm [18]. Thus, each data in the dataset has a uniform background with a high-contrast character. Otsu algorithm performs satisfactorily on our experimental dataset. Some results of thresholding are shown in fig 5. Each sample in the first line is thresholded to give a binary sample in the second line.

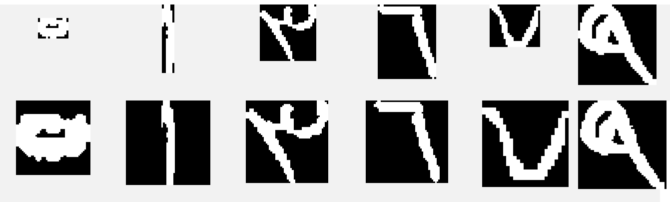


Fig 6: Samples after binarization and size normalization

6. CLASSIFICATION RESULTS

In a HTM network, to increase the performance, the selection of the parameters for the spatial module is of great importance. Parameters Sigma, transitionMemory and grouper algorithm were set to 1, 1 and 16, respectively. MaxDistance in 1 and 3 layer was 1. maxDistance on the low level defines the minimum value the squares of the Euclidean distances between an input image A and all the previously memorized inputs have to take in order for A to be considered a new image. In modified HTM, after feature vectors were derived in layer 3, they were fed to a classifier in top nod, and all-pairs multi-class non-linear SVM classifier was used for classification. Table5 show confusion matrix for handwritten Persian digits in the dataset. We obtained 99.12% accuracy using a few of preprocessing.

Table 1 and table 2 compared accuracy of the proposed method with other methods were used in recent researches on HODA dataset and Farsi handwritten letters datasets, HODA and IFHCDB [19]. Results are for test sets and are averaged over ten runs. In Table below most of the existing works were evaluated on smaller datasets. We used 80,000 data for our experiment. The Highest accuracy was obtained from the work due to Soltanzadeh et al. [27] but they have tested with Only 8,918 data and used 257 dimensional features.

Table 1. Result of different algorithms on HODA digits dataset

Algorithms	Train Size	Test Size	Accuracy (%)
Ziaratban et al. [20]	6000	4000	97.65
Shirali-shahreza et al. [21]	2600	1300	97.80
Dehghan, Faez [22]	6000	4000	97.01
Mowlaei, Faez [23]	2240	1600	92.44
Ebrahimpour et al. [24]	6000	2000	95.3
Javidi and sharifizadeh [25]	6000	2000	98.16
Mozaffari et al. [26]	2240	1600	91.37
Soltanzadeh,Rahmai [27]	4979	3939	99.57
Harifi.,Aghagolzadh [28]	230	500	97.60
Hosseini,Bouzerdm [29]	480	480	92.00
Mozaffari et al. [30]	2240	1600	94.44
Mowlaei et al. [31]	2240	1600	91.88
Sadri et al. [32]	7390	3035	94.14
Ebrahimpour et al. [33]	60000	20000	97.52
Rashnodi et al. [34]	60000	20000	98.94
Rashnodi et al [35]	60000	20000	98.84
Proposed Algorithm	60000	20000	99.12

7. CUNCLUSION

In this study, modified HTM was used for Farsi handwritten digit and letter recognition. A set of features was first extracted from a training set of digit and letter images. A classifier was then constructed over these data and was evaluated over a separate test set. High handwritten digit recognition proves appropriateness of these features for the mentioned task. Modified HTM achieved better performance than other methods. We achieved 99.12% recognition rate using modified HTM. Further, to the best of our knowledge, this work is one of the paramount works, towards the recognition of Persian handwritten digits and letters on a huge dataset.

Table 2. Result of different algorithms on HODA and IFHCDB letter datasets

Algorithms	Dataset	Train Size	Test Size	NO. of Clusters	(%)
Alaei et al. [36]	IFHCDB	36682	15338	8	98.1
Alaei et al. [36]	IFHCDB	36682	15338	32	96.68
Mozaffari et al. [37]	IFHCDB	3200	2880	8	87.26
Dehghan et al. [38]	IFHCDB	1600	1600	20	96.92
Shanbezade et al. [39]	IFHCDB	1600	1200	32	87
Mowlai et al. [40]	IFHCDB	3200	2880	8	32.75
Ziaratban et al. [41]	IFHCDB	11471	7647	8	87.26
Matin Niya,	HODA	70645	1770	36	97.89

Sajedi[42]			6		
Proposed Algorithm	HODA	70645	17706	36	98.93

8. REFERENCES

[1] V. Ghods and M. K. Sohrabi, "Online Farsi digit recognition using their upper half structure," in Sixth International Conference on Graphic and Image Processing (ICGIP 2014), 2015, pp. 94430B-94430B-6.

[2] Ping Zhang, Reliable recognition of handwritten digits using a cascade ensemble classifier system and hybrid features, Ph.D. thesis, Concordia University, Montreal, P.Q., Canada, 2006.

[3] Fabien Lauer, Ching Y. Suen, and Gérard Bloch, "A trainable feature extractor for handwritten digit recognition," Pattern Recognition, vol. 40, no. 6, pp. 1816–1824, 2007.

[4] Patrice Y. Simard, Dave Steinkraus, and John C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," International Conference on Document Analysis and Recognition, vol. 2, pp. 958–963, 2003.

[5] E.M. Kussul, T.N. Baidyk, D.C. Wunsch, O. Makeyev, and A. Martin, "Permutation coding technique for image recognition systems," IEEE Transactions on Neural Networks, vol. 17, no. 6, pp. 1566–1579, 2006.

[6] Price, Ryan William. Hierarchical Temporal Memory Cortical Learning Algorithm for Pattern Recognition on Multi-core Architectures. Diss. Portland State University, 2011.

[7] Hawkins, J., Blakeslee, S. (2004). On intelligence. Henry Holt and Company, New York.

[8] George, D., Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. PLoS Computational Biology 5(10). DOI 10.1371/journal.pcbi.1000532.

[9] George, D., Jaros, B. 2007. The htm learning algorithms. Whitepaper, Numenta Inc.

[10] Numenta (2007). Zeta1 algorithms reference. Document version 1.0.

[11] Chauchard, F., et al. "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes." Chemometrics and Intelligent Laboratory Systems 71.2 (2004): 141-150.

[12] George, D., Hawkins, J.: A hierarchical bayesian model of invariant pattern recognition in the visual cortex. In: Proceedings of 2005 IEEE International Joint Conference on Neural Networks, IJCNN 2005, July 4-August, vol. 3, pp. 1812–1817 (2005)

[13] Numenta (2009). Numenta node algorithms guide, NuPIC 1.7.

[14] Kostavelis, I., Gasteratos, A., 2011. Laboratory of Robotics and Automation.http://utopia.duth.gr/~gkostave/downloads/operation_of_nodes_in_the_first_level.rar

[15] Khosravi, H., Kabir, E.: Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. Pattern Recognit. Lett. 28(10), 1133–1141 (2007).

[16] Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27:1–27:27, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

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

[18] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," Systems, Man and Cybernetics, IEEE Transactions on, vol. 9, no. 1, pp. 62-66, 1979.

[19] S. mozaffari, K. faez, F. Faradji et al., "A comprehensive isolated Farsi/Arabic character database for handwritten OCR research." pp. 385-389.

[20] 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.

[21] M. H. Shirali-Shahreza, K. Faze 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, 1992.

[22] K. Fouladi, B. N. Araabi, and E. Kabir, "A fast and accurate contour-based method for writer-dependent offline handwritten Farsi/Arabic subwords recognition," International Journal on Document Analysis and Recognition (IJDAR), vol. 17, pp. 181-203, 2014.

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

[24] R. Ebrahimpur, A. Esmkhani, and F. Faradji, "Farsi handwritten digit recognition based on mixture of RBF experts," IEICE Electron. Express, vol. 7, no. 14, pp. 1014-1019, 2010.

[25] M. M. Javidi, and f. Sharifizadeh, "A Modified Decision Templates Method for Persian Handwritten Digit Recognition," Journal of American Science, vol. 8, no. 1, pp. 504-512, 2012.

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

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

[28] A. Harifi and A. Aghagolzadeh," A New Pattern for Handwritten Persian/Arabic Digit Recognition",

Journal of Information Technology vol. 3, pp. 249-252, 2004.

[29] M. R. Keyvanpour, R. Azmi, Z. S. M. Tabatabai, and Z. Abdolhosseini, "Handwriting Persian character recognition using optimized structural elements," Global Journal of Information Technology, vol. 4, 2015.

[30] 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.

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

[32] 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.

[33] R. Ebrahimpur, M. R. Moradian, A. Esmkhani et al., “Recognition of Persian handwritten digits using Characterization Loci and Mixture of Experts,” International Journal of Digital Content Technology and its Applications, vol. 3, 2009

[34] Rashnodi, Omid, Hedieh Sajedi, and Mohammad Saniee Abadeh. "Using Box Approach in Persian Handwritten Digits Recognition." International Journal of Computer Applications 32.3 (2011)..

[35] Rashnodi, Omid, Hedieh Sajedi, and Mohammad Saniee Abadeh. "Persian Handwritten Digit Recognition using Support Vector Machines." International Journal of Computer Applications 29.12 (2011).

[36] A. Alaei, P. Nagabhushan, and U. Pal, "A New Two-Stage Scheme for the Recognition of Persian Handwritten Characters." pp. 130-135.

[37] S. Mozaffari, K. Faez, and H. R. Kanan, "Recognition of isolated handwritten Farsi/Arabic alphanumeric using fractal codes." pp. 104-108.

[38] M. Dehghan, and K. Faez, "Farsi handwritten character recognition with moment invariants." pp. 507-510 vol.2.

[39] J. Shanbezadeh, H. Pezashki, and A. Sarrafzadeh, "Features Extraction from Farsi Hand Written Letters." pp. 35-40.

[40] A. Mowlaei, and K. Faez, "Recognition of isolated handwritten Persian/Arabic characters and numerals using support vector machines." pp. 547-554.

[41] M. R. Keyvanpour, R. Azmi, Z. S. M. Tabatabai, and Z. Abdolhosseini, "Handwriting Persian character recognition using optimized structural elements," Global Journal of Information Technology, vol. 4, 2015.

[42] K. Fouladi, B. N. Araabi, and E. Kabir, "A fast and accurate contour-based method for writer-dependent offline handwritten Farsi/Arabic subwords

[43] recognition," International Journal on Document Analysis and Recognition (IJDAR), vol. 17, pp. 181-203, 2014