International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 1 (2015) pp. 1911-1924

http://www.ripublication.com

Fuzzy C Means Based Hybrid Classifiers For Offline Recognition Of Handwritten Indian (Arabic) Numerals

Maen Takruri, Rami Al-Hmouz, Ahmed Al-Hmouz, Mohammad Momani

American University of Ras Al Khaimah, Ras Al Khaimah, UAE
maen.takruri@aurak.ae
King Abdulaziz University, Jeddah, Saudi Arabia
ralhmouz@kau.edu.sa
Middle East University, Amman, Jordan
aa998@uowmail.edu.au
University of Technology, Sydney, Ultimo, NSW 2007, Australia
Mohammad.Momani@uts.edu.au

Abstract

In this study, we present two offline hybrid classification approaches for Indian (Arabic) handwritten numerals in effort to obtain higher reliability and classification rates than those obtained by single classifiers. Both methods work at the pixel level. No feature extraction methods are used as the purpose of this study is to focus on classifiers. The first hybrid classifier introduced is the serial hybrid classifier. It consist of three consecutive single classifiers. The first level is Fuzzy C-Means classifier followed by Support Vector Machine for as second level when more details are required and finally confirmation of classification will be through unique pixels method which forms the third classification level. The second hybrid classifier is the parallel hybrid classifier. It fuses the decisions simultaneously obtained from a Fuzzy C-Means classifier with the decision of a Neural Network to obtain the final decision. Both algorithms are tested on the CENPARNI Indian (Arabic) handwritten numerals dataset. The overall testing accuracy reported is 88%, 89% for the serial hybrid classifier and the parallel hybrid classifier, respectively. The paper also reports the results obtained using different types of single classifiers and compares with the above mentioned hybrid classifiers results. It shows the superiority of the hybrid classifiers over single classifiers.

Keywords-component; OCR; Handwritten Indian (Arabic) numbers; Fuzzy C Means; SVM, NN, Bayesian Fusion.

Introduction

Optical Character Recognition (OCR) is the process of automatic recognition and identification of characters from images. Over the past 50 years, it has gone through major developments. Currently, OCR is implemented in many applications including data entry, signature identification and License Plate Recognition (LPR) [1]. OCR performance primarily depends on the quality of the input image; most of the existing OCR systems work with a very constrained character images and are still unable to provide a reliable accuracy under various conditions; thus, no comparisons can be made with human reading capabilities [2].

Various OCR methods have been explored by scholars, including those based on template matching, Neural Network (NN), Hausdorff distance, Support Vector Machine (SVM), Hidden Markov Models (HMMs) and the probabilistic model. In template matching and Hausdorff distance methods, the character image is compared to templates (character images), and the image that has the minimum distance or the highest match is considered to have the same features of the character image. The distance between two images can be measured by Euclidean distance, Hausdorff distance [3, 4], Chamfer matching [5] and cross correlation. The correlation measure is achieved through Normalized Cross Correlation (NCC).

Support Vector Machine based recognition was investigated by [6]. Four SVMs were used to classify characters and numerals. Results were then fed into 10 SVMs for numerals (0-9) and into 26 SVMs for characters (A-Z). A recognition rate of 97% was reported. SVM with SIFT descriptors was also investigated by [7], however SVM classifier is still very sensitive to noise. HMM was also used for character recognition [8]; however, the result was still not convincing compared with other methods.

Data clustering which is a form of unsupervised learning has also been used in OCR literature for character classification. Basically, clustering is the classification of patterns or objects based on how similar or different they are from each other. Similar objects are considered to belong to the same group (cluster) while others will belong to some other groups based on a similarity measure such as the Euclidian distance. In general, clustering schemes are subdivided into two main categories namely, hard clustering and soft (fuzzy) clustering. Unlike the conventional hard clustering method which restricts each point of the dataset to exclusively belong to one cluster [9], fuzzy clustering allows the points to belong (be a member of) to more than one cluster with a certain degree of membership [10]. In the context of image segmentation, fuzzy clustering is proven to perform better than hard clustering when talking about real life images suffering from issues such as limited spatial resolution, poor contrast, overlapping intensities, and noise and intensity inhomogeneities [10]. Among the fuzzy clustering methods, Fuzzy C-Means (FCM) algorithm [11] is the most popular method used in image segmentation because it has robust characteristics for ambiguity and can retain much more information than hard segmentation methods [12].

FCM was introduced in [13], developed in [14], generalized in [11]. The FCM algorithm and its derivatives have been used successfully in many applications, such as pattern recognition, classification, data mining, and image segmentation [15].The algorithm is an iterative clustering method that produces an optimal c partition by minimizing the weighted within group sum of squared error objective function [10]. According to [16], FCM is the fuzzy variant of K-means as it allows pixels to belong to multiple clusters with varying degrees of membership. It is also considered as an improvement over K-means when the data set cannot clearly subdivide into underlying partitions [15].

Recently, the issue of recognition of handwritten offline Arabic text and numerals has received more attention in literature. Driven by the need to have real-life automatic systems to recognize, digitize, store and analyze handwritten Arabic documents, various efforts were directed towards coming up with a reliable Arabic OCR system that can recognize handwritten numerals, numeral strings, legal amounts, letters, words, dates and symbols in Arabic language. For that reason, several databases for Arabic handwritten recognition were developed. Of these were the IFN/ENIT database which was developed in 2002 [17], and consisted of 26,549 images of handwritten Arabic words, the AHDB database developed in 2003 [18] and consisted of handwritten words describing numbers and quantities used in checks and the Arabic Check Database developed in CENPARNI center in 2003 for the recognition of hand-written Arabic checks [19].

A more recent effort to develop a standard Arabic databases resulted in the CENPARNI Arabic Dataset [20, 21] which aimed to serve as a standard comprehensive Arabic handwriting benchmark database. It includes datasets for Indian (Arabic) isolated handwritten numerals, Indian (Arabic) numeral strings, isolated Arabic letters, Arabic words and special symbols that are used in Arabic documents. It is worth mentioning here that in the Arabic-script based languages, such as Arabic, Urdu, Dari, Farsi, and Pashto, Indian numerals are used, as Arabic numerals are used in Latin scripts [20]. For this reason we refer to the Indian numerals used in Arabic documents as Indian (Arabic) numerals as they slightly differ from the Indian numerals used Urdu, Dari, Farsi, and Pashto. The KHATT database [22] is another comprehensive offline handwritten Arabic text database that is designed for the use in Arabic text recognition research, writer identification and verification research and for Arabic forms analysis research.

A survey of the offline Arabic handwriting recognition efforts was given in [23]. The article discussed the research in the Arabic text handwriting recognition, at the time, in terms of the feature extraction methods, the classification and recognition techniques applied and the available databases. It also provided a comparison of the results of the Arabic text recognition but did not address the problem of numerals recognition. The problem of Indian (Arabic) numerals was addressed by many researchers with focus on the recognition of isolated handwritten numerals [20, 21, 24, 25], connected numerical strings [20, 21], Arabic dates [20,21], and numerals in Arabic legal Documents and Checks [19].

An asymmetrical segmentation pattern to obtain a feature vector for handwritten Indian (Persian/Arabic) numeral was used in [26]. The feature vectors were fed to a Neural Network classifier with a reported recognition rate of 97.6 %. The database used consisted of 730 digits written by 73 participants. A multiple feature/resolution scheme for Indian numeral recognition using Hidden Markov Models (HHMs) was proposed in [27]. The features used included gradient, structural, and concavity features. An average recognition rate of 99% resulted from testing the scheme on a database of 21,120 images obtained from 44 writers with an average of 48 samples per numeral.

Gradient features were extracted in [21] from the Indian (Arabic) handwritten numerals of the CENPARNI Arabic dataset by convolving each pixel neighborhood with Roberts' Cross operators [28, 29] to calculate the gradient strength and gradient direction. Feature vectors were generated as a composition of the gradient strengths accumulated in different directions. The feature vectors were used for training and testing a Radial Basis Function SVM classifier system resulting in a recognition result of 97.29%. The same dataset was used in [20] for training and testing resulting in a recognition rate of 93.60%. The recognition system in [20] used a K-NN classifier with the same features used in [30] however instead of being extracted from handwritten Arabic numbers as in [30] they were extracted from handwritten Indian (Arabic) numbers. The feature vectors used in [30] were reduced resolution version (in the vertical direction and the horizontal) of the skeleton of the input number.

A novel three-level approach for classification of isolated Indian (Arabic) handwritten numerals was introduced in [31]. After being cropped, resized and vectorized, the binary image vector of the handwritten numeral was fed to an FCM classifier stage followed by an SVM classifier stage to reduce the misclassification rate of the FCM. Three SVM blocks were implemented in parallel to differentiate between digit pairs (0,1), (9,6) and (2,4). Finally, unique pixel method was applied as the third classification stage for the cases of (3, 6, 7, 8, 9). The unique pixel method decision was compared with decision of the first stage classifier (FCM) and the second stage classifier (SVM). If their decision did not match with unique pixels method result, the second maximum of the FCM classifier (the one with second degree membership) were then fed to the SVM stage for cases (0,1,4,5, 0) and again to third level of pixel locations for the cases (3,6,7,8,9) and so on.

In this study, we extend our work in [31]. We present two hybrid approaches for the classification of isolated Indian (Arabic) handwritten numerals from images. The first is the serial hybrid classification approach which is the same three-level classification approach (FCM, SVM and Unique pixel method) introduced in [31]. The other one is the parallel hybrid classification approach which is based on Bayesian fusion of the decisions simultaneously given by an FCM classifier and a NN to result in final classification decision. The evaluations of our approaches were conducted on the CENPARNI Arabic database, particularly, the isolated Indian (Arabic) handwritten numerals database [20, 21] which has been widely used in the Indian (Arabic) handwritten numerals recognition literature. The paper also reports the results obtained using different types of single classifiers (on the same dataset) and provides a comparison with the results obtained by the above mentioned hybrid classifiers. It shows the superiority of hybrid classifiers over single classifiers.

The rest of the paper is organized as follows: Sections II and III present our serial and parallel hybrid classification approaches for Indian (Arabic) handwritten numerals, respectively. The evaluations of the proposed approaches are given in section IV. Section V concludes with future work.

Serial Hybrid Classification Approach

Unlike the numbers used in the rest of the world, the Indian (Arabic) numbers have not received much attention in the OCR literature. Here we present an algorithm for the classification and recognition of these numbers. It is a 3-level classification approach [31]. The first classification level uses FCM for clustering, the second level utilizes SVM and the third uses the unique pixel method as explained below:

Fuzzy C-Mean Classifier

Firstly, each digit image is converted to a binary form, cropped, resized to a 20 x 20 matrix and then vectorized before being fed to an FCM classifier. The values for the elements of the 400 pixels vector will be either 0 or 1. Hence, the values of the prototypes for the training set using FCM will also be either 0 or 1 considering two clusters. In this case, the membership to prototypes will also be either 0 or 1.

In this approach, the prototype vector (which is the cluster center) for each Indian (Arabic) digit is forced to be the pixel by pixel average of all images corresponding to that digit across the training set. Figure 1 shows our training set average digit images for the numeric digits from 0-9. If all pixels in the dataset for a given numeric digit are 0 for the same pixel location *i* , then the average is 0 and the opposite applies for 1. In reality, this not true. It can be noticed most pixel location do not have 0 or 1 for all images for location *i* due to the variations in the handwritten digits samples.

We force the prototypes in location *i* to have the value of v_{1i} (mean of pixels) and v_{2i} (1-mean of pixels). Each pixel x_i will be associated to prototypes according to [11]:

$$\eta_{ij1} = \frac{1}{\sum_{l=1}^{2}\left(\frac{\| x_i - \mathbf{v}_{ji} \|}{\| x_i - \mathbf{v}_{li} \|}\right)^{2/(m-1)}} \qquad (1)$$

Where *m* is the Fuzzification Coefficient (we use *m*=2).

For the training dataset, each pixel *i* (*i* =1, 2,..., 400.) in the given numeric images set for digit *j* (*j* =0,1,2,..9) will have memberships to the v_{1i} and v_{2i}=(1- v_{1i}) by η_{ij1} and η_{ij2} = (1-η_{ij1}), respectively. The average memberships n_{ij1} and n_{ij2} are then calculated by:

$$n_{ij1} = \frac{\sum_{k=1}^{N_j} \eta_{ij1}(k)}{N_j}$$

$$n_{ij2} = \frac{\sum_{k=1}^{N_j} (1-\eta_{ij1}(k))}{N_j}$$

Where N_j is the number of images included in the training set corresponding to number (class) *j*.

At the testing stage, the membership of each pixel in a digit image is calculated. This is done for all numbers (0-9). The memberships a_{ij1} and a_{ij2} = (1-a_{ij1}) of each pixel to prototypes v_{1i} and v_{2i} are calculated using (1). Then using the average memberships n_{ij1} and n_{ij2} calculated in the training phase, the class memberships (ω_j) of each image in the testing set are calculated by:

$$\omega_j = \sum_{i=1}^{400} \sum_{j=0}^{9} \max\{\min\left(a_{ij1}, n_{ij1}\right), \min\left(a_{ij2}, n_{ij2}\right)\} \tag{2}$$

The recognized numeric will then be j_0 as follows:

$$j_0 = \arg\max_{j=0,1,2,\ldots 9} \omega_j \tag{3}$$

Support Vector Machine

In order to reduce the misclassification rate of FCM, we use the well-known SVM classifier as level-2 classifier. Three SVMs are implemented to differentiate between the similar Indian (Arabic) digits pairs (0,1), (9,6) and (2,4) which (as will be seen in the evaluations) have the highest FCM misclassification error because of the apparent similarity in their shapes.

Unique Pixel Method

The unique pixel method is applied as the third classification level. It works on determining the pixel areas that are unique to each digit. So if unique pixels areas appear in the image, the digit is directly identified. These areas are used as identifiers for the classification of the corresponding digits. Figure 2 shows the unique pixels that exist in the images for each digit in the training data set.

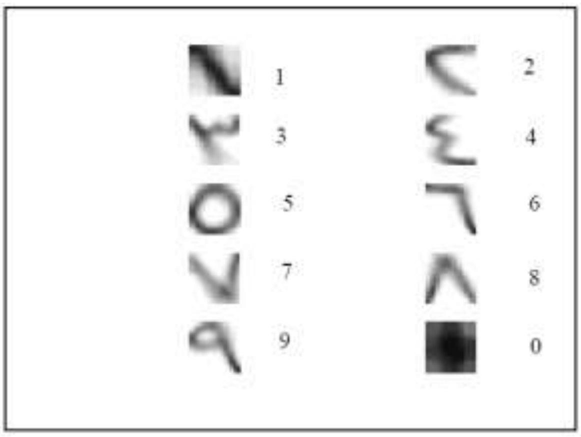

Fig. 1 Averaged images for handwritten Indian (Arabic) digits (0-9)

Fig. 2 Locations of unique pixels for the digits shown in figure 1

From figure 2, it can be noticed that the Indian (Arabic) numbers (3, 6,7,8 and 9) have unique pixels (white areas) that only appear on them. They are unique in the senses that they are different for each digit and therefore can be used as identifiers for the corresponding digits. On the other hand, it can be clearly seen that the images for digits 1,2,4,5 and 0 do not have such pixels and therefore, the unique pixel method cannot be used to identify them.

The unique pixel method decision is compared with decision of the lower classifiers (FCM) and (SVM). If their decision do not match with unique pixels method result, the second maximum of the FCM classifier (the one with second degree membership) will then be fed to SVM for cases (0,1,2,4,6,9) and again to the third level of pixel locations for the cases (3,6,7,8,9) and so on. Figure 3 shows the proposed algorithm.

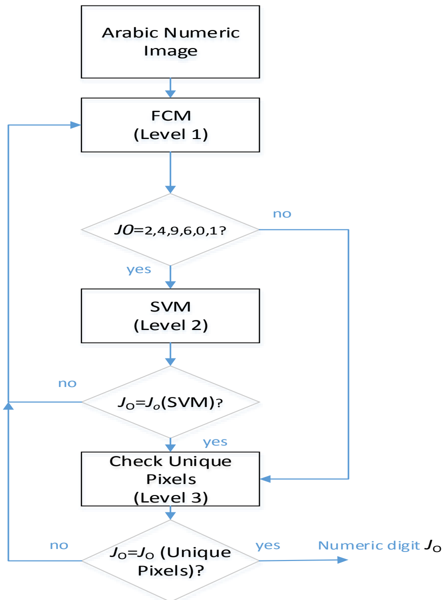

Fig. 3 Flowchart for the Serial Hybrid Classifier (3-level classification) approach

It should be noted that moving from one classifier to another depends on the FCM decision. For example if the FCM classifies the input digit image as 9, then SVM is used to decide whether this digit is 9 or 6. If the SVM decision is 9 then the last level (unique pixels) is used to confirm the previous two classification levels decision 9. If one of the upper levels disagrees with FCM, the next highest membership will be considered and the same process will be carried on. On the other hand, if the decision of FCM is 3, the SVM stage is bypassed and the unique pixels method is used to confirm whether the FCM 3 decision is right. If the unique pixels method decision disagrees with FCM decision, the next highest membership will be considered and the same process will be carried on.

Parallel Hybrid Classification Approach

The parallel hybrid classification approach we present in this paper is based on the principle of Bayesian data fusion. The image of the Indian (Arabic) handwritten numeral is simultaneously fed to an FCM classifier and a feed forward NN. Their decisions are then fused using Bayes rule to obtain the final decision of this hybrid classifier. The Bayesian fusion is explained as follows:

For a certain input numeral the decision i.e. the output class given by the FCM single classifier can be modelled as the probability of the output class given that the classifier is the FCM classifier and is denoted by $p(C_i \mid w_{FCM})$. Correspondingly, the output class given by the NN single classifier can then be modelled as the probability of the output class given that the classifier is the NN classifier and is denoted by $p(C_i \mid w_{NN})$. It follows that the probability of the class C_i given the decision of both classifiers (the class posterior) is denoted by $p(C_i \mid w_{FCM}, w_{NN})$. By Bayes rule:

$$p(C_i \mid w_{FCM}, w_{NN}) = \frac{p(w_{FCM} \mid C_i, w_{NN})\, p(C_i \mid w_{NN})}{p(w_{FCM} \mid w_{NN})} \tag{4}$$

Where C_i denotes the class which can take a value from 0 to 9 and w denotes the type of the classifier. Noting that the decision of the FCM classifier is independent from the decision of the NN classifier, then the following simplifications apply:

$$p(w_{FCM} \mid C_i, w_{NN}) = p(w_{FCM} \mid C_i), \quad p(w_{FCM} \mid w_{NN}) = p(w_{FCM})$$

which reduces (4) to:

$$p(C_i \mid w_{FCM}, w_{NN}) = \frac{p(w_{FCM} \mid C_i)\, p(C_i \mid w_{NN})}{p(w_{FCM})} \tag{5}$$

Expanding the NN related term using Bayes rule:

$$p(w_{FCM} \mid C_i) = \frac{p(C_i \mid w_{FCM})\, p(w_{FCM})}{p(C_i)} \tag{6}$$

Substituting (6) into (5) simplifies (4) to:

$$p(C_i \mid w_{FCM}, w_{NN}) = \frac{p(C_i \mid w_{FCM})\, p(C_i \mid w_{NN})}{p(C_i)} \tag{7}$$

It can be seen from this result that the class posterior is actually equal to the product of the conditional probabilities of the class C_i across the single classifiers, scaled by constant $p(C_i)$. Where $p(C_i)$ serves as a normalization constant to make $p(C_i \mid w_{FCM}, w_{NN})$ a valid probability density function [1, 32]. The class (Decision) with higher probability across the two classifiers is taken as the best classification result $\max_i^N \{p(C_i \mid w_{FCM}, w_{NN})\}$ for the input.

The FCM algorithm will results in ten (first degree) memberships each corresponding to different class i.e. different Indian (Arabic) numeral. We take the degree of membership as the class probability of the FCM decision. Each FCM probability is multiplied with the probability of corresponding class in the NN. The product with higher value represents the final decision of the parallel hybrid classifier.

Evaluation

The algorithms presented in this paper are tested on the CENPARNI Indian (Arabic) handwritten numeral public dataset found here: (http://users.encs. concordia.ca/~kharma/ ExchangeWeb/Databases/ArabicDBases/). This set has 4640 images. The dataset contain 10 folders labeled from 0 to 9. For each folder, 60% of images are used for training and 40% are used for testing. It should be noted here that there are many images in the set that cannot be visually recognized. The FCM classification rates for each digit over the training and testing datasets are shown in table I. The overall classification rates for training and testing dataset are 69.85% and 71.72%, respectively.

The detailed numbers of misclassification of each digit due to the use of FCM over the training set are given in table II. It can be clearly seen that the numeric digits 6, 4, 0 and 1 have the maximum number of misclassifications. It can also be noticed that most of the images of digit 1 are misclassified as 0 and vice versa. Moreover, 122 images of numeric 9 are misclassified as 6.

After applying the SVM and the unique pixel method over the FCM output the number of misclassifications for each digit considerably decreases as shown in table III. The resulting classification rates of our serial hybrid classification approach over the training set and the testing set are shown in table IV with an overall classification rate of 87.46 % over the training data and 88.18 % over the testing data. It should be noted that the lower classifier (FCM) gives a preliminary decision as the confirmation for this decision might come from the later levels, However, for some cases (i.e. 5)

Table I FCM classification rates for each digit over the training set and the test set

	0 (٠)	9 (٩)	8 (٨)	7 (٧)	6 (٦)	5 (٥)	4 (٤)	3 (٣)	2 (٢)	1 (١)	Overall rate
Training	0.78	0.37	0.66	0.67	0.99	0.68	0.97	0.80	0.58	0.46	0.6985
Testing	0.80	0.42	0.78	0.61	0.99	0.76	0.99	0.82	0.60	0.39	0.7172

the final decision comes from FCM. Table V shows the classification levels required for each numeric number to reach to a final decision.

The resulting classification rates of our parallel hybrid classification approach over the training set and the testing set are shown in table VI with an overall classification rate of 91 % over the training data and 89 % over the testing data which shows a little improvement over the results of the serial hybrid classifier.

The same experiment was conducted over several single classifiers namely, feed forward NN, KNN, Classification Trees and Naive Bayes classifier. The test result of the single classifiers are listed in in table VII. It can be seen from the table that the results obtained by single classifiers are less than those obtained by the Hybrid classifiers. It is also clear that the fusion of the NN decision with the FCM decision has resulted in a classification rate that is higher than the classification rate of both the FCM classifier separately and the NN separately. The above experiments show that Hybrid classifiers give better classification rates for the problem of classification of handwritten Indian (Arabic) numerals from images.

Table II FCM Misclassification matrix

		Indian (Arabic) digit in an image									
		0 (٠)	9 (٩)	8 (٨)	7 (٧)	6 (٦)	5 (٥)	4 (٤)	3 (٣)	2 (٢)	1 (١)
Misclassification Result	0 (٠)	0	0	0	3	1	4	2	1	1	31
	9 (٩)	0	0	1	0	122	0	9	1	1	0
	8 (٨)	0	2	0	0	41	0	3	0	0	0
	7 (٧)	0	2	0	0	26	0	14	7	0	0
	6 (٦)	0	0	0	0	0	0	2	0	0	0
	5 (٥)	4	1	9	0	51	0	13	1	2	2
	4 (٤)	1	0	0	1	5	0	0	0	0	0
	3 (٣)	0	2	0	2	28	0	14	0	2	0
	2 (٢)	2	2	0	0	29	1	40	19	0	0
	1 (١)	40	4	1	0	39	0	41	1	4	0
	SUM	47	13	11	6	342	5	138	30	10	33

Table III Serial Hybrid Classifier Misclassification matrix

		Indian (Arabic) digit in an image									
		0 (٠)	9 (٩)	8 (٨)	7 (٧)	6 (٦)	5 (٥)	4 (٤)	3 (٣)	2 (٢)	1 (١)
Misclassification Result	0 (٠)	0	0	0	1	0	4	0	2	1	4
	9 (٩)	7	3	0	0	14	0	7	2	4	0
	8 (٨)	2	11	0	0	2	1	0	21	0	4
	7 (٧)	0	3	0	2	17	0	8	0	10	5
	6 (٦)	1	1	0	1	0	0	0	0	0	1
	5 (٥)	5	0	1	0	0	0	26	4	13	3
	4 (٤)	0	0	0	0	0	0	2	0	0	0
	3 (٣)	0	5	0	0	13	0	9	7	8	6
	2 (٢)	0	1	0	0	0	0	1	1	2	3
	1 (١)	0	0	1	0	0	0	7	1	3	3
	SUM	15	24	2	4	46	5	60	38	41	29

Table IV Serial Hybrid Classifier rates for each digit over the training and test sets

	0 (٠)	9 (٩)	8 (٨)	7 (٧)	6 (٦)	5 (٥)	4 (٤)	3 (٣)	2 (٢)	1 (١)	Overall rate
Training	0.94	0.93	0.94	0.67	0.99	0.80	0.98	0.82	0.82	0.85	0.8746
Testing	0.89	0.96	0.96	0.65	1.00	0.85	0.98	0.87	0.84	0.79	0.8818

Table V The number of classification levels for each Indian (Arabic) number.

Number	FCM	SVM	Unique pixels
0 (٠)	X	X	
9 (٩)	X	X	X
8 (٨)	X		X
7 (٧)	X		X
6 (٦)	X	X	X
5 (٥)	X		
4 (٤)	X	X	
3 (٣)	X		X
2 (٢)	X	X	
1 (١)	X	X	

Table VI Parallel Hybrid Classifier rates for each digit over the training and test sets

	0	9	8	7	6	5	4	3	2	1	Overall rate
Training	0.89	0.96	0.99	0.92	0.97	0.92	0.99	0.95	0.87	0.82	0.91
Testing	0.86	0.92	0.92	0.86	0.95	0.88	0.95	0.92	0.78	0.73	0.89

Table VII Single stage classifier classification rates for each digit over the test set

Classification Method	Result
NN	0.8672
KNN	0.8719
Classification Trees	0.7854
Naive Bayes	0.7854
FCM	0.7172

Conclusion and Future Work

In this paper we have presented two hybrid approaches for the classification of Indian (Arabic) handwritten digits. The first hybrid classifier introduced is serial in the sense that it consist of three consecutive classifiers. The lower level is FCM classifier followed by SVM classifier as second classification level when more details are required and finally the classification is confirmed by the unique pixels method which forms the third classification level. The second hybrid classifier is a parallel one in the sense that it fuses the decision obtained from the FCM classifier with the decision of a Neural Network to obtain the final decision. The overall testing accuracy reported are 88% and 89% for the serial hybrid classifier and the parallel hybrid classifier, respectively.

In future we plan to use other clustering methods such k-means and competitive Neural Networks to compare with the performance of the FCM classifier. We also intend to fuse the output of different classifiers and apply other post processing techniques in order to improve the classification rates.

References

[1] Rami Al-Hmouz, “OCR Based Pixel Fusion,” Journal of Applied Sciences, Vol. 12, No.22, pp. 2319-2325, 2012.

[2] Cheung, A., M. Bennamoun and N.W. Bergmann, 2001,”An Arabic optical character recognition system using recognition-based segmentation,” Pattern Recognition, Vol. 34, pp. 215-233.

[3] Shuang-Tong, T. and L. Wen-Ju, “Number and letter character recognition of vehicle license plate based on edge hausdorff distance,” Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and Technologies, pp: 850-852, Dec, 2005.

[4] Juntanasub, R. and N. Sureerattanan, “Car license plate recognition through Hausdorff distance technique,” Proceedings of the 17th IEEE International Conference on Artificial Intelligence, pp. 647-651, November 14-16, 2005, Hong Kong.

[5] Barrow, H., J.M. Tenenbaum, R.C. Boles and H.C. Wolf, “Parametric correspondence and chamfer matching: two new techniques for image matching,” Proceedings of the 5th International Joint Conference on Artificial Intelligence,. (IJCAI’77), pp.659-663, San Francisco, USA.

[6] Kim, K.K., K.I. Kim, J.B. Kim and H.J. Kim, “Learning-based approach for license plate

recognition," Proceedings of the IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X, Vol. 2, pp: 614-623, December 11-13, 2000,. Sydney, Australia.

[7] Yang, M.N., X.J. LI and X.H. Zhang, "Robust Description Method of SIFT for Features of License Plate Characters," Information Technology Journal, Vol. 10, No. 11, pp: 2189-2195, 2011.

[8] Duan, T.D., Du T.L.H., T.V. Phuoc and N.V. Hoang, "Building an automatic vehicle license-plate recognition system," Proceedings of the International Conference on Computer Science Research, Can Tho, Vietnam,. pp: 59-63, 2005.

[9] Pham, D. L. and Prince, J. L., "Adaptive Fuzzy Segmentation of Magnetic Resonance Images," IEEE Trans. Medical Imaging, Vol. 18, 1999, pp. 737–752.

[10] Yong Yang and Shuying Huang "IMAGE SEGMENTATION BY FUZZY C-MEANS CLUSTERING ALGORITHM WITH A NOVEL PENALTY TERM," Computing and Informatics, Vol. 26, 2007, 17–31.

[11] Bezdek, J.C.,"Pattern Recognition with Fuzzy Objective Function Algorithms," New York: Plenum Press, 1981.

[12] Bezdek, J.C., Hall, L.O. and Clarke, L.P., "Review of MR Image Segmentation Techniques Using Pattern Recognition," Med. Phys., Vol. 20, 1993, pp. 1033–1048.

[13] Ruspini E.H.,"A new approach to clustering," Information and. Control, 15:22-32, 1969.

[14] Dunn J.C, "A fuzzy relative of the ISODATA process and its use in detecting compact Well-Separated clusters," Journal of Cybernetics, Vol. 3, pp: 32-57.

[15] Yang Hong Lei, Peng Jun Huan, Xia Bai Ru and Zhang Ding Xuan, " Remote Sensing Classification Using Fuzzy C-means Clustering with Spatial Constraints Based on Markov Random Field," European Journal of Remote Sensing, Vol. 46, pp: 305-316, 2013.

[16] Hartigan J.A., Wong M.A., " A K-Means clustering Algorithm," Applied Statistics, Vol. 8, No. 1, pp: 100–108, 1979.

[17] G. Dimauro, S. Impedovo, R. Modugno, and G. Pirlo., "A new database for research on bank-check processing", Proc. of the 8th Int. Workshop on Frontiers in Handwriting Recognition, pp 524-528, Canada, 2002.

[18] S. Alma'adeed, D. Elliman, and C.A. Higgins, " A Database for Arabic Handwritten Text Recognition Research", In Proc. Eighth International Workshop of Frontiers in Handwriting Recognition, pp485-589, Canada, 2002.

[19] Y. Al-Ohali, M. Cheriet and C. Suen, " Databases for recognition of handwritten Arabic cheques", Pattern Recognition, Vol. 36, 2003, pp 111-112.

[20] H. Alamri, J. Sadri, N. Nobile, and C. Y. Suen, "A Novel Comprehensive Database for Arabic Off-Line Handwriting Recognition," In Proc. of the II" International Conference on Frontiers in Handwriting Recognition, pp. 664—669, Montreal, Canada, 2008.

[21] Alamri, Huda (2009) *Recognition of off-line arabic handwritten dates and numeral strings.* Masters thesis, Concordia University.

[22] S. Mahmoud, I. Ahmad, W. Al-Khatib, M. Alshayeb,M. T. Parvez, V. Margner and G. A. Fink , " KHATT: An open Arabic offline handwrittentext database", Pattern Recognition, Vol. 47, 2014, pp 1096-1112.

[23] L.M. Lorigo, and V. Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 28, pp. 712-724,2006.

[24] W.M. Pan, T.D. Bui and C.Y. Suen, " Isolated Handwritten Farsi numerals Recognition Using Sparse And Over-Complete Representations", In Proc. 10th international conference on document analysis and recognition, pp. 586-590, 2009.

[25] N. Nobile, C.L. He, M.W. Sagheer, L. Lam and C.Y. Suen, " Digit/Symbol pruning and

Verification for Arabic Handwritten Digit/Symbol Spotting", In Proc. international conference on document analysis and recognition, pp. 648-652, 2011.

[26] A. Harifi and A. Aghagolzadeh, "A New Pattern for Handwritten Persian/Arabic Digit Recognition", In Proc. of World Academy of Science, Engineering and Technology, Vol. 3, pp. 293-296, 2005.

[27] S. M. Awaidah, and S. A. Mahmoud, "A Multiple Feature/Resolution Scheme to Arabic (Indian) Numerals Recognition Using Hidden Markov models," Signal Processing, Vol. 89, pp. 1176-1184, 2009.

[28] L.G. Roberts, "Machine Perception of Three-Dimensional Solids", Optical Electro-Optical Processing of Information, MIT Press, Cambridge, Massachusetts, pp. 159-197, 1965.

[29] J.C. Russ, The Image Processing Handbook, 2nd Edition, CRC Press, Boca Raton, USA, 1995.

[30] J. Sadri, C. Y. Suen, and T. D. Bui., "A New Clustering Method for Improving Plasticity and Stability in Handwritten Character Recognition Systems", Proc. Of Int. Conference on Pattern Recognition, vol.2,pp. 1130– 1133, Hong Kong, 2006.

[31] M. Takruri, R. Al-Hmouz and A. Al-Hmouz, "Three Level Classification: Fuzzy C Means, Support Vector Machine and Unique Pixels for Arabic Handwritten Digits", Proceedings of the World Symposium on Computer Applications & Research (WSCAR), Jan, 2014.

[32] Khushaba, R.N. , Kodagoda, S. , Takruri, M. and Dissanayake, G., "Toward improved control of prosthetic fingers using surfaceelectromyogram (EMG) signals", Expert Systems with Applications, Volume 39, Issue 12, 15 September 2012, pp. 10731-10738