Combining Neural Networks for Arabic Handwriting Recognition

Chergui Leila
Department of Computer Sciences
University Larbi Ben Mhidi
Oum El Bouaghi, Algeria
pgleila@yahoo.fr

Kef Maâmar
Department of Computer Sciences
University Hadj Lakhder
Batna, Algeria
Lm_kef@yahoo.fr

Chikhi Salim
Department of Computer Sciences
University Mentouri
Constantine, Algeria
slchikhim@yahoo.com

Abstract—**Combining classifiers is an approach that has been shown to be useful on numerous occasions when striving for further improvement over the performance of individual classifiers. In this paper we present an off-line Multiple Classifier System (MCS) for Arabic handwriting recognition. The MCS combine two individual recognition systems based on Fuzzy ART network used for the first time in Arabic OCR, and Radial Basis Functions. We use various feature sets based on Hu and Zernike Invariant moments. For deriving the final decision, different combining schemes are applied. The best combination ensemble has a recognition rate of 90,1 %, which is significantly higher than the 84,31% achieved by the best individual classifier. To demonstrate the high performance of the classification system, the results are compared with three research using IFN/ENIT database.**

Keywords-classifier system; Arabic recognition; fuzzy ART network; RBF network; Hu momen; Zernike moment.

I. INTRODUCTION

The research domain of multiple classifier systems (MCS) examines how several classifiers can be applied together to obtain better classification systems. MCS methods may be used to increase the speed of the system or to reduce the time taken for the design of the classification system. There are three basic structures of classifier combination: serial, parallel and conditional. In a serial combination the output of a classifier is the input of the next classifier. Normally a classifier in a serial combination reduces the set of candidate classes so that the final classifier output is only one class. In conditional combination some classifiers are only applied if the output of other classifiers meets a certain condition. In an often used conditional classifier combination the second classifier classifies only the pattern rejected by the first classifier. The second classifier deals with the "difficult" patterns and is normally more complex and slower than the first. In a parallel combination the output of all classifiers is combined in a last step to a single output. This combination is very popular, because each classifier that solves the classification problem may be used (for the serial and conditional combination normally special classifiers are used).

There are many ways to combine in parallel the results of a set of classifiers, depending on the type of the classifiers' output. If the output is only the best ranked class then majority voting can be applied. More sophisticated voting schemes also look at the probability of the classification error for a specific class (Bayesian Combination Rule [14]) and the dependencies between the classifiers (Behavior-Knowledge Space [6]). Some classifiers have a ranked list of classes as output. For them often Borda count [11] or related methods are used. Some classifiers also generate 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. It is also possible to first weigh each classifier according to its individual performance and then apply a combination rule. Another approach is to use the score values output by the individual classifiers as input for a trainable classifier, e.g. a neural-network [3], which acts as the combiner.

In this case study, our research aims at developing an automatic recognition system for off-line Arabic handwritten words using a parallel combination between a Fuzzy Adaptive Resonance Theory and Radial Basis Functions networks, we focus on three combining methods; majority vote, max-rule, and sum-rule. We also applied two types of invariant moments as features.

This paper is organized as follows: Section 2 briefly reviews the Arabic script characteristics. In section 3 related works concerning off-line Arabic handwriting are given. The system recognition phases including pre-processing, feature extraction and classification are detailed in section 4. Experiments and obtained results are presented and discussed in section 5. Finally, we conclude the work and discuss the future directions in the last section.

II. ARABIC CHARACTERISTICS

The recognition of Arabic script has many applications such as mail sorting, bank check reading, and, more recently, the recognition of historical manuscripts. Arabic language provides a rich source of technical challenges for recognition algorithms. The most obvious characteristic of the Arabic language are:

- Arabic script is inherently cursive.
- It is written from right to left.
- The Arabic alphabet has 28 basic letters (figure 1).

978-1-4577-0908-1/11/$26.00 ©2011 IEEE

- The letters can have four different shapes, depending on their position in the word (beginning, middle, end or alone).
- Certain character combinations form new ligature shapes which are often font dependent.
- Many Arabic characters have dots which are positioned at a suitable distance above or below the letter body. Dots can be single, double, or triple.
- Spacing may separate not only words but also certain characters within a word forming sub-words.
- Some Arabic letters may have a zigzag like stroke called Hamza.
- Arabic characters are connected on an imaginary line called baseline.
- Some characters contain closed loops.

Character	Isolated	Initial	Middle	End
Alif	أ	أ	ـا	ـا
Ba'	ب	بـ	ـبـ	ـب
Ta'	ت	تـ	ـتـ	ـت
Tha'	ث	ثـ	ـثـ	ـث
Jeem	ج	جـ	ـجـ	ـج
Ḥa'	ح	حـ	ـحـ	ـح
Kha'	خ	خـ	ـخـ	ـخ
Dal	د	د	ـد	ـد
Thal	ذ	ذ	ـذ	ـذ
Ra'	ر	ر	ـر	ـر
Zy	ز	ز	ـز	ـز
Seen	س	سـ	ـسـ	ـس
Sheen	ش	شـ	ـشـ	ـش
Sad	ص	صـ	ـصـ	ـص
Dhad	ض	ضـ	ـضـ	ـض
T'ah	ط	طـ	ـطـ	ـط
The'ah	ظ	ظـ	ـظـ	ـظ
Ain	ع	عـ	ـعـ	ـع
Gain	غ	غـ	ـغـ	ـغ
Fa	ف	فـ	ـفـ	ـف
Qaf	ق	قـ	ـقـ	ـق
Kaf	ك	كـ	ـكـ	ـك
Lam	ل	لـ	ـلـ	ـل
Meem	م	مـ	ـمـ	ـم
Noon	ن	نـ	ـنـ	ـن
Ha'	ه	هـ	ـهـ	ـه
Waw	و	و	ـو	ـو
Ya	ي	يـ	ـيـ	ـي

Figure 1. Arabic alphabet.

III. RELATED WORKS

Compared to Latin script where a lot of research work is done, the number of work for the combination of Arabic script is quite limited.

One of the first works in this field was given by Farah [8] who introduced a system based on the combination of three Multi Layer Perceptrons for the recognition of Arabic literal amount with a recognition rate of 94% on a small test database containing 4800 words. El-Hajj [7] have used Neural Networks to combine three homogeneous HMM-based classifiers, which have different features as input, they used the IFN/ENIT database achieving a recognition rate of 94,44%. In [17], a strategy for Arabic handwritten word recognition has been proposed by Miled. The idea is based on a sequential hierarchical cooperation of three classifiers, all of a Markovian type. The first classifier is based on a global description of the word using sequential visual indices. The second classifier is associated with an analytic approach that models the characters deprived of their diacritic dots. The third classifier is associated to the sub-word. Different types of combinations methods were tested (measure, rank, hierarchy, class). The rates of recognition of the system exceed 89%. This represents an increase of about 8% with respect to the best performing classifier taken individually. Alma'adeed [2] combined a rule based recognizer with a set of HMMs to recognize words in a bank-check lexicon of 47 words. The rule-based engine used ascenders, descenders, and other structural features to separate the data into groups of words, and an HMM classifier for each group used frame-based features to determine the word. The system was tested on about 4700 words collectively written by 100 writers, excluding about 10% of the words due to errors in baseline detection and pre-processing. A near 60% recognition rate was achieved. Souici, Meslati and Sellami presented in [16] a hybrid approach to the recognition of literal amounts on bank-checks. Three classifiers ran in parallel: neural networks, K-nearest-neighbour, and Fuzzy K-nearest-neighbour. The outputs were combined by word-level score summation. 1200 word by 100 writers were used for training and 3600 words for testing. The recognition rate was 96%. In [4] Burrow applied one K-Nearest Neighbours classification approach to each sub-word, a majority vote is taken on its overall class and repeated for each sub-word sample. First results result in 47% accuracy. By refining the scoring system and adding some features, including the number of dots, the author scores at 74% for sub-word on correctly represented classes.

IV. SYSTEM OVERVIEW

The diagram of the off-line Arabic word recognition system is shown in figure 2.

The recognition process typically consists of three stages: pre-processing, feature extraction, and classification. Pre-processing is to reduce the noise in character image for improving the recognition accuracy. In feature extraction stage, each word is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

For classification of a large category set, a single classifier cannot achieve both high accuracy and high speed. Combining classifiers is chosen here to produce better recognition results.

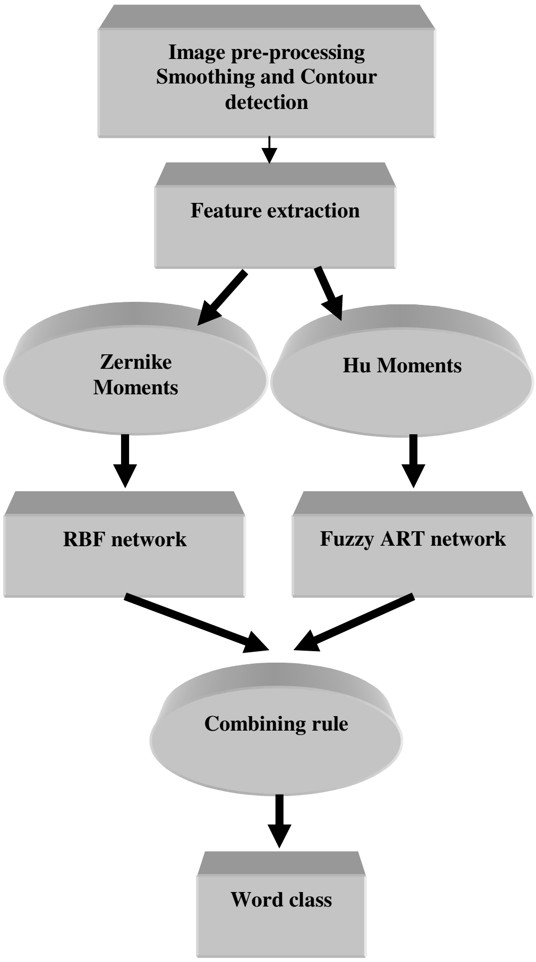

Figure 2. Structure of the proposed recognition system.

A. *Preprocessing*

The original data set is subjected to a number of preliminary processing steps to make it usable by the feature extraction algorithm. In our research we have used: smoothing and contour tracing. Figure 2 and 3 illustrate an example.

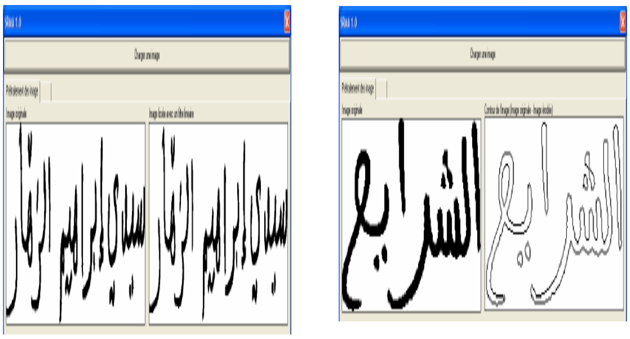

Figure 3. Smoothing and contour detection of word sample.

B. *Feature Extraction*

In our implementation, moment invariants used by Hu [19] have been utilized to build the feature space for Fuzzy ART network. Using nonlinear combinations of geometric moments, Hu derived a set of invariant moments which has the desirable property of being invariant under image translation, scaling and rotation. The central moments, which are invariant under any translation, are defined as:

$$M_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy \, . \qquad (1)$$

Where:

$$\bar{x} = \frac{\overline{M}_{10}}{\overline{M}_{00}}, \qquad \bar{y} = \frac{\overline{M}_{01}}{\overline{M}_{00}} \quad and \qquad . \qquad (2)$$
$$\overline{M}_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy$$

However, for images, the continuous image intensity function f(x,y) is replaced by a matrix, where x and y are the discrete locations of the image pixels. The integrals in equations (1) and (2) are approximated by the summations:

$$M_{pq} = \sum_{x=0}^{m}\sum_{y=0}^{n} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy \, . \qquad (3)$$

$$\overline{M}_{pq} = \sum_{x=0}^{m}\sum_{y=0}^{n} x^p y^q f(x,y)dxdy \, . \qquad (4)$$

Where m and n are dimensions of the image. The set of moment invariants that has been used by Hu are given by:

$$\phi_1 = M_{20} + M_{02} \, . \qquad (5)$$

$$\phi_2 = (M_{20} - M_{02}) + 4M_{11}^2 \, . \qquad (6)$$

$$\phi_3 = (M_{30} - 3M_{12})^2 + (3M_{21} - M_{03})^2 \, . \qquad (7)$$

$$\phi_4 = (M_{30} + M_{12})^2 + (M_{21} + M_{03})^2 \, . \qquad (8)$$

$$\begin{aligned} \phi_5 &= (M_{30} - 3M_{12})(M_{30} + M_{12}) \\ &[(M_{30} + M_{12})^2 - 3(M_{21} + M_{03})^2] + (3M_{12} - M_{03}) \quad . \\ &(M_{21} + M_{03})[3(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] \end{aligned} \qquad (9)$$

$$\phi_6 = (M_{20} - M_{02})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] + 4M_{11}(M_{30} + M_{12})(M_{21} + M_{03}) \quad . \qquad (10)$$

$$\phi_7 = (3M_{21} - M_{03})(M_{30} + M_{12})[(M_{30} + M_{12})^2 - 3(M_{21} + M_{03})^2] + 3(M_{21} - M_{03})(M_{21} + M_{03})^* [3(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2] \quad . \qquad (11)$$

These functions can be normalized to make them invariant under a scale change by using the normalized central moments instead of the central moments. The normalized central moments are defined by:

$$M_{pq} = \frac{M_{pq}}{M_{00}^{a}} \; where \quad a = \frac{(p+q)}{2} + 1 . \qquad (12)$$

These, when substituted into the above equations, will give seven moments which are invariant to translation, scale change and rotation.

The ϕ_s have large dynamic values. Thus, it was found that it was more practical to deal with the logarithms of the absolute value of the ϕ_s [20] thus; the seven moment invariants used in the proposed system are replaced by their logarithmic values. Table 1 shows the rounded values of ϕ obtained for some of the words in the training set.

TABLE I. MOMENT INVARIANT VALUES FOR THREE WORDS

	سيدي الظاهر	بو عثمان	تونس القباضة الأهلية
$\phi1$	1,1510	0,9937	1,2045
$\phi2$	2,2821	1,9367	2,3945
$\phi3$	1,0679	1,1754	1,5419
$\phi4$	0,3176	0,4587	1,5390
$\phi5$	0,5082	1,1193	3,0794
$\phi6$	-0,7184	1,0512	2,7312
$\phi7$	0,9877	1,1312	1,0762

Zernike moments are chosen for RBF classifier; Zernike [5] defined a complete orthogonal set $\{V_{nm}(x, y)\}$ of complex polynomials over the polar coordinate space inside a unit circle $(x^2 + y^2 = 1)$ as follows:

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta} . \qquad (13)$$

Where $j = \sqrt{-1}$, $n \geq 0$, m is a positive or negative integer, $|m| \leq n$, $n - |m|$ is even, ρ is the shortest distance from the origin to (x, y) pixel, θ is the angle between vector ρ and x-axis in counter clockwise direction, and $R_{nm}(\rho)$ is the orthogonal radial polynomial given by:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s . \frac{(n-s)!}{s!\left(\frac{n+|m|}{2} - s\right)!(\frac{n-|m|}{2} - s)!} \rho^{n-2s} . \qquad (14)$$

Note that $R_{n-m}(\rho) = R_{nm}(\rho)$. These polynomials are orthogonal and satisfy the following condition:

$$\int\limits_{x^2+y^2 \leq 1} \int [V_{nm}(x, y)]^* V_{pq}(x, y) \quad dxdy = \frac{\pi}{n+1} \delta_{np} \delta_{mq} . \qquad (15)$$

Where:

$$\delta_{ab} = \begin{cases} 1; if & a = b \\ 0; & otherwise \end{cases} . \qquad (16)$$

Zernike moments are the projection of the image intensity function $f(x, y)$ onto the complex conjugate of the previously defined Zernike polynomial $V_{nm}(\rho, \theta)$, which is defined only over the unit circle :

$$A_{nm} = \frac{n+1}{\pi} \int\limits_{x^2+y^2 \leq 1} \int f(x, y) V_{nm}^{*}(\rho, \theta) dxdy . \qquad (17)$$

For a digital image, Zernike moments are given by:

$$A_{nm} = \frac{n+1}{\pi} \sum_{x} \sum_{y} f(x, y) V_{nm}^{*}(\rho, \theta), x^2 + y^2 \leq 1 . \qquad (18)$$

C. Classification

In this section, we briefly describe the single classifiers based on Radial Basic Function and on Fuzzy ART network.

1) Fuzzy ART network: Adaptive Resonance Theory (ART) is a family of algorithms for incremental unsupervised learning developed by Carpenter and Grossberg [10]. The ART neural network is a typical representative of competitive networks. The Fuzzy ART neural network has two advantages; one is the ability to handle both binary and analogy vectors and the other is faster implementation.

Fuzzy ART is a generalized ART-1 method which is restricted to continuous binary data in the interval of [0,1]. It is similar to many iterative clustering algorithms where each pattern is processed by finding the nearest cluster and then updating that cluster to be closer to the pattern. However, in the framework of Fuzzy ART is a little changed by introducing the concept of resonance so that each case is processed by first finding the nearest cluster seed that resonates with the case, and then updating that cluster seed. Resonance is just a matter of being within a certain threshold of a second similarity measure.

Fuzzy ART takes three input parameters: choice parameter $(\beta > 0)$, vigilance parameter $(0 \le \rho \le 1)$ and learning rate $(0 \le \lambda \le 1)$. The detail of training Fuzzy ART algorithm is given in [5].

2) Radial Basic Function Network: The RBF is a feed-forward network; it has one hidden layer, with each hidden node performing a localized nonlinear function, mostly a Gaussian function. The response values of basis functions are linearly combined by the output nodes. The activation function of output nodes can be either linear or sigmoidal; to facilitate minimum error training by gradient descent and make the output values approximate a posteriori probabilities, we use sigmoid functions in the output layer.

For an RBF network with d input signals, m hidden units, and M output units, assume spherical Gaussian functions for the hidden nodes:

$$\phi_j(x) = \exp\left(-\frac{\|x-\mu_j\|^2}{2\sigma_j^2}\right), \qquad j = 1,\dots,m. \tag{19}$$

Where μ_j is the center Gaussian, and σ_j^2 is the variance. The linear combination of Gaussian functions:

$$\begin{aligned} y_k(x) = \upsilon_k(x) &= \sum_{j=1}^{m} w_{kj}\phi_j(x) + w_{k0} \\ &= \sum_{j=0}^{m} w_{kj}\phi_j(x), \qquad k = 1,\dots,M \end{aligned} \tag{20}$$

Under supervised training of all parameters, the RBF network can achieve a higher classification performance with much fewer hidden nodes. This can compete or even exceed the Multi-Layer Perceptron. The training algorithm is described in [10].

V. EXPERIMENTS

We tested the two classifiers on the IFN/ENIT database [19]. IFN/ENIT was produced by the Institute for Communications Technology at the Technical University of Braunschweig (IFN) and the "Ecole Nationale d'Ingénieurs de Tunis". The total number of binary images of handwritten Tunisian town/village names is 26459. Those names were written by 411 writers, and they were labeled according to 946 name classes.

A. Effect of System Parameters

Concerning Fuzzy ART Network, we study the effect of varying vigilance parameter ρ according to the number of class. Figure 4 shows the number of class detected by diverse vigilance parameters $\rho \in [0,1]$. We observed that if the vigilance used is smaller, the number of class=1. If ρ=0.6, the number of classes achieved is 100. Furthermore, if ρ=0.9, the number of classes found surpass one hundred. We use ρ=0.6 as our benchmark for the following experiments due to its performance found.

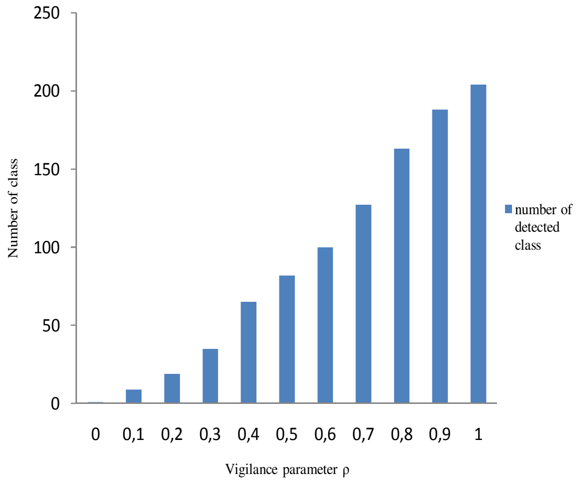

Figure 4. Number of classes detected according to various vigilance parameter.

B. Classifier Combination

We investigate basically three different combination strategies for combining our off-line recognizers: majority vote, max-rule, and sum-rule. Majority vote chooses the class receiving the most votes from all classifiers. The max-rule takes the class with the maximum output value among each classifier, while the sum-rule sums up the output for each class and selects the one with the highest sum.

The plots in figure 5 show the improvement of accuracy rate results by using combining rules. As it is shown, using majority vote rule gives the best recognition rate compared to the other used rules.

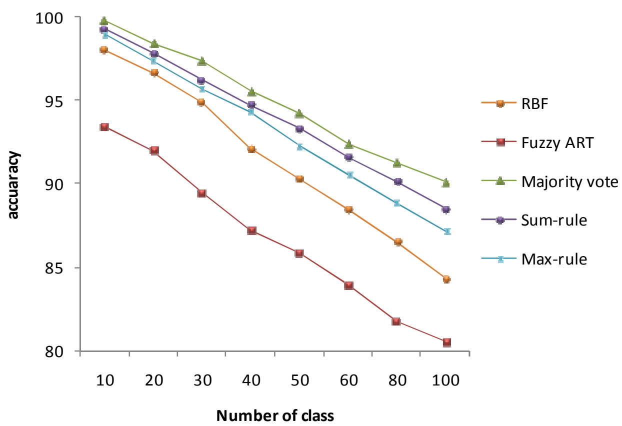

Figure 5. Fuzzy ART, MLP and Fuzzy ART+MLP classifier accuracy.

C. Comparison with Other Systems

The IFN/ENIT database is available to the scientific community and this makes system comparison possible. As mentioned in the abstract, we compared the developed system to three other systems tested also on IFN/EINIT database; the ones of El-hajj [7], Miled [17] and Burrow [4], it may be noted from table 2 that the highest accuracy was obtained by El-Hajj following our system, this is due to the use of a segmentation phase, on the other side our system achieves a good accuracy compared with Burrow's and Miled's systems which prove the performance of combining an RBF with a Fuzzy ART network.

TABLE II. COMPARISON RESULTS

Systems	Classifiers	Combining rule	Accuracy
El-Hajj's system	3 HMMs	Neural Network	94,44%
Miled's system	3 HMMs	Majority vote	89%
Burrow's system	Several K-NN	Majority vote	74%
Our System	RBF+Fuzzy ART	Majority vote	90,1%

VI. CONCLUSION

Handwritten Arabic word recognition is still a very difficult problem. Classifier combination is considered as a promising way for solving such complex pattern recognition problem.

We have shown the efficiency of the presented combining schemes of Fuzzy ART and RBF network for Arabic cursive word recognition. So all the combination approaches are superior to the best individual classifier, and the majority vote outperforms the other classifier combination scheme by 90,1% accuracy rate.

Moment invariants are invariant to translation, rotation and scaling which omit the use of some pre-processing operations like normalization and skew detection and correction.

The performance of Radial Basis Function Neural Networks on classification is always excellent, compared with other networks. Concerning Fuzzy ART network, it does not demand a priori knowledge of the fixed number of necessary classes.

An interesting way to improve the performance of the hybrid RBF-Fuzzy ART recognizer, presented in this paper, consists in adding other kind of classifier to the combining scheme like HMM classifier. Also we can apply structural features.

REFERENCES

[1] N. Azizi, N. Farah, M. Sellami, and A. Ennaji, "Using diversity in classifier set selection for Arabic handwritten recognition," Springer, Multi Classifier System, 2010, pp.235-244.

[2] S. Alma'adeed, C. Higgens, and D. Elliman "Recognition of Off-Line Handwritten Arabic Words Using Hidden Markov Model Approach," *IEEE*, Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02), 2002.

[3] H. Bunke, and A. Kandel, Hybrid Methods in Pattern Recognition. World Scientific, Singapore, 2002.

[4] P. Burrow, Arabic Handwriting Recognition. Report of Master of Science School of Informatics, University of Edinburgh, 2004.

[5] L. Chergui, and M. Benmohammed, "ART network for Arabic handwritten recognition system", in Proc. ACIT: The 9 th International Arab Conference on Information Technology, Tunisia, 2008.

[6] M. Cheriet, N. Kharma, C.L. Liu, and C.Y. Suen, Character Recognition Systems, Willey, USA, 2007.

[7] R. El-Hajj, C. Mokbel, and L. Likforman-Sulem, "Combination of HMM-Based classifiers for the recognition of Arabic handwritten words," Document Analysis and Recognition, 2007.

[8] N. Farah, M.T. Khadir, and M. Sellami, "Artificial neural network fusion: Application to Arabic words recognition," in Proc. European Symposium on Artificial Neural Networks, Bruges, 2005, pp. 151-156.

[9] V. Govindaraju, and S. Sethur, Gide to OCR for Indic Scripts, Document Recognition and Retrieval. Springer, New-York, 2009.

[10] D. Graupe, Principles of Artificial Neural Networks. World Scientific, Singapore, 2007.

[11] M. Haidl, and J. Kittler, Multiple Classifier Systems. Springer, Germany, 2007.

[12] V. Kecman, Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models, The MIT Press, USA, 2001.

[13] Y. Kessentini, T. Burger, and T. Paquet, "Evidential combination of multiple HMM classifiers for multi-script handwriting recognition," Springer, IPMU, 2010, pp. 445-454.

[14] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience. New Jersey. USA, 2004.

[15] M. Liwicki, and H. Bunke, Recognition of Whiteboard Notes Online, Offline and Combination. World Scientific, Singapore, 2008.

[16] L. M. Lorgio, and V. Govindaraju, "Off-line Arabic handwriting recognition: a survey," IEEE on Pattern Analysis and Machine Intelligence, vol.28, no. 5, 2006, pp. 712-724.

[17] H. Miled, C. Olivier, M. Cheriet, and Y. Lecourtier, "Coupling observation/letter for a Markovian modilization applied to the recognition of Arabic handwriting," in Proc. Fourth Int'l Conf. Document Analysis and Recognition, 1997,pp. 580-583.

[18] J.J. Oliveira, M.N. Kapp, C.O. Freitas, J.M. Carvallo, and R. Sabourin, "Handwritten recognition with multiple classifiers for restricted lexicon," IEEE. Symposium on Computer Graphics and Image Processing, 2004, pp. 1530-1534.

[19] M. Pechwitz , S.S. Maddouri, V. Margner, N. Ellouze, and H.Amiri, "IFN/ENIT Database of handwritten Arabic words," in Proc. CIFED, 2002 pp. 129-136.

[20] M. Sarfraz, Computer Aided, Intelligent Recognition, Techniques and Applications. John Wiley & Sons, England, 2005.

[21] S. Theodoidis, and K. Kontrommbas, Pattern Recognition, Elsevier, San Diego, USA, 2003.

[22] Q. Wang, and S. Ju, "A mixed classifier based on combination of HMM and KNN," IEEE Trans. Fourth International Conference on Natural Computation, vol. 4, July 2008, pp. 32-42.