**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 3, No. 1, July 2016, pp. 235 ~ 239
DOI: 10.11591/ijeecs.v3.i1.pp235-239


# Tifinagh Characters Recognition Using Simple Geometric Shapes

**M Boutaounte, Y Ouadid**
Faculty of Sciences and Technology, Beni-Mellal, Morocco
Corresponding autor, e-mail: boutaounte.mehdi@gmail.com

***Abstract***

*In this paper we present method of features extraction based on decomposition of the characters into several simple geometric shapes (segment, arc) by detecting the branch points and end points, as it explained follow a new methods are using to treated the obtained information in order to decide if the characters need to add more key points also in this step we extracte the type of the shapes (segment or arc) and the orientation. The next step of characters recognition different methods are used such as neural network (NN), K-mean and support vector machine (SVM) classifier. The results shown in this paper are obtained using the IRCAM database.*

***Keywords:*** *character recognition, neural network, K-mean, SVM, features extraction,*



## 1. Introduction

Optical character recognition (OCR) is the most leading field of research treated around the world, but all this research target languages commonly used such as Latin and arebic languages and ignores other as the Tifinagh which is a Hamito-Semitic language derived from ancient Berber. It includes a variety of dialects present from Morocco to Egypt, passing through Algeria, Tunisia, Mali, Niger and Libya. Those characters present a challenge to research in order to create an OCR system that give which gives satisfactory results due to the similarity in shape of the charaters, and the processing time caused by the high number of characters (33 characters).

In this paper, we start by presenting a review of some proposed Tifinagh character recognition systems such as, Essaady, et al., [1] proposed an analytical method in used finite automaton for recognition and Freeman coding as feature extraction method. Bencharef and al [2] suggest the Riemannian metric as a feature these metrics descriptors are reliable toward change scale. El Ayachi, et al., [3, 4], shows two methods of recognition both of them use the Multilayer Neuron Networks and Dynamic Programming for the classification, Moments invariants & Walsh transform for feature extraction. In their work neural networks combined with Walsh transform showed fairly interesting results. Oujaoura, et al., [5] proposed a Tifinagh character recognition system based on Bayesian Networks classification and combination of three features extractor GIST descriptors, texture and Walsh but have some problems in term of CPU time and recognition rate. Amrouch, et al., [6] where they use continuous Hidden Markov Models and directional feature vector sequence. These models are not very discrimination since each hidden Markov models uses the learning of a single character.

In order to resolve these problems of characters recognition in general and not only for tifinagh case, some works have proposed to combine descriptors such as Oujaoura, et al., [7] and Moudni, et al., [8] other are using Fuzzy pattern recognition Zhiyi, et al., [9], in which each characters mainly classified by direct methods according to maximum membership maximum membership that pose a problem in terms of processing time due to the size of the descriptor vector. For this reason, we propose a structural representation of characters, not only based on point feature extraction like Lei Guo, et al., [10] but also describes the types of characters geomitrique components and its orientation extracting using a new approch based in analysing the angle between key points. The majority of work related to Tifinagh characters are using neural network as classifier such as Khadija, et al., [11] but in this work more than one classifier are used to prove the performance of this system of recognition.

*Received February 2, 2016; Revised May 25, 2016; Accepted June 10, 2016*

This paper is organized as follows: Section 2 presents two techniques for features extraction. The SVM is presented in section 3 the Experimental results and discussions are in section 4 then the conclusion in section 5.

## 2. Features Extraction

All images in database are pre-treated using techniques like normalization and thinning. In normalization step the unnecessary area is removed, for the thinning the algorithm of Zhang and Wang [12] was adopted due to its robustness and speed. This is a parallel algorithm in a single iteration that produces perfectly skeletons 8-connected and which operates the collisions (Figure 1).

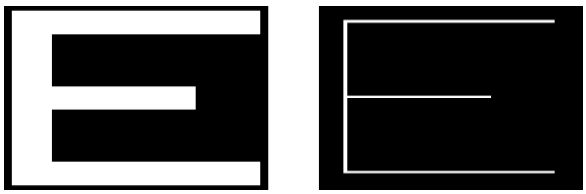

Figure 1. Skeletonized Image Character

The proposed recognition system (Figure 2) consists of three main axes, the pre-processing step, the isolated character recognition block and the charactere recognition block. In other words the preprocessing step, feature extraction and the recognition.

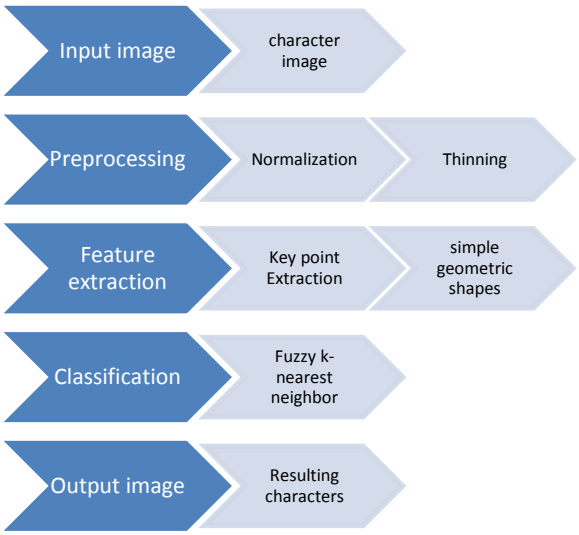


Figure 2. Proposed Recognition System

The objective of this work is to represent the skeleton of the characters by key points, which will be transformed later in geometric simple shape such as lines or arcs. We started by extracting primary key points which are “branch” and “end” points as shown in Figure 3, The next step is to find secondary key points in which there is a direction change by using a new technique based in calculation of the angle between axes and extracted segment. This step requires the extraction of segment and arc between detected points to make the decision of adding a secondary point or not.

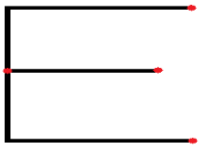

Figure 3. Branch and End Points

For each connected two key points we draw a segment A that connects them, then the distance between this segment A and the black pixels that connect the two points in image is calculated. If the distance is equal to 0 that mean that the geometric shaps between the two points is a simple segment, if not we have two case it can be an arc or more than one connected segment (Figure 4).

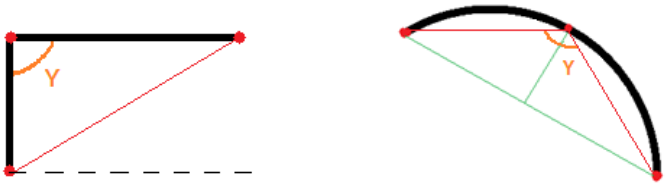


Figure 4. Determination of the Y Angle

To resolve this issue, we introduce an angle Y as shown in Figure 4, and we set a threshold for this angle to determine if it is an arc or not.

An other parameter that we introduce to detect the orientation of every segment and arc is the X angle between the segment A and the horizontal axes, shown in Figure 5.

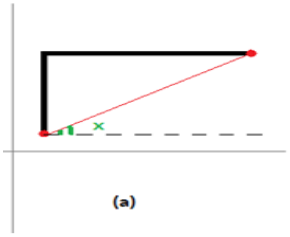


Figure 5. Determination of the X Angle

The figure below shows the process from the beginning up to extracting the key points and the simple geometric shapes (Figure 6).

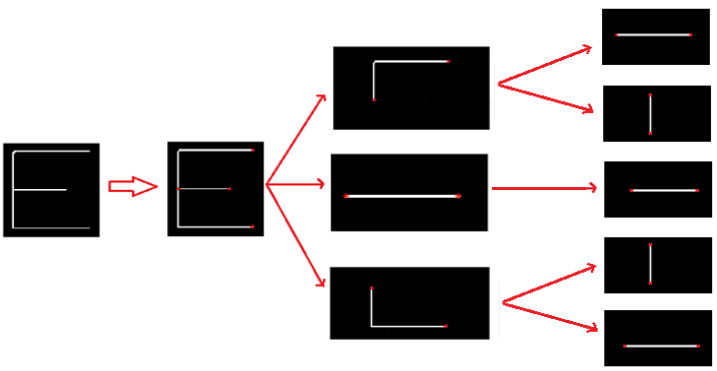

Figure 6. Extracting Key Points and Simple Geometric Shaps

The information collected in this step will represente our feature vector as follows:

1. The number of key points for each type ( end , branch and inflection)
2. The number of simple geometric shaps detected classifie by type and orientation.

## 3. Classifier Learning and Classification

### 3.1. Support Vector Machine

SVM is a technique motivated by statistical learning theory and has been successful applied to numerous classification tasks. The key idea is to separate two classes with a decision surface that has maximum margin (Figure 7) [13].

However using of SVMs is still limited to a small number of researchers. A possible reason is that the training algorithms for SVMs been slow, especially for major problems. Another explanation is that the SVM training algorithms are complex, subtle and hard to engineer a way to implement.

Let $P_i$, be a set of training vectors where i=1,2…,N with corresponding binary labels $S_i = 1$for the positive ans 0 for not . In classification an SVM assigns a label S to a test vector T by evaluating (1).

$$f(T) = \sum_i \propto_i S_i K(T, P_i) + b \quad (1)$$

The weights αi and the bias b are SVM parameters and adopted during training by maximizing [14] for this work the proposed method is given by grouping the training data in terms of classes and dividing them training data of each clas into two parts 70% for training and the rest for validation. Useing discriminant function with a low computation cost for each class. (In our case the Euclidean discriminant function is used) we evaluate the performance of discriminant functions by using an N*M matrix in where each row corresponds to a class whose patterns can be classified to.

### 3.2. Experimental Results and Discussions

The OCR system was subjected to different set of input characters images, in order to determine its recognition efficiency. The IRCAM Database composed of 3300 images, 100 by characters is used. For each character we extract the feature as it is clarify in the previous section .In this experience, 2000 character images are used for training and the rest to testing the performance method. Table 1 shows the obtained results using two types of features, first the key points that give a rate of 80% secondly the key points and the simple geometric shapes that constitute the character by using it the recongition rate rise to 98%. The key points extracted and the connexion between it used in detection of shapes type can be used also in graph theory for graph creation.

Table 1. Recognition Accuracy of SVM

| | | Recognation rate % | |
|---|---|---|---|
| | feature type | Key points | Key points and shapes |
| Adopted Method | Training data | 2000 | 2000 |
| | Test data | 1300 | 1300 |
| | Recognition rate | 82,50 | 98,94 |
| | Erreur rate | 17,50 | 1,06 |

Using other classifier such as neural network give a recognition rate between 75% and 95% depend on type of features used it was only key points or key pints and shapes but using the SVM we have reached the rate of 98%. By againstmost of existing work using the description as gist, surfing which is a very large size requires more time in the extraction of features and learning phase.

The algorithm presented here achieves high identification rate, as well as low false alarm rates. On the training set, each characters is represented by only 100 images in database a percentage of 70% it give only 70 image for training for each characters, which is not

sufficient to achieve high performance when SVM is used. Wich force us to look for a new database with more elements for high performance.

It was observed that the developed Yoruba OCR system's performance unit is quick in extraction of features and good recognition rate that can be increased by using a large data base.

## 4. Conclusion

In this work we presented a new technique for Tifinagh character recognition. In the features extraction stage we used the key points to extract geometric shapes in each character and detect their orientation. In addition, this method gives right sets with a small number of classes for treaning using the SVM classifier.

The obtained resultat confirms that we have implemented an efficient method for features extraction that can be used with other classifier such as neural network, k-nearest neighbor …, to reduce the treaning time that SVM method required and also we can use it for building graph and use for recognition the methods of graph matching. Considering the fact that it is the first time we use this feature whit SVM classifier applied in Tifinagh characters, the results are encouraging.

## References

[1] Es Saady Y, Rachidi A, El Yassa M, Mammass D. Printed Amazighe Character Recognition by a Syntactic Approach using Finite Automata. *Journal of Graphics, Vision and Image Processing*. 2010; 10(2): 1-8.
[2] Bencharef O, Fakir M, Idrissi N, Bouikhalen B, Minaoui B. *Application de la géométrie riemannienne à la reconnaissance des caractères Tifinagh*. Third international symposium on Automatic Amazigh processing (SITACAM'11). Agadir, Morocco. 2011.
[3] El Ayachi R, Fakir M, Bouikhalene B, Safi S. Offline printed amazighe scripts recognition. *Journal of Theoretical and Applied Information Technology*. 2010; 20(2).
[4] El Ayachi R, Fakir M, Bouikhalene B. Recognition of Tifinagh Characters Using Dynamic Programming & Neural Network. *Book of Document Recognition and Understanding (InTech)*. 2011: 35-57.
[5] Oujaoura M, Minaoui B, Fakir M. Walsh, Texture and GIST Descriptors with Bayesian Networks for Recognition of Tifinagh Characters. *International Journal of Computer Applications*. 2013; 81(12): 39-46.
[6] Amrouch M, Es-saady Y, Rachidi A, El Yassa M, Mammass D. Handwritten Amazighe Character Recognition System Based on Continuous HMMs and Directional Features. *International Journal of Modern Engineering Research (IJMER)*. 2012; 2(2): 436-441.
[7] Oujaoura M, El Ayachi R, Minaoui B, Fakir M, Bouikhalene B, Bencharef O. *Invariant Descriptors and Classifiers Combination for Recognition of Isolated Printed Tifinagh Characters*. Third international symposium on Automatic Amazigh processing (SITACAM' 13). Beni-Mellal, Morocco. 2013.
[8] Moudni H, Er-rouidi M, Oujaoura M, Bencharef O. *Recognition of Amazigh characters using SURF & GIST descriptors.* Third international symposium on Automatic Amazigh processing (SITACAM' 13). Beni-Mellal, Morocco. 2013.
[9] Zhiyi Ruan, Ying Zou, Dongming Hong, Lurong Wu. Visualization of License Plate Recognition System. *Indonesian Journal of Electrical Engineering and Computer Science (TELKOMNIKA).* 2013; 11(11): 6714-6721.
[10] Lei Guo. Characters Feature Extraction based on Neat Oracle Bone Rubbings. *Indonesian Journal of Electrical Engineering and Computer Science (TELKOMNIKA).* 2013; 11(9): 5427-5434.
[11] Khadija El Gajoui, Fadoua Ataa Allah. *Optical character recognition for multilingual documents: Amazigh-French.* Second World Conference on Agadir Complex Systems (WCCS). Agadir, Morocco. 2014.
[12] Zhang YY, Wang PSP. *A Modified Parallel Thinning Algorithm*. 9th International Conference on Pattern Recognition. Rome, Italy. 1988: 1023-1025.
[13] Vapnik V, Vladimir N. The Nature of Statistical Learning Theory. New York: Springer-Verlag. 1995.
[14] Bahlmann C, Haasdonk B, Burkhardt H. *On-line Handwriting Recognition with Support Vector Machines A Kernel Approach*. Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02). Ontario, Canada. 2002.