

Evaluation of Features Extraction and Classification Techniques for Offline Handwritten Tifinagh Recognition

Mustapha Amrouch

Received: 11 December 2015 Accepted: 5 January 2016 Published: 15 January 2016

Abstract

This paper presents a review on different features extraction and classification methods for off-line handwritten Amazigh characters (called Tifinagh) recognition. The features extraction methods are discussed based on Statistical, Structural, Global transformation and moments. Although a number of techniques are available for feature extraction and classification, but the choice of an excellent technique decides the degree of accuracy of recognition. A series of experiments were performed on AMHCD database allowing to evaluate the effectiveness of different techniques of extraction features based on Hidden Markov models, Neural network and Support vector Machine classifiers. The statistical techniques give encouraging results.

Index terms— handwritten recognition, tifinagh characters, extraction features (statistical, structural and global transformation), classification (HMM, MLP, SVM).

1 I. INTRODUCTION

Feature extraction in handwriting recognition is a very important field of image processing and object recognition. Fundamental component of characters are called features. The basic task of feature extraction and selection is to find out a group of the most effective features for classification; that is, compressing from high-dimensional feature space to low-dimensional feature space, so as to design classifier effectively.

Due to the nature of handwriting with its high degree of variability and imprecision obtaining these features, is a difficult task. Feature extraction methods are based on 3 types of features [1]: continuous signal contains more information that can be represented for the purpose of classification [4].

In this paper our study was conducted to evaluate and examine the main approaches classes of extraction features on the Tifinagh script.

The majority of characters of this script are formed by loops, lines and curves (figure ??), this make it difficult to describe and sensitive to noise, the main problem is how to extract features. This may be solved by the selection of the useful primitives customarily defined in the automatic character recognition.

Figure ??: Some Tifinagh characters from AMHCD database [5] Recently the recognition of handwritten Tifinagh characters (Figure ??) is the subject of several researches. These studies have been published in the literature. Among these researches, we find ([6] [7] [??8] [9] [10] [11]).

All of previous cited works used a particular type of extraction features technique. [6] and [10] used the invariant moments as pattern sensitive features in classification and recognition. [7] and [8] used statistical techniques by applying respectively zoning method and Freeman code to form the vector characteristics, whereas [9] used the Hough transform and the extracted features are structural based on the horizontal and vertical centreline of the letter in [11].

To evaluate the efficiency and the relevance of each type of extracted features we have used several methods of classification (Neural Networks, Hidden Markov Model and Support Vector Machine) for the recognition of Tifinagh characters.

The remainder of this paper is organized as follows. Section (2) presents the multiple techniques used to extract features from an image of Tifinagh letter after the preprocessing step. Section (3) is focused on the classification

step. In section (4) we present the experimental results of several techniques used. The paper finally concludes with an analysis of the results and an introduction of future work.

2 EXTRACTION FEATURES

After a number of preprocessing operations such as binarization, noise reduction, skeletonization and normalization, a feature extraction method is applied to extract the most relevant characteristic of the character to recognize. The performance of a character recognition system largely depends on the quality and the relevance of the extracted features.

Features of a character can be classified into three main classes: Statistical features, Structural or topological features and Global transformations

3 a) Statistical Features

Statistical features are obtained from the arrangement of points constituting the character matrix. These features can be easily detected as compared to topological features. A number of techniques are used for feature extraction; some of these techniques used in this work are:

Zoning According to this technique the character matrix is divided into small portions or zones (figure 2(a)). The densities of pixels in each zone are calculated and used as features; more details about zoning methods for handwritten character recognition are given in [14].

4 Diagonal based

Diagonal features extraction [15][16] scheme for recognizing offline handwritten characters is proposed in this work. Every character image of size 100x100 is divided into 100 equal zones, each of size 10x10 pixels (figure 2(b)). The features are extracted from each zone pixels by moving along the diagonals of its respective 10x10 pixels. Each zone has 19 diagonal lines and the foreground pixels present along each diagonal line is summed to get a single sub-feature, thus 19 subfeatures are obtained from each zone, and then are averaged to form a single feature value placed in the corresponding zone (figure ?? (a)). This procedure is sequentially repeated for all zones. There could be some zones whose diagonals are empty of foreground pixels, the feature value corresponding to these zone are zero. Finally 100 features are extracted for each character figure ??(c). Structural features are based on topological and geometrical properties of the character, such as aspect ratio, cross points, loops, branch points, strokes and their directions, inflection between two points, horizontal curves at top or bottom, etc.

In this study we used the Geometric features technique proposed in [17], this technique extracts the geometric features of the character contour. These features are based on the basic line types that form the character skeleton.

The image is divided into windows of equal size, and the feature is done on individual windows, for the system implemented, the image was zoned into equal sized windows.

To extract different line segments in a particular zone, the entire skeleton in that zone should be traversed. For this purpose, certain pixels in the character skeleton were defined as starters, intersections and minor starters (figure 3). After zonal feature extraction, certain features were extracted for the entire image based on the regional properties namely: Euler Number, Regional area, Eccentricity.

5 Global Journal of Computer Science and Technology

Volume XVI Issue V Version I ii.

6 a) Moment and Global Transformations

The global transformations are generally widely used previously in the signal processing field. Their goals are change the image representation space(character or word)to facilitate the extraction of relevant features. There are many techniques used in handwritten recognition, in this work we have choice the Zernik moments and Gabor filter.

i. **Zernike Moments** Zernike moments are used in pattern recognition applications as invariant descriptors of the image shape. They have been proven to be superior to moment functions such as geometric moments in terms of their feature representation capabilities and robustness in the presence of image quantization error and noise [18] They provide a compact way of describing an object's overall shape using a small set of values.

ii. **Gabor Filter** Tifnagh character image features are extracted in this part using Gabor filters which can be written as a two dimensional Gabor function $g(x, y)$, its Fourier transform $G(u, v)$ as given in Equations below [20]:(1) (2)

Where σ_x, σ_y are the variances of x and y along the x, y axis, respectively; $u = \frac{1}{2} \sigma_x^2$ and $v = \frac{1}{2} \sigma_y^2$.

After filtering the given input image, statistical features such as the mean and the variance of the image are computed. The extracted feature vector is constructed from the means and variances of all filtered images.

The Gabor filters are applied using the different orientations and scales. The mean μ and the standard deviation σ for each filtered image are then computed to form the character feature vector.

7 III. CLASSIFICATION

Classification is the process of assigning the sensed data to their corresponding class with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. Some classification techniques used in this work are:

8 IV. EXPERIMENTAL RESULTS

A series of experiments have been performed to evaluate the effectiveness of different techniques of extraction features and classification. These experiments were performed on database of isolated Amazigh handwritten characters (AMHCD), 4200 character images from the portion of AMHCD were used in our experiment, 3100 character images were used for training and 930 character images were used to test identification performance. ? Hidden Markov Model :are a powerful tool frequently used in handwritten text recognition [22][23], and also in other fields related to pattern recognition and computational linguistics, like speech recognition, machine translation, Parts-Of-Speech tagging and information retrieval. ? Support Vector Machines:Support Vector Machines (SVMs) are a set of related supervised learning methods which can be used for both classification and regression [24]. The table shows the comparison of recognition rates between statistical, geometric, global transformations and Moments methods for extraction features using three divers classifier; NN,HMM and SVM.

9 The table below shows the experimental results of the different techniques of extraction features and classification's methods:

As can be seen in table above, the results of recognition rate are varied according to extraction features technique used.

If we compared the results, we find that discrimination capability of statistical methods is better, whereas the Gabor filter and Zernike moments which are invariant to translation and rotation are limited for selection the pertinent features due to the similarity of Tifinagh characters (e.g. and , and).Structural technique gives an important resultsopening the way to a set of combination of statistic and geometric methods to integrate both the peculiarities of the text and the pixel distribution characteristics in the character image.

After analysing the result files that describe the target and actual outputs, we found that for some particular characters, the classification rate is poor. It can be explained that feature extraction techniques are influenced by many factors such as the variations of characters, the order of the strokes always different for different writers. Also, the form of the strokes can be varied. For example, the straight strokes can be curved as bows. Also the similarity of characters influence clearly the results, some characters were easily recognized as other particular characters such as the and , and , and . On other hand, the results are influenced mainly by the classifier performance, it is observed that recognition rate using HMMs are low compared to SVM and NN due the major problem of HMMs which is the estimation of emission probabilities, this confirms that HMMs are powerful to model sequences but still limited compared to NN and SVM in classification. To improve the results obtained using HMMs, it is recommended to use a hybrid classifier.

10 V. CONCLUSION

Feature extraction is an important phase in text recognition systems and for many pattern recognition problems.

In this paper, we have evaluated the feature extraction techniques for offline character recognition of Tifinagh script using various classifiers, the best recognition rate was achieved using statistical techniques. We noticed that the success rate of any recognition system depends not only on the features extraction but it depends on several reasons such as the recognizer technique, the pre-processing stage.

The work done is a first step for several perspectives. We try to improve the recognition rate by combining several classes of features to give a more general description of the character and classification techniques for a better representation and the speed of the system. We try to extend the approach to therecognition of words, sentences and texts and to other scripts, then exploit the results to develop a contextual recognition system. ¹

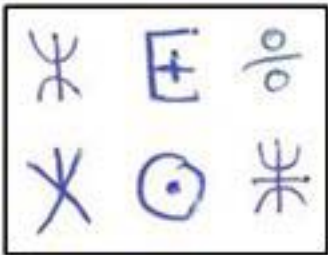


Figure 1:

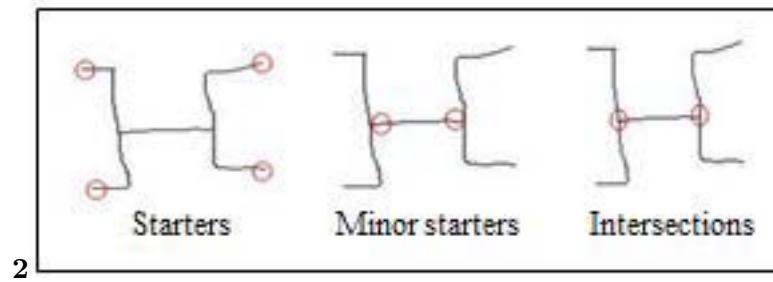


Figure 2: Figure 2 :

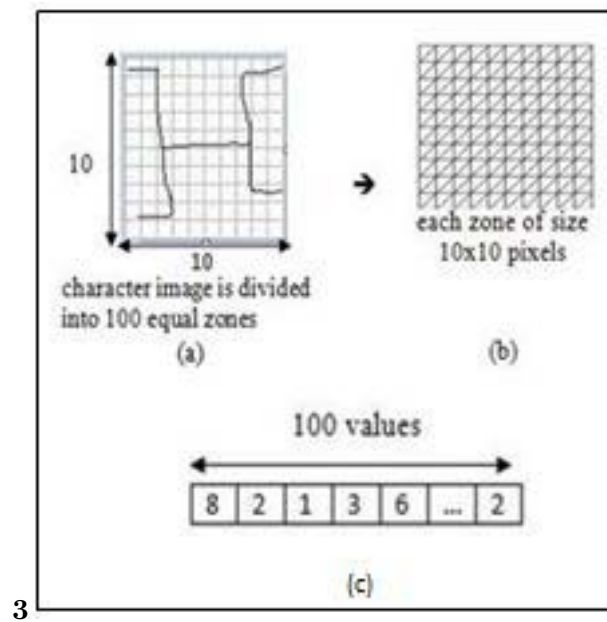


Figure 3: Figure 3 :

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right]$$

Figure 4:

1

[Note: T.R: Training Rate; R.R: Recognition Rate 39 Year 2016]

Figure 5: Table 1 :

-
- [Qacimya and Hammouch] , B El Qacimya , Hammouch .
- [Grosicki et al. ()] , E Grosicki , H El-Abed , « Icdar . 2011. 2011. French Handwriting Recognition Competition », ICDAR.
- [Angadi et al. (2015)] , S A Angadi , H Sharanabasavaraj , "structural Features For Recogni-Tion Of Hand Written Kannada Chara-Cter Based On Angadi , Svm . *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)* April 2015. 5 (2) .
- [Chong et al. ()] 'A comparative analysis of algorithms for fast computation of Zernike moments'. Chee-Way Chong , P Raveendran , R Mukundan . *Pattern Recognition Journal* 2003. 36 p. .
- [Zhang and Suen ()] 'A fast parallel Algorithm for Thinning Digital Patterns'. T Y Zhang , C Y Suen . *Image processing and computer vision* 1984.
- [Kerroum ()] 'A review of feature extraction techniques for hand written Arabic text recognition'. M A Kerroum . *Electrical and Information Technologies (ICEIT)*, 2015.
- [Saady et al. (2011)] 'AMHCD: A Database for Amazigh Handwritten Character Recognition Research'. Y Es Saady , Ali Rachidi , Mostafa El Yassa , Driss Mammass . *International Journal of Computer Applications* August 2011. 27 (4) p. 44.
- [Arora1 et al. ()] *Application of Statistical Features in Handwritten Devnagari Character Recognition*, S Arora1 , D Bhattacharjee2 , M Nasipuri2 , D K Basu2 , M Kundu2 . 2010.
- [Al and Mahmoud ()] *Arabic Character Recognition Using Gabor Filters*, Hamdi Al , -Jamimi , Sabri Mahmoud . 2010. Dhahran, Saudi Arabia. Information and Computer Science, King Fahd University of Petroleum and Minerals
- [Haidar and Fakir] *BENCHAREF Hybridation des modèles de Markov cachés et de la logique floue pour la reconnaissance des caractères Tifinagh manuscrits*, A Haidar , M Fakir , O . (5ème conférence internationale sur les TIC pour l'amazighe 2012)
- [Pradeep et al. (2010)] 'Diagonal Feature Extraction Based Handwritten Character System Using Neural Network'. J Pradeep , E Srinivasan , S Himavathi . *International Journal of Computer Applications* October 2010. 8 (9) p. .
- [Dinesh Dileep A feature extraction technique based on character geometry for character recognition ()] *Dinesh Dileep A feature extraction technique based on character geometry for character recognition*, 2012.
- [Behrouz et al. ()] 'Farsi Handwritten Word Recognition Using Discrete HMM and Self-Organizing Feature Map'. Behrouz , Vaseghi1 , Somayeh , Hashemi . *International Congress on Informatics, Environment, Energy and Applications-IEEA 2012 IPCSIT*, (Singapore) 2012. 2012. 2012. IACSIT Press. 38.
- [El] *Fegh Handwritten Arabic Words Recognition using Multi Layer Perceptron and Zernik Moments*, I El .
- [César De ()] *Handwriting Recognition Revisited: Kernel Support Vector Machines*, Souza César De . 2012.
- [Kessab et al. ()] *Handwriting Moroccan regions recogniti on using Tifinagh character*, B El Kessab , C Daoui , B Bouikhalene , R Salouan . 2015.
- [Das ()] 'Handwritten Arabic Numeral Recognition using a Multi Layer Perceptron'. Nibaran Das . *Computer Science and Engineering Department* 2006.
- [Hirwan and Gonnade (2014)] 'Handwritten Character Recognition System Using Neural Network'. A Hirwan , S Gonnade . *International Journal of Advance Research in Computer Science and Management Studies* February 2014. 2 (2) .
- [Mane and Ragha ()] Vanita Mane , Leena Ragha . *Handwritten Character Recognition using Elastic Matching and PCA" International Conference on Advances in Computing, Communication and Control*, 2009. ACM. p. .
- [Mustapha AMROUCH Reconnaissance des caractères imprimés et manuscrits, textes et documents basés sur les modèles de Markov cachés, 2012. (Thèse de doctorat)
- [Abaynarh and Zenkour ()] 'Offline Handwritten Characters Recognition Using Moments Features and Neural Networks'. Mohamed Abaynarh , Lahbib Zenkour . *Computer Technology and Application* 2015. 6.
- [Rachid El Ayachi, Mohamed Fakir and Belaid Bouikhalene "Recognition of Tifinaghe Characters Using Dynamic Programming
- 'Rachid El Ayachi, Mohamed Fakir and Belaid Bouikhalene "Recognition of Tifinaghe Characters Using Dynamic Programming & Neural Net work'. *Global Journal of Computer Science and Technology* 2011. 10.
- [Alkhateeb et al. ()] 'Word-based Handwritten Arabic Scripts Recognition using DCT Features and Neural network Classifier'. J H Alkhateeb , R Jinchang , J Jianmin , S S Ipson , H El-Abed . *5th International Multi-Conference on Systems, Signals and Devices*, 2008. p. .

- 200 [Youssef Es Saady Amazigh Handwritten Character Recognition based on Horizontal and Vertical ()] *Youssef*
201 *Es Saady Amazigh Handwritten Character Recognition based on Horizontal and Vertical*, 2011. Centerline of
202 Character.
- 203 [Impedovo N and Pirlo (2014)] ‘Zoning methods for handwritten character recognition: A survey’. D Impedovo
204 N , G Pirlo . *Pages 969-981, Handwriting Recognition and other PR Applications*, March 2014. 47.