

Random Forest for video Text Amazigh

Youssef Rachidi

Laboratoire Image et Reconnaissance de Formes – Systèmes Intelligents et Communicants
(IRF – SIC), Université Ibn Zohr Agadir, Maroc

y.rachidi@uiz.ac.ma

Abstract— In this paper; we introduce a system of automatic recognition of Video Text Amazigh based on the Random Forest. After doing some pretreatments on the video and picture, the text is segmented into lines and then into characters. In the stage of characteristics extraction, we are representing the input data into the vector of primitives. These characteristics are linked to pixels' densities and they are extracted on binary pictures. In the classification stage, we examine four classification methods with two different classifiers types namely the convolutional neural network (CNN) and the Random Forest method. We carried out the experiments with a database containing 3300 samples collected from different writers. The experimental results show that our proposed OCR system is very efficient and provides good recognition accuracy rate of handwriting characters images acquired via Video camera phone.

Key-Words: Pretreatments, Video Text Amazigh, Mobile phone, OCR, CNN, Random Forest

I. INTRODUCTION

The automatic recognition of handwritten or printed Amazigh characters remains a subject of research and experimentation. The problem is not yet solved despite the fact that results have reached fairly high rates in some applications [1]. Some attempts have been done to improve the current situation [1]. In this context, we have employed a recognition system of Amazigh handwritten characters extracted from video taken by camera phone [2]. Indeed, in the primitives' extraction stage, our approach is based on primitives of the Zoning types [3], of distance profile feature [4], Projection histogram and Gray Level Co-occurrence Matrix (GLCM) technique [5]. These primitives will supply a Convolutional Neural Networks and Random Forest Method in the learning and recognizing phases. Video text handwritten Amazigh, segmented and isolated characters acquired by camera phone, obtained an encouraging results on the majority of this characters.

Habitually, the phases form the structures of handwriting recognition system are: Pre-processing, Segmentation, Feature extraction, Classification and Post-processing [2].

In this paper, our objective is mainly interested in the development of Video Text handwriting Amazigh recognition system and Improvement of the Recognition Rate by CNN in some characteristics extraction, in which the images from video.

The paper is organized as follows. In section II, the proposed the pre-processing and gives descriptions of the methods that we used throughout the OCR process, which includes the following stages: Binarization, Noise removing, skew detection and correction and Segmentation. The feature extraction procedure adopted in the system is detailed in the section III. Section IV describes the classification and recognition using CNN and Random Forest. Section V presents the experimental results and comparative analysis. Finally, the paper is concluded in section VI.

II. PRE-PROCESSING

The procedure of preprocessing which refines the scanned input image from video includes several steps: Binarization, for transforming gray-scale images in to black and white images, noises removal, and skew correction performed to align the input paper document with the coordinate system of the scanner and segmentation into isolated characters [1].

2.1 Binarization and Noise Remo- oval

We used the Sauvola method for binarization [6] this method of thresholding is performed as a preprocessing step to remove the background noise from the picture prior to extraction of characters and recognition of text from video.

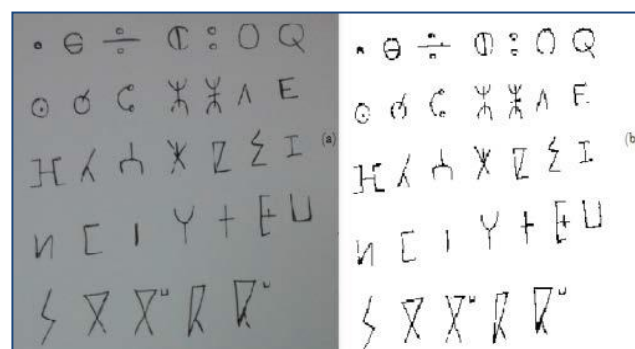


Fig1. (a) Example of an input image, (b) Thresholded image with Sauvola method.

Noise which is in the images is one of the big difficulties in optical character recognition process. The aim of this part is to remove and eliminate this obstacle; there are several methods that allow us to overcome this problem. In this work we decided to use the morphology operations to detect and delete small areas of less than 30 pixels[2].

2.3 Skew detection and correction

Skew correction methods are used to align the paper document with the coordinate system of the scanner. Main approaches for skew detection include line correlation [7], projection profiles [8], Hough transform [9], etc. For this purpose two steps are applied. First, the skew angle is estimated. Second, the input image is rotated by the estimated skew angle. In this paper, we use the Hough transform to estimate a skew angle θ_s and to rotate the image by θ_s in the opposite direction.

2.4 Segmentation

Next step for OCR is the Segmentation of the image. In This paper we propose a segmentation algorithm, in which text is easily segmented into Lines and Words using the traditional vertical and horizontal projection[10].

2.4.1 Line Segmentation

Once the image of the text cleaned, the text is segmented into lines. This is used to divide text of document into individual lines for further preprocessing. For this, we used analysis techniques of horizontal projection histogram of the pixels in order to distinguish areas of high density (lines) of low-density areas (the spaces between the lines) (see Fig.2). These techniques were often used to extract lines in printed texts[1].

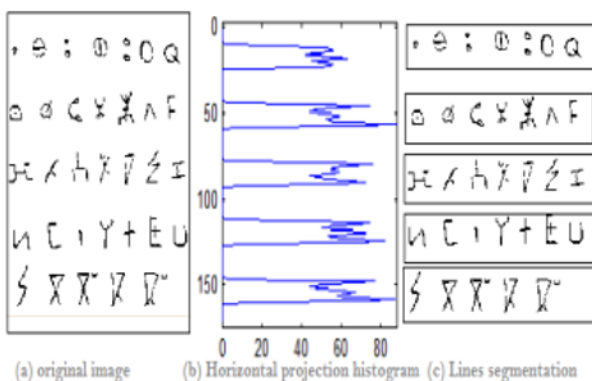


Fig.2 Lines segmentation

2.4.2 Characters Segmentation

We used in this part the vertical projection histogram to segment each text line of characters. Fig.3 shows a text line, the vertical histogram and the result of segmentation into characters [2].

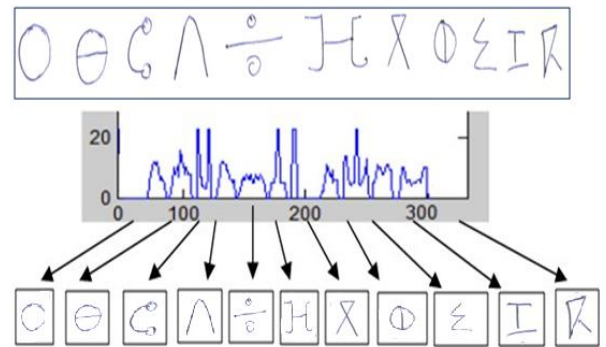


Fig.3 Characters segmentation

III. FEATURE EXTRACTION

In This part we present some feature extraction methods for recognition of segmented (isolated) characters [11]. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance in character recognition systems. Different feature extraction methods are designed for different representations of the characters, such as solid binary characters, character contours, skeletons (thinned characters) or gray-level sub-images of each individual character[11], In this paper, we have tested four methods: the Zoning types, Distance profile feature, Projection histogram and Gray Level Co-occurrence Matrix (GLCM) technique.

3.1 Zoning:

The zoning technique [3] is a statistical region-based feature extraction, it aim is to get the local characteristics in lieu of global characteristic. Therefore, according to the size normalized character image (60 x 50 pixels), we divided it into 30 (6 x 5) zones of 10 x 10 pixels size, then we calculated the densities of pixels in each zone, finally we are getting 30 features.

3.2 Projection histogram:

Projection histogram descriptor is a statistical feature; According to this feature we have used two direction of projection horizontal traversing. The horizontal histogram of the character Amazigh computed by counting the number of black pixels in each row. At the last we will have 60 features depending on the direction projection.

3.3 Gray Level Co-occurrence Matrix:

Gray Level Co-occurrence Matrix (GLCM) technique is an approach for extracting statistical texture features that have been proposed by Haralick [5]. The main principle of GLCM is to counts the number of times various combinations of pixel gray levels occur in a given image. Haralick defines 14 statistical

features measured from the GLCM. In this work, five important features are used namely energy, contrast, correlation, entropy and homogeneity.

3.4 Distance profile:

In distance profile feature [4] the distance (number of pixels) between the bounding box of image and the first pixel of foreground will be calculated. We have employed two types of profiles sides left and top. Concerning left profile, it is extracted by counting the distance from the left bounding box to the nearest foreground pixels in each row. Then as well, top profile, it is extracted by counting the distance from the top bounding box to the nearest foreground pixels in each column.

Table 1 Combination of the different feature vectors

Feature Method	Contained feature	Size
FM1	Zoning	30
FM2	GLCM	5
FM3	Distance Profile (Left + Top)	110
FM4	Projection Histogram Horizontal	60

IV. CLASSIFICATION

In the complete process of system recognition of forms, the classification plays an important role by pronouncing on the membership of a shape in a class. The main idea of the classification is to attribute an example (A form) not known about one Class predefined from the description in parameters of the form. Several surrounding areas of classification are used in the field of recognition of forms which are more or less good adapted to the recognition of the writing.

In litterateur, there are many types of classifiers that have been implemented in handwritten optical Amazigh character recognition problems. Among them, in this paper we have used two classifiers: the Convolutif Neural Network (CNN) and Random Forest.

4.1 Convolutif Neural Network

Convolutional networks are derived from perceptron architectures Multi Layer Perceptron (MLP), however they use shared weights, related to the convolution window, which allow them an implicit extraction of local features.

The difference of Convolutional neuron networks compared to conventional networks of MLP type, let us analyze the principle of recognition on the character G ("yag" in Amazigh), Fig.4 A neuron of an MLP is fully connected to all the neurons of the previous layer while for a convolutional network, a neuron is connected to a subset of neurons of the previous layer. Each neuron can be seen as a unit for detecting a local characteristic, a particular structural singularity such as the detection of a vertical or horizontal line, or even a loop. Along the trajectory,

The matrix of weights corresponding to the sliding window is identical (notion of shared weights): same detection, same convolution

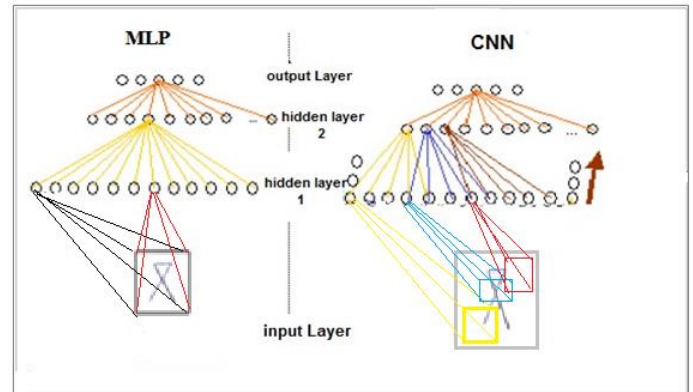


Fig.4 The difference of Convolutional neuron networks compared to conventional networks of MLP type

4.2 Random Forest

Random forest is an ensemble training algorithm that constructs multiple decision trees. It suppresses over-fitting to the training samples by random selection of training samples for tree construction in the same way as is done in bagging (Breiman,1996)[12], (Breiman,1999)[13], resulting in construction of a classifier that is robust against noise. Also, random selection of features to be used at splitting nodes enables fast training, even if the dimensionality of the feature vector is large [1].

• Algorithm

$z = \{(x_1, y_1), \dots (x_n, y_n)\}$ learning sample, x_i describes nominal variables p explanatory [20]:

1. for $b=1$ to B (B number of trees)
 - (a) Draw a bootstrap sample z_b of size N from the training data
 - (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
 - i. Select m variables at random from the p variable
 - ii. Pick the variable/split-point among them
 - iii. Split the node into two daughter nodes
2. Output the ensemble of tree $\{T_b\}_1^B$

To make a prediction at a new point x :

Regression:

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

Classification: let $\hat{C}_b(x)$ be the class prediction of the b th random-forest tree. Then

$$\hat{C}_{rf}^B(x) = \text{majority vote } \{\hat{C}_b(x)\}_1^B \quad (2)$$

Why Random Forest works [20]

Mean Squared Error = Variance + Bias²

If trees are sufficiently deep, they have very small bias
 How could we improve the variance?

$$\text{var} \left(\frac{1}{B} \sum_{i=1}^B T(c) \right) = \frac{1}{B^2} \sum_{i=1}^B \sum_{j=1}^B \text{Cov}(T_i(x), T_j(x)) \quad (3)$$

V. EXPERIMENTAL RESULTS

Due to the absence of standard database of handwritten Amazigh characters acquired by camera phone, we have constructed our own database of upper-case Amazigh character images obtained by Video camera phone. The database contains 100 samples of 33 classes, collected from 10 different writers. As a result the database consists of 3300 samples.

The samples are divided randomly into two set, one for training stage, we have used 85 % (2805 samples) and the other for testing stage, we have used 15 % (495 samples). We have tested the proposed system on database of handwritten Amazigh characters acquired by camera phone the HUAWEI Y9 of this characteristics; 16Megapixels.

For classification stage we have used two classifiers: the Convolutif Neural Network (CNN) and the Random Forest for each classifier we employed a set of different features extraction methods.




The Zoning feature extraction (FM1) provides higher recognition and learning rate by Random Forest with the achievement of a Recognition rate of 94.02 %. Also FM 4 and FM 3 give some encouraging results. According to the results of Convolutif Neural Network the hybrid method FM 2 achieves a very good recognition and training rate: 96, 19 % of Learning rate and 94, 01 % of Recognition rate.

Table.2 Results of different single feature vectors using Convolutif Neural Network and Random Forest classifiers

Classifier Feature Vector	Convolutif Neural Network		Random Forest (N=600)	
	Learning R.	Recognition R.	Learning R.	Recognition R.
FM1	95.66 %	93.36 %	95.26 %	94.02 %
FM2	96.19 %	94.01 %	94.54 %	92.64 %
FM3	93.41 %	91.59 %	93.94 %	93.22 %
FM4	95.10 %	93.70 %	93.62 %	93.13 %

Out of the 495 Read-only characters, 392 were recognized, representing a recognition rate of 94.02%. With respect to the rate obtained for each letter, the best result achieved with this approach was 98.89%, for the character ('Ha' in Amazigh). Table 3 below shows the recognition rate obtained on certain characters.

Table.3 below shows the recognition rate obtained on certain characters.

Characters	Recognition Rate
	94.39 %
	98.89 %
	93.72 %

The recognition errors are high for the letter ('Rr' in Amazigh), which is explained in particular by the insufficiency of the characteristics used to better describe each character during phase of extraction the primitives, and to the initial data used during the learning step. A good estimate of this data can reduce the error rate of our system.

VI. CONCLUSION

In this paper, we have presented a system of handwriting Video Text Amazigh recognition based on the method Random Forest and Convolutif Neural Network. Several features have been studied and compared; as a result we've chosen Sauvola [10] method duo its ability to remove the noise. The experiments carried out in database were performed on a database obtained by Video camera phone with applying different classifiers and for each classifier we have tested a set of single feature methods.

The results obtained in this paper that has been compared and analyzed have shown that Random Forest with Zoning feature is the best in terms of recognition accuracy rate and GLCM technique provide higher recognition rate by CNN.

In future work, we will add other features methods that improve the results for some characters for example, minimize the length of execution of program which to calculate the recognition rate.

REFERENCES

- [1] Wei Chen ,Xiaoshen Xie, Jiale Wang, Biswajeet Pradhan ,Haoyuan Hong, Dieu Tien Bui , Zhao Duan, Jianquan Ma, 'A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility', Vol 151, pp 147-160 , April 2017.
- [2] Hassan El Bahi, Zouhir Mahani and Abdelkarim Zatni "A robust system for printed and handwritten character recognition of images obtained by camera phone" .*Published in WSEAS Transactions on Signal Processing, Volume 11*, 2015, pp.9-22
- [3] Elima Hussain, Abdul Hannan, Kishore Kashyap, "A Zoning based Feature Extraction method for Recognition of Handwritten Assamese Characters" *IJCST Vol. 6, Issue 2, April - June 2015*

- [4] Dongdong Guan ; Deliang Xiang ; Xiaoan Tang ; Li Wang ; Gangyao Kuang “Covariance of Textural features: A New Feature descriptor for SAR image Classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, issue 10, pp. 3932-3942, Oct 2019.
- [5] J. Sauvola and M. Pietikainen, “Adaptive Document Image Binarization,” *Pattern Recognition* 33(2), pp. 225–236, 2000
- [6] SAB Haji, A James, S Chandran, “A Novel Segmentation and Skew Correction Approach for Handwritten Malayalam Documents”, Vol 24, pp 1341-1348 , 2016
- [7] E. Grilli, F. Menna, F. Remondino, “A REVIEW OF POINT CLOUDS SEGMENTATION AND CLASSIFICATION ALGORITHMS”, Volume XLII-2/W3, March 2017 Nafplio, Greece.
- [8] Y.Boukharouba, ‘A New algorithm for skew correction and baseline detection based on the randomized Hough Transform, *JKSU-CIS*, Vol 29, issue 1, pp 29-38, January 2017
- [9] Archana A. Shinde, D.G.Chougule, “Text Pre-processing and Text Segmentation for OCR” *IJCSET /January 2012/ Vol 2, Issue 1, 810-812*
- [10] Y.rachidi, Z.mahani ; “Recognition image Obtained by camera phone of character Arabic”, *IJDR*, Vol 08, March 2018.
- [11] Breiman, L. ; Using adaptive bagging to debias regressions. In Technical Report. Statistics Dept. UCB.
- [12] Saad Albawi , Tareq Abed Mohamed, Saad Al-zawi ,”Understanding of a convolution neural network”, *ICET*, 21-23 Aug 2017.