Research Inventy: International Journal of Engineering And Science
Vol.10, Issue 6 (June 2020), PP 01-04
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Recognition System of Character Latin Acquired by a Smartphone based on Features Horizontal Centerline

Rachidi Youssef [1], Ikermane Mohamed [2]
[1]*IRF –SIC Laboratory, University Ibn Zohr Agadir, Morocco*
[2] *Laboratory LABSI, University Ibn Zohr Agadir, Morocco*

Abstract: *In this paper; we introduce a system recognition of characters Latin handwriting and Printed recognition based on horizontal centerline in non-constrictive pictures that are stemmed from the terminals Mobile phone . After doing some pretreatments on the picture, the text is segmented into lines and then into characters. The positions of the horizontal centerline of the character are used to obtain a set of independent and dependent features to those lines. These characteristics are linked to pixels densities and they are extracted on binary pictures. Finally, a multilayer perceptron is used for character classification. The system was tested on two bases of the Latin characters: on a printed database of Latin characters and on another one for handwritten characters created locally. The correct average recognition rate obtained using 10-cross-validation was 98.36 % for the 2600 images of Latin printed characters and 97.09% for the 3900 Latin handwritten characters.*

Keywords*: Mobile phone ; Handwritten Character Recognition; Character Latin; horizontal centerline; MLP; Segmentation.*

Date of Submission: 10-06-2020 Date of Acceptance: 27-06-2020

I. INTRODUCTION

The automatic recognition of handwritten or printed characters Latin 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]. Someattempts have been done to improve the current situation [1]. In this context, we have employed a recognition system of handwritten and printed Latin characters extracted from a picture taken by camera phone [2]. Indeed, in the primitives' extraction stage, our approach is based on Parameters such as horizontal centerlines are used to derive a subset of dependent centerline features from the segmented characters [3].

These primitives will supply a multilayer perceptron in the learning and recognizing phases. On a database of handwritten Latin, segmented and isolated characters acquired by camera phone, obtained an encouraging results on the majority of this characters. The database contains 150 samples of 26 classes, collected from 5 different writers. As a result the database consists of 3900 samples of handwritten and 100 samples of 26 classes of printed characters. For classification stage we have used the Neural Network.

Finally a multilayer perceptron with one hidden layer are used for the training step of extracted features and the recognition of characters.

The paper is organized as follow: Section 2 displays the system architecture of the proposed method and we will deeply discuss, respectively, pre-processing and feature extraction. In Section 3, we present the training and recognition steps. Results are given in section 4. Finally, conclusions and future work are discussed in section 5.

Figure1. Text Recognition of Smartphone.

II. OUR SYSTEM ARCHITECTURE

Generally, a system of off-line handwritten character recognition includes three stages: pre-processing, feature extraction, and classification. The process of handwriting recognition involves extraction of some defined characteristics called features permitting to classify an unknown character into one of the known classes [4]. The pre-processing is primarily used to reduce variations of handwritten characters and printed characters, to correct the skew of the text lines and to segment the text into isolated characters. The feature extraction step is essential for efficient data representation and extracting meaningful features for later processing. The classifier assigns the characters to one of classes.

Pre-Processing

The procedure of preprocessing which refines the scanned input image 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 [4].

Features Extraction

After pre-processing, a feature extraction method is applied to extract the most relevant characteristic of the character to recognize. 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, The performance of a character recognition system largely depends on the quality and the relevance of the extracted features. In this paper, the extracted features consist of the letter Features based on the horizontal centerline: This step separates the image of the character into two zones: an upper zone that corresponds to the area above the baseline, which is the horizontal central line, and a lower zone which corresponds to the area below the baseline. Fig. 2 shows the line position of writing on a few Latin characters [5][6][7].

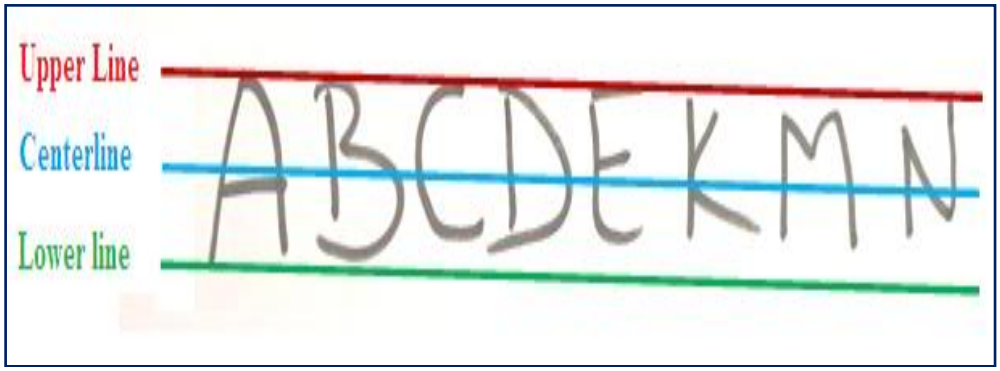

Figure 2. Line position of writing on a few Latin characters.

To create the feature vector, the letter is then scanned from left to right and from top to bottom with a sliding window [8][9]. The image is divided into horizontal frames (Fig. 3). The height and width of the frame are constant and are considered as parameters. The window height varies according to each image [13].

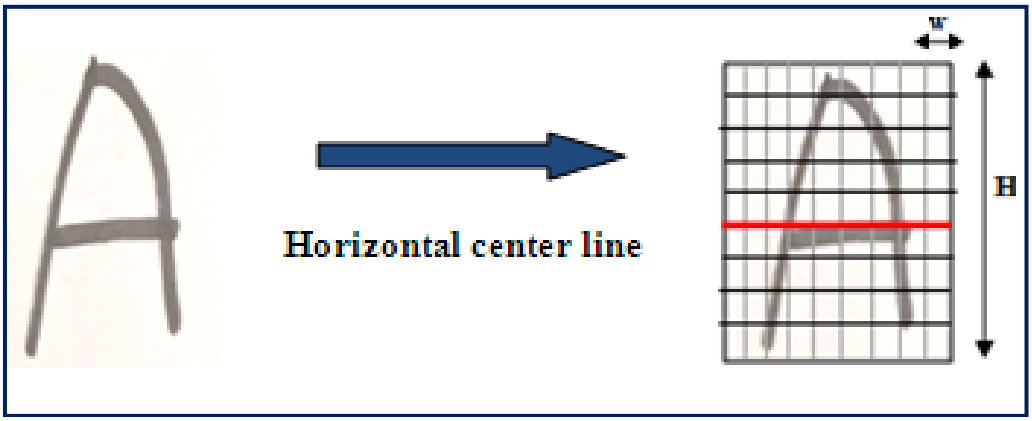

Figure 3. Dividing letters into horizontal frames

III. TRAINING AND RECOGNITION

In this paper, we choose a multi-layer perceptron architecture using the back propagation with momentum learning scheme[10][11][12]. Before feature extraction, each letter is normalized to 60×50 pixels. Then, the different sets of features (45 features based on the horizontal centerline) are used in the training and testing modules. There are several available architectures and learning methods which may be selected depending on a given problem. The multilayer perceptron architecture consists of an input layer of information processing nodes, a hidden layer with hidden nodes and an output node which consists of output nodes that usually equal the number of data classes. The number of neuron outputs is 26. The numbers of hidden nodes vary across different applications and this number must be experimentally determined. In our case, we adopted a number of hidden layer neurons equivalent to the number of attributes and classes divided by two. The network learned on the entire training set using the back propagation method, and was then tested (with the validation set) for its performance with a limit of 1000 epochs. The network is trained with 0.3 learning rate and 0.2 momentums constant. This multilayer perceptron implementation is available in the Weka collection of machine learning algorithms.

IV. RESULTS AND ANALYSIS

We have tested the proposed system on two bases: a printed database of Latin characters [5] and another one for handwritten characters presented previously. The tests were performed according to the integration of features dependent and independent on the horizontal centerline. Also, we have used a 15-fold cross validation scheme for recognition result evaluation. For 15 fold cross validation, 85% of the data was used for training and this performance was tested on the remaining 15%. We divided database in 15 major groups and measured recognition accuracy for each group separately. The recognition rates of all the 15 test subsets of the dataset are averaged to get the recognition result. Table 1 shows the experiments results using ten-fold cross-validation on the base of printed Latin characters and on the base of handwritten characters.

In table 1,for the base of Handwitten Latin characters, the recognition rate is 97.09% when integrating the features based on the position of the horizontal centerline and and increases to 90.38% when use features independent on the centerline . This demonstrates that the features based on the position of central lines offer a significant improvement in the recognition performance.

For the base of Printed Latin characters, the recognition rate is 98.36% when integrating the features based on the position of the horizontal centerline and and increases to 94,25% when use features independent on the centerline .

Table 1. Recognition Results Using Ten-fold Cross-validation

Features Integrated	Handwitten Latin Characters		Printed Latin Characters	
	Database zise	Rec.Rate %	Database zise	Rec.Rate %
Features dependent and independent on the horizontal centerline	3900 characters	97,09%	2600 Characters	98,36%
Features independent on the centerline	3900 characters	90,38%	2600 Characters	94,25%

IV. CONCLUSION AND PERSPECTIVES

In this paper, we have presented a system for automatic recognition of Latin handwriting and Latin printed acquired by Smartphone, based on the position of the horizontal centerline of each character. Several features have been studied and compared. The importance of using the position of the centerline in the image of the character has been proved. The extracted features are based on the density of pixels derived from a sliding window. The developed system was tested on two databases acquired by Smartphone, a database of printed Latin patterns and another base of handwriting Latin character developed locally. These results show a significant improvement in recognition rate when integrating the features dependent on the horizontal centerline. In future work, we will add other features that improve the results for some characters, such as information on the possible inclination of writing using the windows inclined.

REFERENCES

[1]. R. M. Bozinovic and S. N. Srihari, "Off-line cursive script word recognition", IEEE Trans. on Pattern Anal. Mach. Intell., vol. 11, no. 1, 1989, pp. 68-83, Jan.

[2]. I. Bazzi, R. Schwatz and J. Makhoul: "An Omni font Open-Vocabulary OCR System for English and Arabic. IEEE Transactions on pattern Analysis and Machine Intelligence". vol. 21, no 6. 1999, pp. 495-504..

[3]. Z. Zhang, L. Jin, K. Ding, X. Gao: Character-SIFT: "A Novel Feature for Offline Handwritten Chinese Character Recognition", 10th International Conference on Document Analysis and Recognition, 2009, pp. 763- 767.

[4]. **Youssef Rachidi**. Improvement of the Recognition Rate by Random Forest, Int. Journal of Engineering Research and Application Vol. 7, Issue 2, (Part -4) February 2017, pp.37-43

[5]. Y. Ait ouguengay, M. Taalabi, „„Elaboration d"un réseau de neurones artificiels pour la reconnaissance optique de la graphie amazighe: Phase d"apprentissage", Systèmes intelligents-Théories et applications, 2009.

[6]. R. El Yachi and M.Fakir, "Recognition of Tifinaghe Characters using Neural Network", International Conference on Multimedia Computing and Systems, Actes dE ICMCS"09, Ouarzazate, Maroc, 2009.

[7]. Y. Es Saady, A. Rachidi, M. El Yassa, D. Mammass, "Reconnaissance Automatique de l"Ecriture Amazighe à base de Ligne Centrale de l"Écriture", 4ème Atelier international sur l"amazighe et les TIC, 2011, IRCAM,Maroc.

[8]. M. Amrouch, Y. Es Saady, A. Rachidi, M. Elyassa, D. Mammass, "Printed Amazigh Character Recognition by a Hybrid Approach Based on Hidden Markov Models and the Hough Transform", International Conference on Multimedia Computing and Systems, Actes de ICMCS"09, Ouarzazate, Maroc, 2009.

[9]. M. Amrouch, A. Rachidi, M. Elyassa, D. Mammass, "Handwritten Amazigh Character Recognition Based On Hidden Markov Models, ICGST-GVIP Journal, Vol.10, Issue 5, pp.11-18, 2010.

[10]. Y. Es Saady, A. Rachidi, M. El Yassa, D. Mammass , " Printed Amazigh Character Recognition by a Syntactic Approach using Finite Automata", ICGST-GVIP Journal, Vol.10, Issue 2, 2010, pp.1-8

[11]. R. El Yachi, K. Moro, M. Fakir, B. Bouikhalene, "On the Recognition of Tifinaghe Scripts", Journal of Theoretical and Applied Information Technology, Vol.20, No.2, 2010, pp.61-66.

[12]. R. El-Hajj, L. Likforman-Sulem, C. Mokbel, "Arabic handwriting recognition using baseline dependent features and Hidden Markov Modeling", ICDAR 05, Seoul, Corée du Sud, 2005.

[13]. M. Razzak, M. Sher, S. Hussain, "Locally baseline detection for online Arabic script based languages character recognition", International Journal of the Physical Sciences Vol. 5(7), 2010, pp. 955-959.

Rachidi Youssef, et. al. "Recognition System of Character Latin Acquired by a Smartphone based on Features Horizontal Centerline." *International Journal of Engineering and Science*, vol. 10, no. 06, 2020, pp. 01-04.