International Conference of Science, Engineering & Environmental Technology (ICONSEET), 2(13): 95-103, 2017
ISSN 0794-9650

www.repcomseet.com

An Offline Yorùbá Handwritten Character Recognition System Using Support Vector Machine

Oladele, M.O.[1], Adepoju, T.M.[1], Omidiora, E.O.[2], Sobowale, A.A.[1], Olatoke, O.A. [2], Ayeleso E.C.[3]

[1] Department of Computer Engineering Technology, The Federal Polytechnic, Ede, Osun State.
[2] Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso.
[3] Department of Computer Science, The Federal Polytechnic, Ede, Osun State.

Abstract: Handwriting recognition is one of the most fascinating and challenging research areas in the field of image processing and pattern recognition which emanates from the need for humans to automate the recognition of handwritten text and enable the computer to receive and interpret them. Several handwritten character recognition systems had been developed to recognize characters or texts of some languages such as English, Arabic, Sinhala and so on, however, it was observed that there are few character recognition system for Yorùbá language. This paper developed a Support Vector Machine based Yorùbá character recognition system to recognize Yorùbá alphabets. The process is divided into four stages, the preprocessing, segmentation, feature extraction, and classification stage. The developed system was experimented with 600 handwritten images for Yorùbá alphabets. 480 was used for training and 120 was used for testing. The results showed a training time of 45.842 seconds, recognition rate of 76.7% and rejection rate of 23.3%.

Keywords: Character Recognition, Image Processing, Pattern Recognition, Support Vector Machine (SVM).

1 Introduction

Character is the basic building block of any language which is used to develop different language structures. Characters are alphabets and the structures developed are the words, strings, sentences, paragraphs and so on (Le Cun *et al.*, 1990). Character recognition also known as optical character recognition is the recognition of optically processed characters. The purpose of character recognition is to interpret input as a sequence of characters from an already existing set of characters (Kader and Deb, 2012).

Handwritten character recognition is the process of converting handwritten text into a form that can be read by the computer. The major problem in handwritten character recognition system is the variation of the handwriting styles of individuals, which can be completely different for different writers (Patel and Thakkar, 2015). Handwritten character recognition system can be divided into two categories namely the online character recognition and the offline character recognition.

Online character recognition is the conversion of text written on a digitizer or PDA automatically where the sensor picks up the pen - tip movements and the pen-up/pen-down switching. The signal obtained from the pen - tip movements is converted into letter codes that can be used by the system and text processing applications. In offline character recognition, the image of the written text is scanned and sensed offline by optical scanning (optical character recognition) or intelligent character recognition (Tawde and Kundargi, 2013).

Support Vector Machine (SVM) is a classifier that separates classes in feature space. It is based on statistical learning theory developed by the Russian scientist Vladimir Naumovich Vapnik in 1962 (Thomé, 2012). It is used to identify a set of linearly separable hyperplanes which are linear functions of the feature space. Among the separable hyperplanes only one hyperplane is chosen and placed such that the distance between the classes is maximum (Kansham and Renu, 2014). SVM has very high accuracy rate for two class problem (Yes or No) but it can also be modified to classify multiclass problem.

The advantages of the character recognition system are that it can save both time and effort when converting to a digital version of the document and also provide a fast and reliable alternative to manual typing (Kader and Deb, 2012). Automated recognition of handwritten characters can be applied in several areas such as Postal and Banking for reading of addressed packages and cheques respectively (Omidiora, Adeyanju, and Fenwa, 2013).

This paper proposes an offline handwritten Yorùbá character recognition system using SVM to recognize Yorùbá upper case letters. Twenty-four (24) uppercase Yorùbá alphabets were considered in this paper. Twenty (20) of these characters were used for training and five (5) characters were used to test the system and evaluate the performance based on recognition rate and rejection rate.

2 Literature Review

There are several research works on handwritten character recognition but there are few ones on Yorùbá character recognition. Few of the existing works on handwritten character recognition are highlighted in this sub-section. Kessentini, Paquet and Benhamadou presented a multi-stream approach for offline handwritten word recognition (Kessentini, Paquet and Benhamadou, 2010). Low level feature streams namely, density based features and contour based features were combined and Hidden Markov Model (HMM) was used as the classifier. The approach achieved a recognition rate of 89.8% using a lexicon of 196 words. Ibraheem and Odejobi (2011) developed a system to recognise handwritten character of Yoruba Upper case letters using Bayesian and decision tree. A recognition rate of 94.44% was achieved. The research work focused on six Yoruba upper case characters only.

Fenwa, Omidiora and Fakolujo (2012) presented a hybrid feature extraction techniques using Geometrical and Statistical features. A hybridized classification model was developed to train the neural network using modified counter propagation and modified back propagation learning algorithms. A recognition rate of 96% was achieved. Ajao, Olabiyisi, Omidiora and Odejobi (2015), presented an evaluation of preprocessing attributes of Yoruba handwriting word recognition. The approach is aimed at assessing the intrinsic measure of some of the preprocessing stages. From the experiment carried out, it was observed that the entropy measure of handwritten word is higher than the typewritten word.

3 Materials and Methods

The system was developed and simulated using MATLAB 2013a. Twenty-five (25) different handwriting styles for 24 Yorùbá alphabets were used which gives a total of 600. The Yorùbá alphabet “GB” was not included in the dataset but it was classified as two separate characters “G” and “B” because there was no Unicode representation of the alphabet. The Unicode standard, version 4.1 was used to display the alphabets especially the characters with under dot (Ẹ, Ọ, Ṣ).

3.1 Design Approach

The complete framework for this paper is presented in figure 1. The first step is the acquisition of handwritten characters. The second step is the preprocessing stage where the image is converted to a gray image, filtered, converted to black and white image and dilated. The pre-processed image was segmented and the feature extraction stage extracts the features that were used as input to the classifier. The statistical feature extraction method was used for the experiment. The recognition stage recognises the input character.

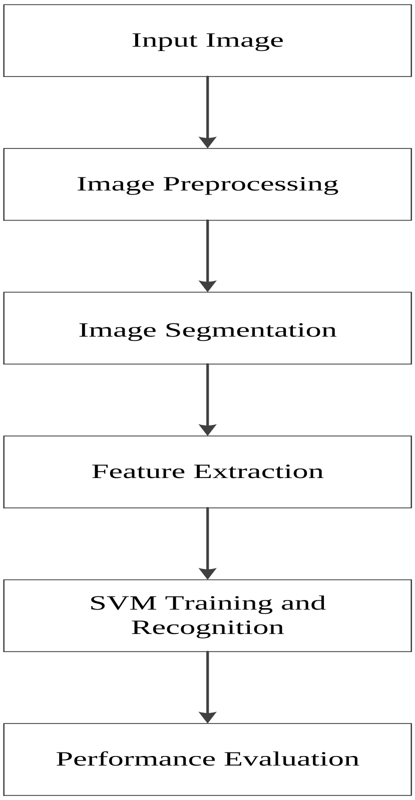


Figure 1: Framework for Yorùbá Handwritten Character Recognition System

3.1.1 Data Acquisition

There is a need to have a database for Yorùbá characters to test the performance of the system. For this work, twenty-five different Yorùbá characters were gotten from different individuals using the Paint application. Twenty images were used to train the SVM classifier and five was used to test for the accuracy of the system. Figure 2 shows the sample images that were acquired.

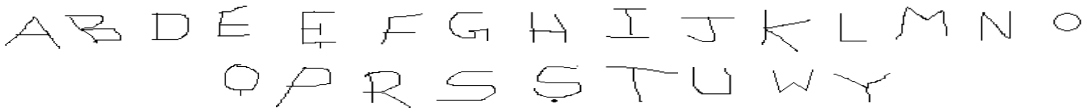


Figure 2: Acquired sample images

3.1.2 Image Pre-processing

Image pre-processing helps to prepare an image for subsequent processing. It helps in increasing the recognition accuracy by removing unwanted information that would have been extracted as features. Pre-processing stage involves the following stages grayscale conversion, image filtering, edge detection, binary image conversion, image dilation and image filling.

Grayscale conversion was used to reduce the number of bits per pixels, image filtering helps to remove the noise generated from the handwritten image. Edge detection helps to identify edges and points, binary image conversion converts the image to black and white which is a combination of zeros and ones. Image dilation helps to bold the image and image filling fills the holes in the input character. Image dilation and filling is also known as morphological processing.

3.1.3 Image Segmentation and Cropping

Image segmentation involves separating the characters in the image. It helps the classifier to extract features from each individual character. A boundary was set on the character and cropped accordingly such that only the handwritten image was left with the region which contains the necessary features needed for feature extraction. Cropping helps to remove unnecessary features such as the background thereby leaving just the relevant features. Figure 2 shows a sample of cropped image.

Figure 2: Sample of the cropped input image

3.1.4 Feature Extraction

Feature extraction was used to extract properties that can identify a character uniquely and also to extract the properties that can differentiate between similar characters. It distinguishes an area corresponding to a letter from an area corresponding to other letters. It is an essential stage in handwriting character recognition as its effective functioning improves the recognition rate and reduces the misclassification. It is used to extract the features of individual characters which is used to train the system. Statistical feature extraction technique was used to extract the characters. The major statistical features used in this work are zoning, projections and profiles, and crossings and distances.

3.1.5 Training and Recognition

The training and recognition stage is the last stage for the developed system. The training and recognition was done using SVM. Twenty images for Yorùbá upper case characters were used for training and five for testing. The parameters used for the SVM classifier was the RBF kernel function, 10 - fold cross validation and the one vs one model.

For the recognition stage, the feature vector obtained from the feature extraction stage was used for recognition. The feature vectors of the input handwritten image were compared with the stored feature vectors of the training images to get the result. The Unicode of the characters were used to get the result since the standard keyboard does not have the Yorùbá alphabets. The flowchart of the system is shown in figure 3.

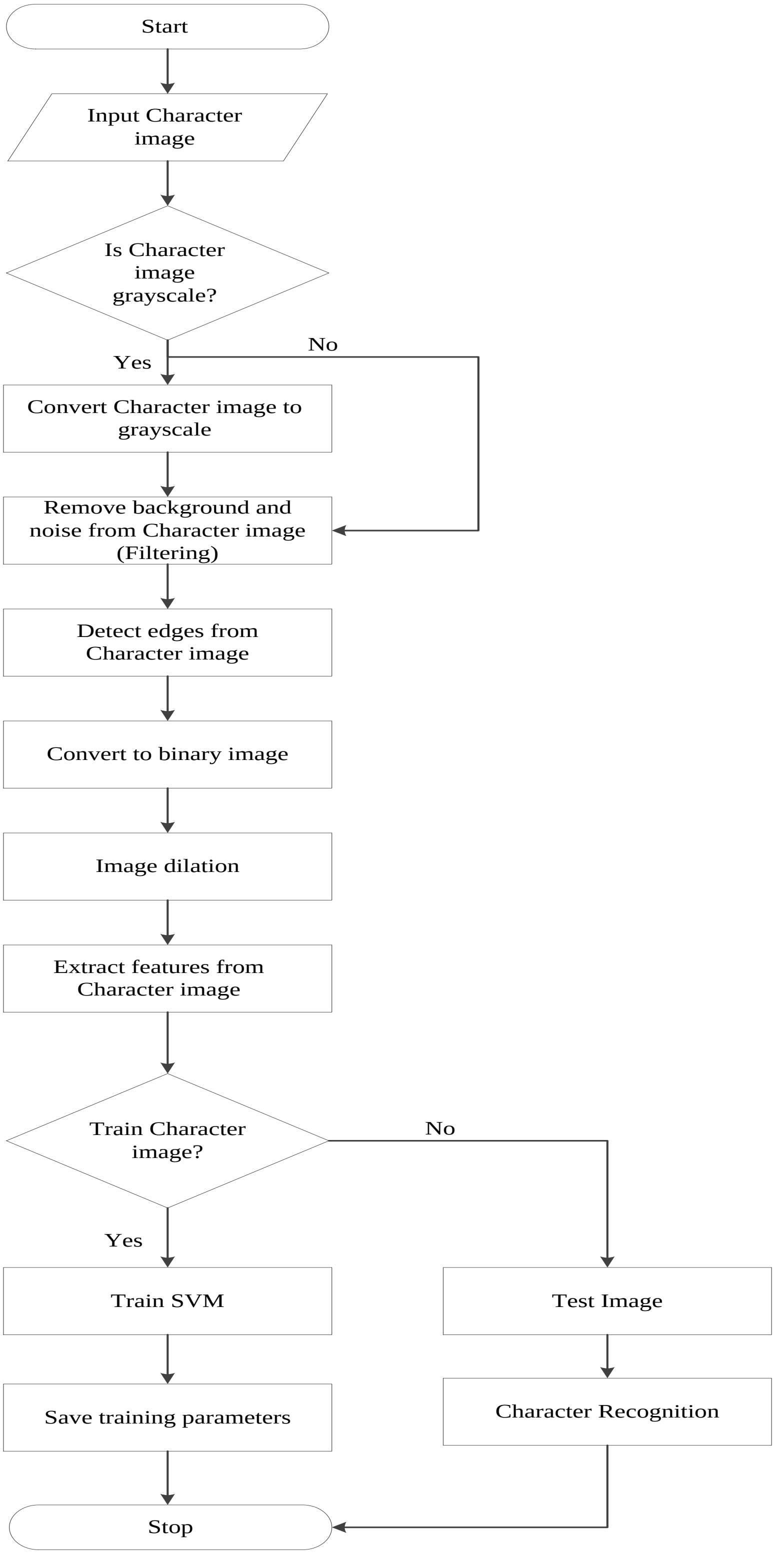

Start
Input Character image
Is Character image grayscale?
Yes
No
Convert Character image to grayscale
Remove background and noise from Character image (Filtering)
Detect edges from Character image
Convert to binary image
Image dilation
Extract features from Character image
Train Character image?
No
Yes
Train SVM
Test Image
Save training parameters
Character Recognition
Stop

Figure 3: Flowchart of the Yorùbá Character Recognition System

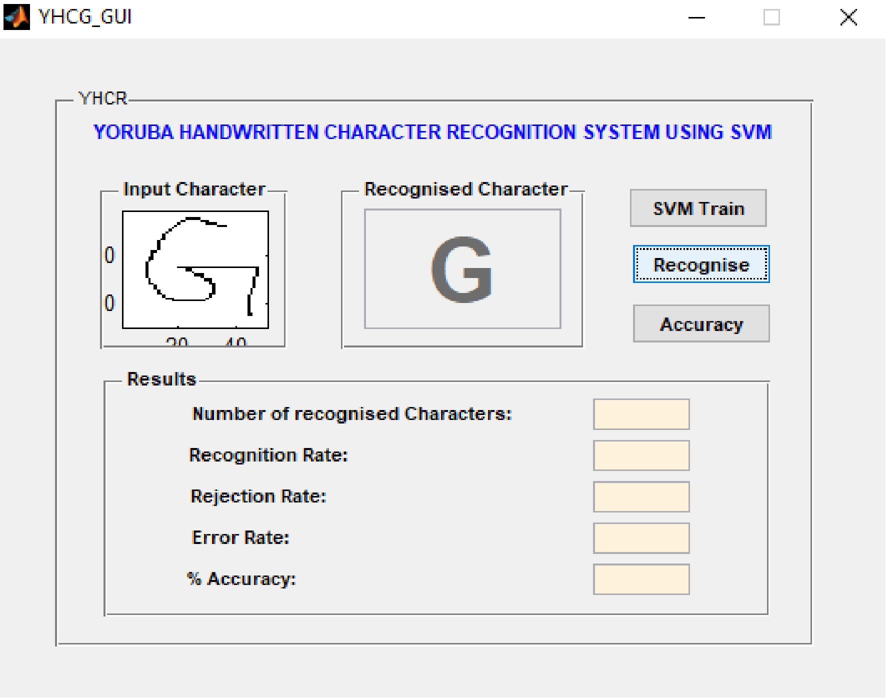
4 Results and Discussion

The Yorùbá handwritten character recognition system was simulated and tested on a 6GB RAM, Intel core i5 and 2.40GHZ CPU speed HP pavilion laptop computer and the results obtained from the developed system shows a total training time of 45.842 seconds with 480 handwritten characters. 120 characters (20% of the handwritten samples) were used to test the system.

A total of 92 characters were correctly recognised. 7 characters were not recognised and 21 characters were incorrectly recognised which shows a recognition rate of 76.7% and a rejection rate of 23.3%. Table 1 shows the results on the different characters. Figure 4 shows the output of the developed system.

Table 1: Results of Yorùbá handwritten character recognition system

Character	**Testing Samples**	**Correctly recognised**	**Incorrectly recognised**	**Not recognised**	**Recognition Rate**
A	5	3	2	0	60%
B	5	5	0	0	100%
D	5	5	0	0	100%
E	5	3	2	0	60%
Ẹ	5	3	2	0	60%
F	5	5	0	0	100%
G	5	4	1	0	80%
H	5	2	3	0	40%
I	5	5	0	0	100%
J	5	3	0	2	60%
K	5	5	0	0	100%
L	5	1	4	0	20%
M	5	5	0	0	100%
N	5	3	2	0	60%
O	5	3	0	2	60%
Ọ	5	3	2	0	60%
P	5	4	0	1	80%
R	5	4	1	0	80%
S	5	4	1	0	80%
Ṣ	5	4	1	0	80%
T	5	5	0	0	100%
U	5	5	0	0	100%
W	5	4	0	1	80%
Y	5	4	0	1	80%



YHCG_GUI

YHCR

YORUBA HANDWRITTEN CHARACTER RECOGNITION SYSTEM USING SVM

Input Character

Recognised Character

Results

Number of recognised Characters:

Recognition Rate:

Rejection Rate:

Error Rate:

% Accuracy:

Figure 4: Output of the Yorùbá handwritten character recognition system

5 Conclusion

A system for recognizing Yorùbá handwritten characters was presented. The system was developed using the SVM classifier. Both the training and testing were done using SVM. The performance was evaluated based on recognition rate and rejection rate. The result shows that the recognition rate of the system was 76.7% while the rejection rate was 23.3%.

6 Recommendations

The work can be extended to lower case Yorùbá characters and also to words. Also, other classifiers can be used to improve the recognition rate and reduce the rejection rate. Finally, Yorùbá handwriting character database can be designed such that large data will be available for testing.

References

[1] Ajao, J.F., Olabiyisi, S.O., Omidiora, E.O. and Odejobi, O.A. (2015), "Yoruba Handwriting Word Recognition Quality Evaluation of Preprocessing Attributes using Information Theory Approach", International Journal of Applied Information Systems (IJAIS), Vol. 9, No. 1. pp. 18 - 23.

[2] Fenwa, O.D., Omidiora, E.O. and Fakolujo, O.A. (2012), "Development of a Feature Extraction Technique for Online Character Recognition System", *Innovative Systems Design and Engineering*, Vol. 3, No. 3.

[3] Ibraheem, A.O., and Odejobi, O.A. (2011), "A System for the Recognition of Handwritten Yoruba Characters", AGIS, Ethiopia.

[4] Kader, M. F. and Deb, K. (2012), "Neural Network-Based English Alphanumeric Character Recognition", *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, Vol. 2, No. 4.

[5] Kansham A. M, Renu, D. (2014), "Recognition of Cheising Iyek/Eeyek-Manipuri digits using Support Vector Machines*" International Journal of Computer Science & Information Technology*, Vol. 1 No. 2, pp 1-6.

[6] Kessentini, Y., Paquet, T., and Benhamadou, A. (2010), "Offline Handwritten Word Recognition using Multi-Stream Hidden Markov Models", *Pattern Recognition Letters*, Vol. 30, No. 1, pp. 60 - 70.

[7] Le Cun, Y., Boaer, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D. (1990), "Handwritten zip code recognition with multilayer networks", *International Conference on Pattern Recognition*, pp. 35-44.

[8] Omidiora, E.O., Adeyanju, I.A. and Fenwa, O.D. (2013), "Comparison of Machine Learning Classifiers for Recognition of Online and Offline Handwritten Digits", *Journal of Computer Engineering and Intelligent Systems*, Vol. 4, No. 13.

[9] Patel, M. and Thakkar, S.P. (2015), "Handwritten Character Recognition in English: A Survey", *International Journal of Advanced Research in Computer and Communication* Engineering**,** Vol. 4, No. 2.

[10] Tawde, G.Y. and Kundargi, J.M. (2013), "An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting", *International Journal of Engineering Research and Applications (IJERA)*, Vol. 3, No. 1.

[11] Thomé, A.C. (2014), "SVM Classifiers – Concepts and Applications to Character Recognition", *Advances in Character Recognition*, pp. 25 – 50.