Study and Analysis of Multilingual Hand Written Characters Recognition Using SVM Classifier

UJWAL SINGH VOHRA, SHRI PRAKASH DWIVEDI and H.L. MANDORIA

Department of Information Technology G. B. Pant University of Agriculture & Technology, Pantnagar, India.

ABSTRACT

Day by day the researchers are trying to make such characters recognition system that can be able to detect the writing and languages of individuals, for this multilingual handwritten characters recognition is such system that playing a vital role for recognizing the characters written in different languages and in different styles. The research work presented in this thesis aims to do the study and analysis to recognize the multilingual handwritten characters with a high level of accuracy and for this purpose the classifier that we are using is support vector machine. Here in this work we have used the languages like Hindi and English and along with this we have taken special characters and numerals and tried to recognize them with our system.

Keywords: OCR, Handwritten Characters,Recognition,SVM.

INTODUCTION

Multilingual Hand written Characters Recognition

While processing handwritten characters if a document contains more than one handwritten language characters to recognize then it refers as Multilingual Handwritten Characters Recognition[11]. Like here in the figure 1.1 we have taken Hindi and English as an input language characters and in figure 1.2 the Multilingual Handwritten Characters Recognition process is shown.

Previous Work

As per some research papers initially the Character recognitionstarted in late 1970s, there are many research that has been done in the field of Characters recognitionand HandwrittenCharacters Recognition till yet.In the field of character recognition many researchers has provided their opinions and proved it with their results. Deshpande (2008) in this paper they proposed Gaussian filter to provide feature vector for the dimension of 200 (5x5x8). And the accuracy achieved was 94% with Support vector Machines (SVM) as the classifier. DOGRA (2012) in this paper they have suggested

an automatic recognition method of OCR. They have attempted handwritten Hindi characters recognition and the classifier used was the SVM and the feature extraction method was diagonal feature extraction. The accuracy achieved by this method was 93.06 % [6].Bansal (2014)have proposed an efficient method for recognition Hindi handwritten numerals, with the use of energy and chain codes. For classification purpose SVM is used for classification. The average recognition of 90.1 % is achieved using four segment methods. Singh (2015)in this paper they got 97.61% result using SVM and ANN for the Handwritten Devanagari Character recognition. The accuracy/recognition rate obtained for different size of images by ANN and SVM classifiers. Farkya (2015) in this paper the a whole document was scanned and passed to

Fig. 1.1: Multilingual Handwritten Characters Image

OCR then the rate of recognition slightly decreases to 96 %.

Work flow of Multilingual Handwritten Characters recognition system

In this system the users have to upload a picture for which they want to perform the recognition task. After uploading the picture the recognition task is performed on the image[12]. While performing the recognition task the preprocessing, segmentation, feature extraction and classification tasks with using the SVM classifier are being performed on the image and the system finally produce an output of the recognition process. Here for skew detection we have taken randon transformations function and the segmentation of characters is done using horizontal histogram analysis and vertical histogram analysis methods.

Support Vector Machine (SVM) classifier

It is a supervised type of classification method in the discipline of statistical learning theory which has been successfully applied in the domains of bioinformatics and pattern recognition applications like text, face and voice recognition. In order to maximize the distance of the separating boundary between the two or more classes SVM tries to maximize the distance of the separating planes from each of the feature vector whether the feature vector belongs to class c_1, c_2 and so on. In SVM the optimization criteria is width of the margin between the classes i.e., empty area around the decision boundary specified by the distance to the
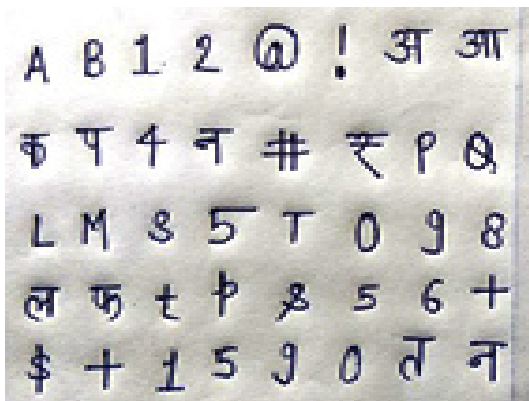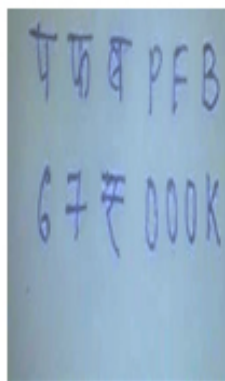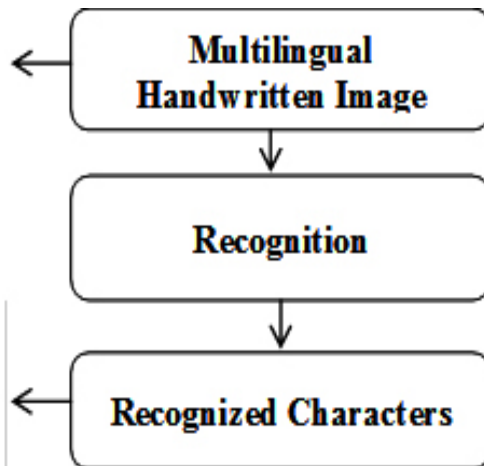

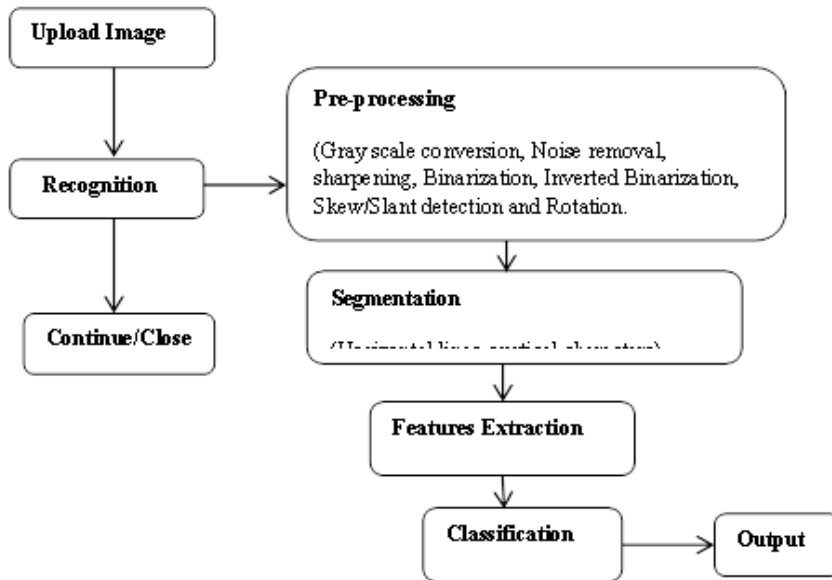**Fig. 1.2: Multilingual Handwritten Characters recognition**

Fig. 1.3: Pictorial representation of Multilingual Handwritten Characters recognition system

nearest training patterns. These patterns, called support vectors, finally describe a classification function, by maximizing the margin their number is minimized. Many researchers applied SVM successfully viz., Arjun Singh et al.[10] , C. V. Jawahar et al.[8], SandhyaArora et al.[2], P. S Deshpande et al.[5], ChanchalBansal et al.[4], SaurabhFarkya et al.[7], for characters recogiton etc.

RESULTS AND DISCUSSIONS

Various handwriting samples are collected and captured/scanned for applying further processing on them. By clicking on the upload image the user can select image of various formats (like . png, .jpg, .jpeg) from the desired location and based on the selected image various operations are performed on the image so that the handwritten characters those are present in the image can be extracted from the image and can be recognized easily by the system. Here we have taken four different cases for the recognition of Multilingual Handwritten Characters Image and finally the output of each case is shown in the notepad file.

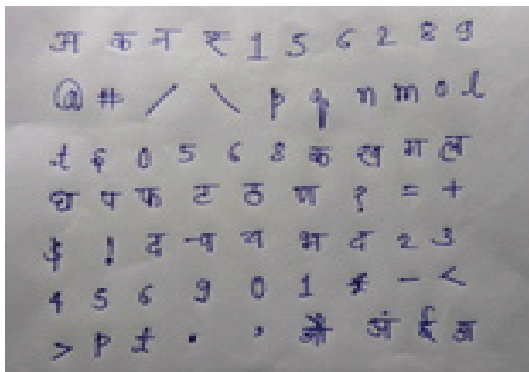Case 1: Multilingual Handwritten Characters Image.
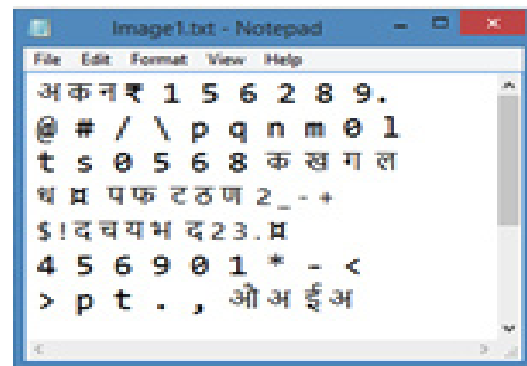

Fig. 1.4: Input Image
Accuracy(%) = 96.96

Fig. 1.5: Recognized Characters

Case 2: Multilingual Handwritten Characters With Skew.

Case 3 :Multilingual Handwritten Characters Large Noisy Image.

Case 4: Large Handwritten Document Image

Fig. 1.6: Input Image
Accuracy(%) = 94.28

Fig. 1.7: Recognized Characters

Fig. 1.8: Input Image
ACCURACY(%) = 92.0

Fig. 1.9: Recognized Characters

Fig. 1.10: Input Image
Accuracy(%) = 96.95

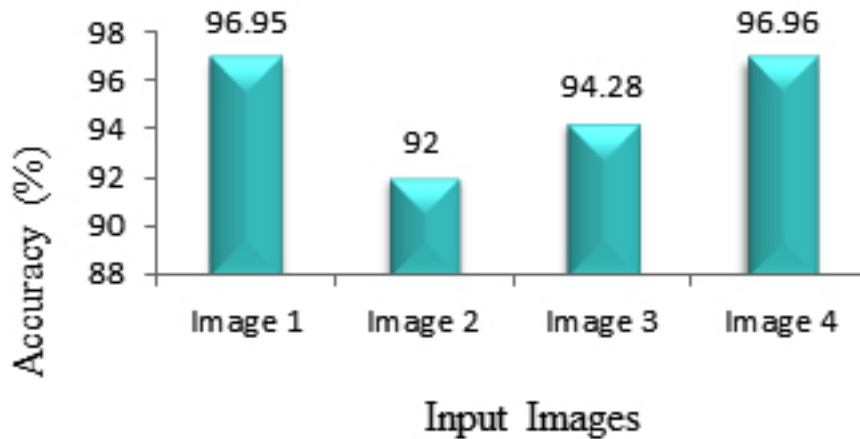
Fig. 1.11: Recognized Characters

Fig. 1.12: Accuracy Grap

The above graphical representation of the accuracy as shown in the figure 1.12 is taken from the values of the accuracy of various images taken in different cases. In this at the x axis the number of images taken for recognition is taken, and on the y axis the accuracy of different images is taken in percentage.

CONCLUSION

While processing the handwritten characters various degradations are found in multilingual handwritten characters scanned/captured images. Some of them are broken lines incomplete characters, different style of writings. Degradations may be due to the defect in papers, writer handwritings, during the capturing or digitization of the image. While processing the characters (like ! ; : " ? = % etc.) we found that for some unknown characters the system recognized them as noise and the noise is depicted as ¤.After the study we found the average accuracy of 95.048 % for Hindi,English,Numerals and special characters in a single captured/scanned. We have taken 50 samples per characters from different writers for training and50 samples per characters for testing with theSVM Classifier. This study can further be implemented using the HOG features with SVM for handwritten script recognition of different languages combined in a single document image.

REFERENCES

1. Arica, N., Fatos, T. and Yarman-Vural. 2001. An overview of character recognition focused on off-line handwriting.*Transactions on systems, man, and cybernetics—part c: applications and reviews*, IEEE, 31:2.
2. Arora, S. 2010. Performance comparison of svm and ann for handwritten devnagari character recognition. *International Journal of Computer Science,*IJCSI,7, 3.
3. Arya, S., C., Singh, R., S. and Mandoria, H., L. 2015. Image Denoising in Hand Written Document for Degraded Documents using Wiener Filter Algorithm.*International Journal for Research in Emerging Science and Technology*, 2, 7.
4. Bansal, C. and Khan, A. 2014. Handwritten numeral recognition using svm and chain code,*IJARET,* 2: VII.
5. Deshpande, P., S., Malik, L., Arora, S., 2008. Recognition of hand written devnagari characters with percentage component regular expression matching and classification tree.*TENCON*, IEEE,2159-3442, 1 – 4.
6. Dogra, S. and Prakash, C. 2012 .Pehchaan: hindi handwritten character recognition

system based on svm. *International Journal on Computer Science and Engineering , IJCSE*,4.

7.   Farkya, s., Surampudi,G. and Kothari, A. 2015. Hindi Speech Synthesis by concatenation of recognized Hand written Devnagri Script using Support Vector Machines Classifier. IEEE *ICCSP conference.*2,4,0893 – 0898.

8.   Jawahar, C.V., Kumar, P. and Kiran, S.S.K.2003. A bilingual ocr for hindi-telugu documents and its applications. *Seventh International Conference on Document Analysis and Recognition (ICDAR)*, 1 ,408 - 412.

9.   Otsu, N.,1979. A threshold selection method from gray-level histograms. *Transactionson Systems, Man, and Cybernetics SMC-9*

IEEE,1,62–66.

10.   Singh,A. and Maring, K., A. 2015. Handwritten devanagari character recognition usingsvm and ann.*International Journal of Advanced Research in Computer and Communication Engineering,4,8.*

11.   Vohra,U.S.,Dwivedi, S.P., Mandoria, H.L.2016. An analytical study of handwritten character recognition.*i-manager's Journal on Pattern Recognition, 2, 4.*

12.   SonalPaliwal, Rajesh Shyam Singh & H. L. Mandoria. A survey on various text detection and extraction techniques from videos and images.*International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*. 6, 3, 1-10.