A Review: Handwritten Character Recognition System

Mahesh R. Barkul[1], S. S. Lokhande[2]

M.E. Student, E&TC, SCOE, Pune, India [1]

Professor, E&TC, SCOE, Pune, India [2]

Abstract: This method presents an innovative technique to deals with the problem of offline handwritten character recognition of English character. This system is very useful to the people which have a lot of paper work on a daily basis. Handwritten character recognition systems provides the solution for how to convert handwritten words into computer readable format. Image enhancement and noise removal are done at the pre-processing phase. After enhancing the image, segmentation of the character will be done using different methods like Horizontal and Vertical Projection Profile method. Feature extraction will be done by using Scale Invariant Feature Transform (SIFT), chain codes, zoning etc. The Support Vector Machine (SVM), Artificial Neural Network (ANN), Euclidian distance methods will be used to recognize the characters. By following above mentioned processes handwritten article images will be converted into an editable text files.

Keywords: Feature extraction, handwritten character recognition, segmentation, SIFT, Support Vector Machine.

I. INTRODUCTION

The purpose of this study is to expose technologies and tools (that are currently openly used in the community) which can be used to recognize handwritten characters. Written characters can be in different shapes and sizes depending on the individual who wrote it. This system will focus on identifying the individual characters contained in those document images in those numerous forms of styles. Handwritten character recognition falls into two categories called Offline character recognition and online character recognition. [1] The general approach towards this problem involves Image Processing techniques. The system mainly focuses on handwriting and offline character recognition. Developing an efficient handwritten character recognition involves a number of challenges. Scanning of the document image may not be perfect which needs special pre-processing methods to be used. When handwriting is concerned different writers have different handwriting styles. Recognizing characters still doesn't achieve 100% accuracy level. There can be errors in identified characters.

The main objective of this system is to provide a perfect solution for how to convert handwritten words into computer readable format. To produce the facility of segmentation, feature extraction and recognition of handwritten characters. To measure performance metric of handwritten character recognition like false acceptance ratio (FAR), true acceptance ratio (TAR), in different complex environment. To recognize words in the article by achieving high accuracy levels. Other objectives are to reduce the waste of time involved in typing articles for newspapers, to produce the facility of editing the article after the conversion by providing an editor, to recognize words in the article by achieving high accuracy levels [1].Character recognition is an art of detecting segmenting and identifying characters from image.

More sophistically character recognition is process of detecting extracting and recognizing characters from input image and converts it into equivalent machine editable form. Character recognition is getting more popular since last ten years due to its huge application range. Conversion of handwritten characters now becomes very important for making different important documents related to our history, such as manuscripts, into machine editable form so that it can be easily accessed and preserved. [3].

There are two different types of character recognition process. Offline character recognition and online character recognition. In offline character recognition system, document is first scanned, binirized, stored in computer and then it is processed. While in case of online character recognition system, processing of character is done while it is under creation. External factors like pressure speed of writing, stroke making etc. does not have any influence in case of offline system but they have great impact on online system.

Accuracy of handwritten character recognition system is not 100 percent due to presence of large variation in shape, scale, style, orientation etc. Character processing systems are domain and application specific, it is not possible to design general system which is used to process all kind of scripts and languages.

Lots of work has been done on European languages and Arabic language. Whereas domestic languages like Hindi, Bangla, Tamil, Marathi etc. are very less explored due to limited usage. In this paper, our focus is to carry out deep study on handwritten character recognition methods.

Any character recognition system goes under following steps, i.e. Image acquisition, Pre-processing, Segmentation, Feature extraction, classification and post processing.

II. LITERATURE SURVEY

D.M.D.S.S Dassanayake proposed Panhinda - Offline Character Recognition System for Handwritten Articles [1]. This system gives an innovative technique to recognize handwritten documents. This system is useful for those people which have a lot of paper work. The proposed handwritten character recognition system was implemented with the capability of extracting the content of an image where the mentioned content is a hand written set of words or characters. The process of conversion runs as a background process without any involvement of the user. Once the conversion is completed, User gets the capability of editing the converted text as he prefers with the aid of the editor. This method describes the techniques for enhancing the quality of the image, character segmentation, character recognition and digital dictionaries. Lighting conditions, noise removal and angle effects are done at the pre-processing phase. This gives quality binarized image. After having a good quality binarized image, character segmentation will be done using Horizontal and Vertical Projection Profile method. The Support Vector Machine technique will be used to classify and recognize the characters. Digital dictionary was used to capture the conflicts of the output. Error correction was done by using a combined model of noisy channel model and natural language model. By following above mentioned processes handwritten article image will be converted into an editable text file. Xiangqian Wu proposed Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform [2].This method proposes a text-independent writer identification method based on scale invariant feature transform (SIFT), which includes three stages of training, enrolment, and identification. In all stages, an isotropic Laplacian of Gaussian (LoG) filter is first used to segment the handwriting image into word regions.

Then, the SIFT descriptors (SDs) of word regions and the corresponding scales and orientations are extracted. In the training stage, an SIFT descriptor codebook is constructed by clustering the SIFT descriptors of training samples. In the enrolment stage, the SDs of the input handwriting are adopted to form an SIFT descriptor signature (SDS) by looking up the SIFT descriptor codebook and the scale and orientations are utilized to generate a scale and orientation histogram (SOH). In the identification stage, the SDS and SOH of the input handwriting are extracted and matched with the enrolled ones for identification. J. Pradeep proposed Performance Analysis of Hybrid Feature Extraction Technique for Recognizing English Handwritten Characters [4]. In this method, an off-line handwritten English character recognition system using neural network classifiers and hybrid feature extraction technique are proposed. A hybrid feature extraction method combines the diagonal and directional based features. The proposed system suitably combines the features of the handwritten characters to increase the recognition accuracy. Neural Network (NN) topologies, namely, back propagation neural network and radial basis function network are used to classify the characters.

The k-nearest neighbor network is also built for comparison. The Feed forward neural network topology exhibits the highest recognition accuracy and is identified to be the most suitable classifier. The proposed system will have applications for postal and parcel address recognition and conversion of any handwritten document into editable text form. Qiu-Feng Wang proposed Handwritten Chinese Text Recognition by Integrating Multiple Contexts [5].This system presents very effective approach for the offline recognition of handwritten Chinese texts. Under the general integrated segmentation-and-recognition framework with character over segmentation, they investigate three important issues: candidate path evaluation, path search, and parameter estimation. For path evaluation, multiple contexts (character recognition scores, geometric and linguistic contexts) are combined from the Bayesian decision view, and convert the classifier outputs to posterior probabilities via confidence transformation. In path search, they used a refined beam search algorithm to improve the search efficiency and, meanwhile, use a candidate character augmentation strategy to improve the recognition accuracy. Abdelhak Boukharouba, Abdelhak Bennia proposed novel feature extraction technique for the recognition of handwritten digits [6]. This system presents an efficient handwritten digit recognition system based on support vector machines (SVM). A novel feature set based on transition information in the vertical and horizontal directions of a digit image combined with the famous Freeman chain code is proposed. This technique does not require any normalization of digits and this is the advantage of this feature extraction algorithm. These features are very simple to implement compared to other methods. They evaluated this scheme on 80,000 handwritten samples of Persian numerals and they have achieved very promising results. Parshuram M. Kamblea, Ravinda S. Hegadi proposed Handwritten Marathi character recognition using R-HOG Feature [7]. They used the Rectangle Histogram Oriented Gradient representation as the basis for extraction of features. To make the system suitable for real-time applications these algorithms require a few simple arithmetic operations per image pixel. Database consists of 40 basic handwritten Marathi characters with 8000 samples of each character. Among these 10 samples of each character from different writers are collected. All sample images of handwritten Marathi characters are normalized to 20×20 pixel size.

The description of the algorithm and experiment with the database is presented in this paper. Experimental results using feed-forward Artificial Neural Network (FFANN) classification and Support Vector Machines (SVM) techniques are presented. Results demonstrate high performance of these features when classified using feed-forward Artificial Neural network, classification.

III.METHODOLOGY

Block diagram is of general character recognition system is shown in fig. 1.These steps are required for all character recognition systems.

A. *Image acquisition:*

Images for handwritten character recognition system can be acquired capturing photograph of document or by scanning handwritten document. This is also known as digitization process.

B. *Pre-processing:*

Pre-processing involves series of operations performed to enhance the quality of image to make it suitable for segmentation. It involves noise removal generated during document scanning or while capturing the photograph. Proper filters like Gaussian filter mean filter, min-max filter, etc. can be applied to remove noise from document. Binarization process converts gray scale or colored image to black and white image. Binary morphological operations like closing, opening, thinning, etc. may be applied to enhance the visibility and structural information of character. If document is scanned then it may not be always horizontally aligned, so we need to align it by performing slant angle correction. Sometimes input document may be too large so it may be resized in smaller size s to improve speed of processing. However reducing dimension below some specific level may remove some useful features too.
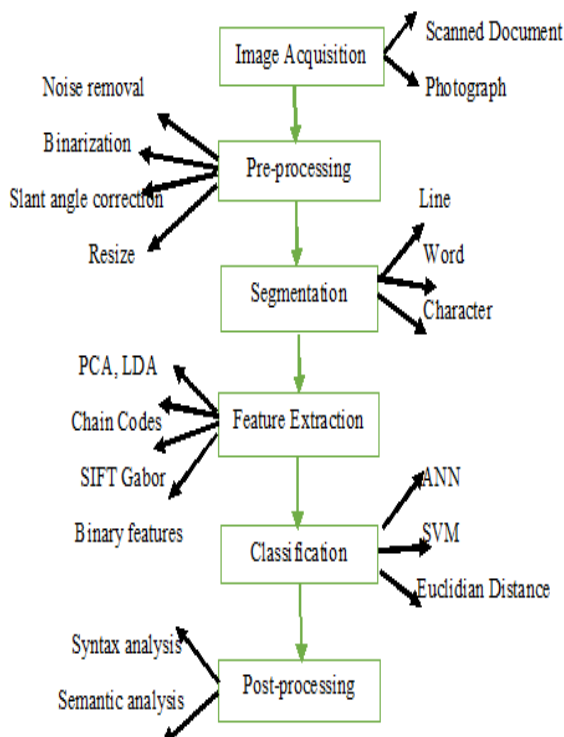

Fig. 1 Block diagram of character recognition system [3]

C. *Segmentation:*

Generally document is processed in hierarchical way. At first level row histogram is used for segmentation of lines. After extracting the lines words are extracted using column histogram and then characters are extracted from words. Accuracy of handwritten character recognition system is highly depends on accuracy of segmentation.

D. *Feature Extraction:*

Feature extraction is very important part of any handwritten character recognition system. Feature extraction techniques like Scale Invariant Feature Extraction (SIFT), [8] Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), zoning, Gradient based features Principle Component Analysis (PCA), Chain Code (CC), can be used to extract the features of individual characters. These extracted features are used to train the system.

E. *Classification:*

When input image is presented to handwritten character recognition system, its features are extracted and given as an input to the different classifier like artificial neural network or support vector machine. Classifiers are used to compare the input feature with stored pattern and find out the proper matching class for the given input. For classification SVM, [9] ANN and Euclidian distance techniques are used.

F. *Post processing:*

This step is not compulsory, but sometimes it helps to improve the accuracy of recognition

IV. CONCLUSION

A solution for offline handwritten character recognition is presented by different systems. This work can be extended from recognizing single character or set of character to document processing. Post processing is also crucial to differentiate similar structured characters like M and N, Q and O like pairs in English language. To design fully automated system which can be used to handle all kind of variability is not possible. Handwritten character recognition is very important research area of pattern recognition. Each and every step directly influence to the accuracy of system, like feature extraction, segmentation, pre-processing, training methods etc. all. So all these areas are open for independent research. Many more things can be improved in each of these steps. Hybrid model can be implemented in which more than one feature extraction methods can be used to discriminate characters properly. Developing a mobile application, enabling the successful recognition for cursive handwriting and implementing the system which enables to run it on any Operating System are the future objectives of the system.

REFERENCES

[1] D.M.D.S.S Dassanayake, R.A.D.D Yasara, H.S.R. Fonseka, E.A Heshan Sandeepa and L.Seneviratne "Panhinda - Offline Character Recognition System for Handwritten Articles" Sri Lanka Institute of Information Technology, Sri Lanka, 2013

[2] Xiangqian Wu, Youbao Tang, and Wei Bu "Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform" ieee   transactions on information forensics and security, vol. 9, no. 3, March 2014

[3] Mansi Shah and Gordhan B Jethava "A Literature Review on Hand Written Character Recognition" Indian streams research journal vol -3, issue –2, march.2013

[4] J.Pradeep*, E.Srinivasan*, S.Himavathi "Performance Analysis of Hybrid Feature Extraction Technique for Recognizing English Handwritten Characters" 978-1-4673-4805-8/12 2012 IEEE

[5] Qiu-Feng Wang, Fei Yin, and Cheng-Lin Liu "Handwritten Chinese Text Recognition by Integrating Multiple Contexts" ieee transactions on pattern analysis and machine intelligence, vol. 34, NO. 8, AUGUST 2012.

[6] Abdelhak Boukharouba, Abdelhak Bennia   "Novel feature extraction technique for the recognition of handwritten digits" Saudi Computer Society, king Saud University Applied Computing and informatics 2015

[7] Parshuram M. Kamblea, Ravinda S. Hegadi "Handwritten Marathi character recognition using R-HOG" International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

[8] David G. Lowe "Distinctive Image Features from Scale-Invariant Keypoints" Computer Science Department University of British Columbia Vancouver, B.C., Canada January 5, 2004

[9] Dino Isa, Lam Hong Lee, V.P. Kallimani, and R. RajKumar, "Text Document Preprocessing with the Bayes Formula for Classification Using the Support Vector Machine" IEEE transactions on knowledge and data engineering, vol. 20, no. 9, September 2008

BIOGRAPHIES

Mahesh R. Barkul, received the B.E. degree in Electronics & Telecommunication engineering from VPCOE Baramati, Pune, India, in 2013. He is currently pursuing the M.E. in Digital Systems from SCOE Pune, India. His current research interest includes image processing and pattern recognition.

Dr. (Mrs.) S. S. Lokhande, is currently a professor in SCOE, Pune. Her current research interest includes neural network, signal processing and devices & circuits.