

ABSTRACT:

Handwriting recognition has gained a lot of attention in the field of pattern recognition and machine learning due to its application in various fields. Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR) has specific domain to apply. Various techniques have been proposed to for character recognition in handwriting recognition system. Even though, sufficient studies and papers describes the techniques for converting textual content from a paper document into machine readable form. In coming days, character recognition system might serve as a key factor to create a paperless environment by digitizing and processing existing paper documents. This paper presents a detailed review in the field of Handwritten Character Recognition.

INTRODUCTION:

Character recognition is a fundamental, but most challenging in the field of pattern recognition with large number of useful applications. It has been an intense field of research since the early days of computer science due to it being a natural way of interactions between computers and humans. More precisely Character recognition is the process of detecting and recognizing characters from the input image and converts it into ASCII or other equivalent machine editable form [1][2].

The technique by which a computer system can recognize characters and other symbols written by hand in natural handwriting is called handwriting recognition system. Handwriting recognition is classified into offline handwriting recognition and online handwriting recognition [3]. If handwriting is scanned and then understood by the computer, it is called offline handwriting recognition. In case, the handwriting is recognized while writing through touch pad using stylus pen, it is called online handwriting recognition. From the classifier perspective, character recognition systems are classified into two main categories i.e. segmentation free (global) and segmentation based (analytic). The segmentation free also known as the holistic approach to recognize the character without segmenting it into subunits or characters. Each word is represented as a set of global features, e.g. ascender, loops, cusp, etc. Whereas segmentation based approach [4]; each word/ligature is segmented into subunits either uniform or non-uniform and subunits are considered independently.

Handwritten character processing systems are domain and application specific, like it is not possible to design a generic system which can process all kinds of handwritten scripts and language. Lots of work has been done on European languages and Arabic (Urdu) language. Whereas domestic languages like Hindi, Punjabi, Bangla, Tamil, Gujarati etc. are very less explored due to limited usage.

In this paper, the section II describes the basic working principle of character recognition followed by a detailed literature survey. Next, in section IV conclusion has been made.

WORKING PRINCIPLE:

Normally handwritten recognition is divided into six phases which are image acquisition, pre-processing, segmentation, feature extraction, classification and post processing. The block diagram of the basic character recognition is shown in fig1.

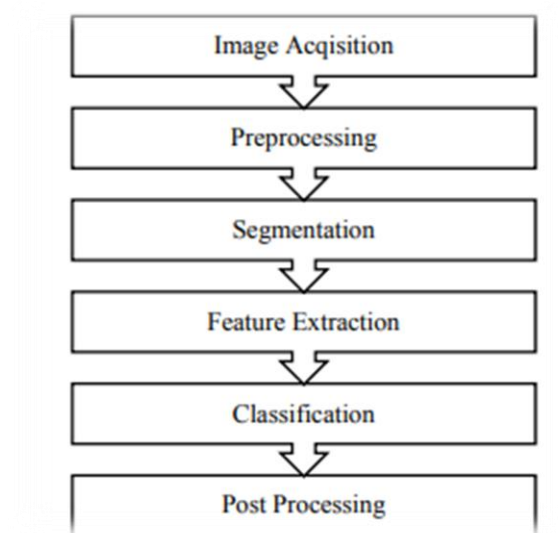


Figure 1: Block Diagram of Character Recognition

A. Image Acquisition

Digitized/Digital Image is initially taken as input. The most common of these devices is the electronic tablet or digitizer. These devices use a pen that is digital in nature. Input images for handwritten characters can also be taken by using other methods such as scanners, photographs or by directly writing in the computer by using a stylus.

B. Preprocessing :

Pre-processing is the basic phase of character recognition and it's crucial for good recognition rate. The main objective of pre-processing steps is to normalize strokes and remove variations that would otherwise complicate recognition and reduce the recognition rate. These variations or distortions include the irregular size of text, missing points during pen movement collections, jitter present in text, left or right bend in handwriting and uneven distances of points from neighbouring positions. Pre-processing includes five common steps, namely, size normalization and centering, interpolating missing points, smoothing, slant correction and resampling of points.

C.Segmentation:

Segmentation is done by separation of the individual characters of an image. Generally document is processed in a hierarchical way. At first level lines are segmented using row histogram. From each row, words are extracted using column histogram and finally characters are extracted from words.

D. Feature Extraction:

The main aim of feature extraction phase is to extract that pattern which is most pertinent for classification. Feature extraction techniques like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Chain Code

(CC), Scale Invariant Feature Extraction (SIFT), zoning, Gradient based features, Histogram might be applied to extract the features of individual characters. These features are used to train the system.

E. Classification:

When input image is presented to HCR system, its features are extracted and given as an input to the trained classifier like artificial neural network or support vector machine. Classifiers compare the input feature with stored pattern and find out the best matching class for input.

F. Post processing:

Post-processing refers to the procedure of correcting misclassified results by applying linguistic knowledge. Postprocessing is processing of the output from shape recognition. Language information can increase the accuracy obtained by pure shape recognition. For handwriting input, some shape recognizers yield a single string of characters, while others yield a number of alternatives for each character, often with a measure of confidence for each alternative.

III. LITERATURE REVIEW

An early notable attempt in the area of character recognition research is by Grimsdale in 1959. The origin of a great deal of research work in the early sixties was based on an approach known as analysis-by-synthesis method suggested by Eden in 1968. The great importance of Eden's work was that he formally proved that all handwritten characters are formed by a finite number of schematic features, a point that was implicitly included in previous works. This notion was later used in all methods in syntactic (structural) approaches of character recognition.

K. Gaurav, Bhatia P. K. [5] Et al, this paper deals with the various pre-processing techniques involved in the character recognition with different kind of images ranges from a simple handwritten form based documents and documents containing colored and complex background and varied intensities. In this, different preprocessing techniques like skew detection and correction, image enhancement techniques of contrast stretching, binarization, noise removal techniques, normalization and segmentation, morphological processing techniques are discussed. It was concluded that using a single technique for preprocessing, we can't completely process the image. However, even after applying all the said techniques might not possible to achieve the full accuracy in a preprocessing system.

Salvador España-Boquera et al [6], in this paper hybrid Hidden Markov Model (HMM) model is proposed for recognizing unconstrained offline handwritten texts. In this, the structural part of the optical model has been modelled with Markov chains, and a Multilayer Perceptron is used to estimate the emission probabilities.

In this paper, different techniques are applied to remove slope and slant from handwritten text and to normalize the size of text images with supervised learning methods. The key features of this recognition system were to develop a system having high accuracy in preprocessing and recognition, which are both based on ANNs.

In [7], a modified quadratic classifier based scheme to recognize the offline handwritten numerals of six popular Indian scripts is proposed.

Multilayer perceptron has been used for recognizing Handwritten English characters [8]. The features are extracted from Boundary tracing and their Fourier Descriptors. The character is identified by analysing its shape and comparing its features that distinguish each character. Also an analysis has been carried out to determine the number of hidden layer nodes to achieve high performance of the back propagation network. A recognition accuracy of 94% has been reported for Handwritten English characters with less training time.

M. Hanmandlu, O.V. Ramana Murthy[20] have presented in their study the recognition of handwritten Hindi and English numerals by representing them in the form of exponential membership functions which serve as a fuzzy model. The recognition is carried out by modifying the exponential membership functions fitted to the fuzzy sets. These fuzzy sets are derived from features consisting of normalized distances obtained using the Box approach. The membership function is modified by two structural parameters that are estimated by optimizing the entropy subject to the attainment of membership function to unity. The overall recognition rate is found to be 95% for Hindi numerals and 98.4% for English numerals.

TABLE I
COMPARISON BETWEEN DIFFERENT TECHNIQUES

Method	Accuracy	Purpose
Hand printed symbol recognition. ^[5]	97% overall.	Extract the geometrical, topological and local measurements required to identify the character.
OCR for cursive handwriting. ^[6]	88.8% for lexicon size 40,000.	To implement segmentation and recognition algorithms for cursive handwriting.
Recognition of handwritten numerals based upon fuzzy model. ^[14]	95% for Hindi and 98.4% for English numerals overall.	The aim is to utilize the fuzzy technique to recognize handwritten numerals for Hindi and English numerals.
Combining decision of multiple connectionist classifiers for Devanagari numeral recognition. ^[7]	89.6% overall.	To use a reliable and an efficient technique for classifying numerals.
Hill climbing algorithm for handwritten character recognition. ^[10]	93% for uppercase letters.	To implement hill climbing algorithm for selecting feature subset.
Optimization of feature selection for recognition of Arabic characters. ^[11]	88% for numbers and 70% for letters.	To apply a method of selecting the features in an optimized way.
Handwritten numeral recognition for six popular Indian scripts. ^[12]	99.56% for Devanagari, 98.99% for Bangla, 99.37% for Telugu, 98.40% for Oriya, 98.71% for Kannada and 98.51% for Tamil overall.	To find out the recognition rate for the six popular Indian scripts.

Conclusion:

The paper discusses in detail all advances in the area of handwritten character recognition. The most accurate solution provided in this area directly or indirectly depends upon the quality as well as the nature of the material to be read. Various techniques have been described in this paper for character recognition in handwriting recognition system. A sort comparison is shown between the different methods proposed so far in table 1. From the study done so far, it is analysed that the selection of the classification as well as the feature extraction techniques needs to be proper in order to attain good rate in recognizing the character. Studies in the paper reveals that there is still scope of enhancing the algorithms as well as enhancing the rate of recognition of characters.