Character Recognition of Offline Handwritten English Scripts: A Review

V. S. Dhaka
School of Engineering and Technology, Jaipur National University, Jaipur, India
vijaypal.dhaka@gmail.com
Manoj Kumar
School of Engineering and Technology, Jaipur National University Jaipur, India
manoj186@gmail.com
Harshit Sharma
School of Engineering and Technology, Jaipur National University Jaipur, India
harsh.sharma198@gmail.com

----------ABSTRACT----------

Character recognition is a process by which computer recognizes letters, numbers or symbols and turn them into digital form that a computer can use. In today's environment character recognition has gained lot of concentration in the field of pattern recognition. Handwritten character recognition is useful in cheque processing in banks,form processing systems and many more. Character recognition is one of the well-liked and challenging area of research. In future character recognition create paperless environment. In this paper we describe the detail study on existing method for handwritten character recognition. We provide a literature review on various techniques used in offline english character recognition.

***Keywords*: Character; Character recognition; Preprocessing; Segmentation; Feature extraction; Classification; neural network; Convolution neural network.**

I. INTRODUCTION

Handwriting is one of the most important mean in daily communication. During the last years, many popular studies and applications merged for bank check processing, mailed envelops reading, and handwritten text recognition in documents and videos[1]. Charcter recognition is a process to identify individual printes symbol. These symbol may be alphabatic, numeric, punctuation etc. These symbols may be printed either by hand or mechanically in a variety of different font and sizes. More precisely character recognition is process of detecting and recognizing character from input image and convert it into ASCII or other equivalent machine editable form[2].

The task of recognition broadly separated into two categories:handwritten data and machine printed data. Machine printed character are uniform and unique. While handwritten character are non uniform and their size, shape depends on the pen used by the writer. Handwriting of same person also may vary depending upon situation in which person writing. Various writing styles lead to the distortion in patterns from the standard patterns used to train the system, giving fake results. So it is difficult task to design system,which is capable to identifying the character with great accuracy.

Character recognition system is classified into two categories:
(a) Offline character recognition
(b) Online character recognition

In offline character recognition system document is first generated, digitized, store in computer and then it is processed. It is not real time system. In online character recognition character is processed while it was under creation. In handheld devices online character recognition is performed. Online character recognition is real time system.

Table 1: Comparasion of online and offline

S.No	Comparasion	Online Character	Offline Character
1.	Availability of no. of pen stroke	Yes	No
2.	Raw data required	Sample/sec	Dots/inch
3.	Way of writing	Using digital pen on LCD	Paper document
4.	Recognition Rate	Higher	Lower
5.	Accuracy	Higher	Lower
6.	Real time contextual information	Yes	No

Off-line handwritten character recognition refers to the process of recognizing characters in a document that have been scanned from a surface such as a sheet of paper and are stored digitally in gray scale format. The storage of scanned documents have to be bulky in size and many

processing applications as searching for a content, editing, maintenance are either hard or impossible. Such documents require human beings to process them manually, for example, postman's manual processing for recognition and sorting of postal addresses and zip code. Character recognition systems translate such scanned images of printed, typewritten or handwritten documents into machine encoded text. This translated machine encoded text can be easily edited, searched and can be processed in many other ways according to requirements. It also requires tinny size for storage in comparison to scanned documents.

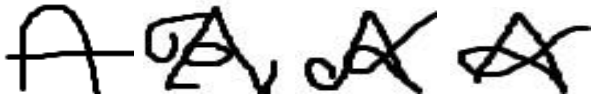

Fig.1. Different Style Of Writing

The online mode of recognition is mostly used to recognize only handwritten characters. In this the handwriting is captured and stored in digital form via different means. Usually, a special pen is used in conjunction with an electronic surface. As the pen moves across the surface, the two- dimensional coordinates of successive points are represented as a function of time and are stored in order. Recently, due to increased use of handheld devices online handwritten recognition attracted attention of worldwide researchers. This online handwritten recognition aims to provide natural interface to users to type on screen by handwriting on a pad instead of by typing using keyboard. The online handwriting recognition has great potential to improve user and computer communication.

In online handwriting recognition, it is very natural for the user to detect and correct misrecognized characters on the spot by verifying the recognition results as they appear. The user is encouraged to modify his writing style so as to improve recognition accuracy. Also, a machine can be trained to a particular user's style. Samples of his misrecognized characters are stored to aid subsequent recognition. Thus both writer adaptation and machine adaptation is possible. Accuracy of offline charater recognition is not 100 percent due to occurrence of large variation in shape, scale, style, orientation etc. The online methods have been shown to be superior to their offline counterparts in recognizing handwritten characters due to the temporal information available with the former [3].

A typical handwritten charcter recognition system have following steps:

1. Data acquisition
2. Pre-processing
3. Segmentation
4. Feature extraction
5. Classifier
6. Post-processing

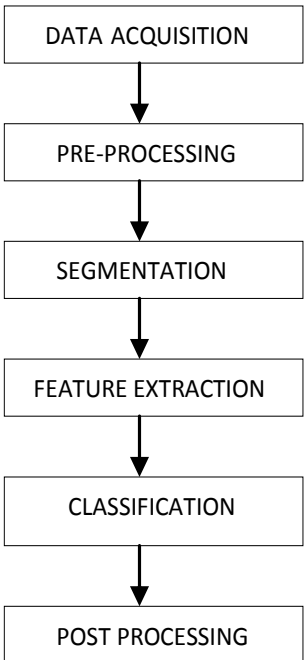

Fig.2. Flow Chart Of Handwritten Charcter Recognition

Charcter recognition getting more and more attention due to its wide range of application. Character processing systems are domain and application specific, like there is not possible to design a generic system which can process all time of languages and scripts. Lot of work done on Europian languages and Arabic language. Whereas domestic languages like Hndi, Punjabi, Marathi, Tamil, Telgu, Bengla, Gujrati etc. are very less explored due to its limited usage. In this paper we discussed English charcter recognition.

II. METHODOLOGIES FOR OFFLINE CHARACTER RECOGNITION SYSTEM

In this section of paper the existing methodologies to develop the stages of the character recognition system are presented.

A. DATA AQUISITION

The automatic character recognition systems advancement is evolved in two categories according to the approach of data acquisition:

On-line character recognition systems
Off-line character recognition systems

The character recognition off-line, captures the statistics from document through optical scanners or cameras whereas the recognition systems on-line, make use of the digitizers which directly captures writing through the order of the strokes, speed, pen- up and pen- down information.

B. PRE-PROCESSING

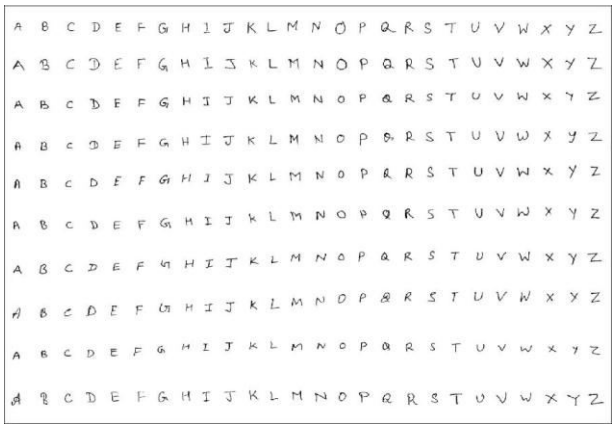

Fig.3. Character Set Of Handwritten Charcter

The raw data, depending on the data acquisition type, is subjected to a number of preliminary processing steps to make it usable in the descriptive stages of character analysis [4]. In above figure a set of handwritten character is taken. In preprocessing image is taken and is converted to gray scale image. The gray scale image is then converted to binary image. This process is called Digitization of image.

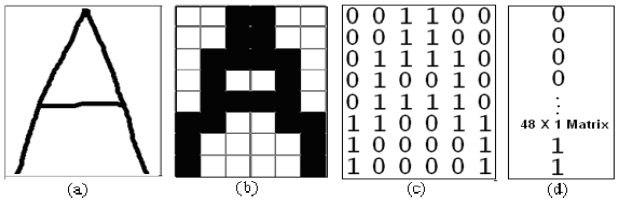

Fig.4. Digitization Process

Basically any scanner is not perfect; the scanned image may have some noise. This noise may be due to some unnecessary information available in the image. Various steps are following in preprosessing technique:

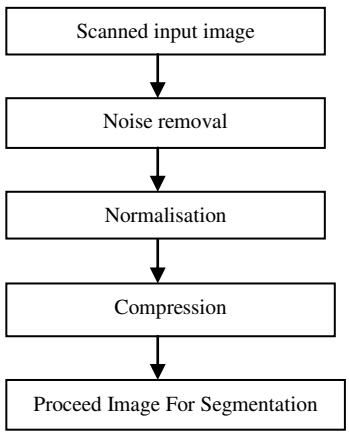

Fig.5. Preprocessing Step

The main objective of pre-processing is to denoise and enhance the quality of scanned digit image.

Noise removal: Optical scanning devices introduces some noises like, disconnected line segments, bumps and gaps in lines, filled loops etc. It is necessary to remove all these noise elements prior to the character recognition been developed by various researchers [22],[23].

Normalisation: The main component of the pre-processing stage is normalization, which attempts to remove some of the variations in the images, which do not affect the identity of the word. Handwritten image normalization from a scanned image includes several steps, which usually begin with image cleaning, skew correction, line detection, slant and slope removal and character size normalization [24].

Compression: Space domain techniques are required for compression. Two important techniques are thresholding and thinning [25, 26]. Thresholding reduces the storage requirements and increases the speed of processing by converting the gray-scale or colour images to binary image by taking a threshold value. Thinning extracts the shape information of the characters.

C. SEGMENTATION

In the segmentation stage, an image consisting of a sequence of characters is decomposed into sub-images of individual characters [5]. The main goal is to divide an image into parts that have a strong correlation with objects or areas of the real world contained in the image. Segmentation is very important for recognition system. Segmentation is an important stage because the extent one can reach in separation of words, lines, or characters directly affects the recognition rate of the script. Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. In Character Recognition techniques, the Segmentation is the most important process. Segmentation is done to make the separation between the individual characters of an image. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. Script segmentation is done by executing the following operations: Line segmentation, Word segmentation and character
segmentation.

D. FEATURE EXTRACTION

Feature extraction is the method to retrieve the most significant data from the raw data. The main objective of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements. Feature extraction is the heart of any pattern recognition application. Feature extraction techniques like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), 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. The features extracted are given as input to the classification stage and the output of this stage is a recognised character. The selection of the combination of feature-classifier contributes to the performance of the system. Several research works has been focussing toward evolving such methods to reduce the processing time and providing higher recognition .

Feature extraction based on three types of feature:

1.) Statistical Feature

These features are derived from statistical distribution points. They provide high speed and low complexity and take care of style variation. Zoning, characteristic loci, crossing and distance are the main statistical features.

2.) Structural Features

Structural features are based on topological and geometrical properties of character, such as, aspect ratio, cross point, loops, branch points, strokes and thire directions, inflection between two points, horizontal curve on top or bottom, etc. The representation of this type may also encode some knowledge about the structure of the object or may provide some knowledge as to what sort of components make up that object.

3.) Global transformation and series expansion

A continuous signal generally contains more information than needs to be represented for the purpose of classification[4]. This may be true for discrete approximations of continuous signals as well. One way to represent a signal is by a linear combination of a series of simpler well-defined functions. The coefficients of the linear combination give a compact encoding known as transformation or/and series expansion. Deformations like translation and rotations are invariant under global transformation and series expansion. Gabor transformation, Fourier transformation and wavelet transformation are common transform and series expansion method used in character recognition method.

E. CLASSIFICATION AND RECOGNITION

The classification phase is the decision making part of the recognition system. The performance of a classifier based on the quality of the features. This stage uses the features extracted in the previous stage to identify the character. 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.

There are four classification method:

a.) Template Matching: The simplest way of character recognition is based on matching the stored prototypes against the character or word to be recognized. Generally speaking, matching operation determines the degree of similarity between two vectors (group of pixels, curvature, shapes etc.) in the feature space. Matching techniques can be studied in three classes Direct Matching, Deformable Templates and Elastic Matching and Relaxation Matching.

b.) Statisticle Techniques: Statistical decision theory is concerned with statistical decision functions and a set of optimality criteria, which maximizes the likelihood of the observed pattern given the model of a certain class. Statistical techniques are, mainly, based on three major assumptions:

i. Feature set distribution is Gaussian or in the worst case uniform,
ii. There are sufficient statistics available for each class,
iii. Given a group of images {I}, one is able to extract a set of features {fi}∈ F, i ∈ {1,...,n}, which represents each distinct class of patterns. The measurements taken from n-features of every word unit can be thought to represent an n-dimensional vector space and the vector, whose coordinates correspond to the taken measurements, shows the original word unit.

c.) Structural Techniques: The recursive description of a complex pattern in terms of simpler patterns based on the shape of the object was the initial idea behind the creation of structural pattern recognition. These patterns are used to describe and classify the characters in the CR systems.

d.) Neural Networks: A neural network is defined as a computing architecture that consists of massively parallel interconnection of adaptive 'neural' processors. Because of its parallel nature, it can perform computations at a higher rate compared to the classical techniques. It can easily adapt to changes in the data and learn the input signals characteristics, because of its adaptive nature. A neural network is consists of many nodes. The output from one node is fed to another one in the network and the final decision depends on the complex interaction of all nodes. In spite of the different underlying principles, it can be shown that most of the neural network architectures are equivalent to statistical pattern recognition methods [20]. Several approaches exist for training of neural networks [21]. These include the boltzman, error correction, hebbian and competitive learning. They cover binary and continuous valued input, in addition with supervised and unsupervised learning. On the other hand, neural network architectures can be classified into two major groups, namely, feed-forward and feedback (recurrent) networks. The most familiar neural networks used in the CR systems are the multilayer perceptron of the feed forward networks and the kohonen's Self Organizing Map (SOM) of the feedback networks.

F.POST-PROCESSING

Post-processing stage is the last stage of the proposed recognition system. It prints the corresponding recognized characters in the structured text form. System results

usually contain errors because of character classification and segmentation problems. For the correction of recognition errors, CR systems apply contextual post-processing techniques. The two most common post-processing techniques for error correction are dictionary lookup and statistical approach. The advantage of statistical approach over dictionary-based methods is computational time and memory utilization.The simplest way of incorporating the context information is the utilization of a dictionary for correcting the minor mistakes.

III. DESIGN OF CHARCTER RECOGNITION

Various approaches used for the design of OCR systems are discussed below:

MATRIX MATCHING : Matrix Matching converts each character into a pattern within a matrix, and then compares the pattern with an index of known characters. Its recognition is strongest on monotype and uniform single column pages.

FUZZY LOGIC**:** Fuzzy logic is a multi-valued logic that allows intermediate values to be defined between conventional evaluations like yes/no, true/false, black/ white etc. An attempt is made to attribute a more human-like way of logical thinking in the programming of computers. Fuzzy logic is used when answers do not have a distinct true or false value and there is uncertainly involved.

FEATURE EXTRACTION**:** This method defines each character by the presence or absence of key features, including height, width, density, loops, lines, stems and other character traits. Feature extraction is a perfect approach for OCR of magazines, laser print and high quality images.

STRUCTURAL ANALYSIS**:** Structural Analysis identifies characters by examining their sub featuresshape of the image, sub-vertical and horizontal histograms. Its character repair capability is great for low quality text and newsprints.

NEURAL NETWORK**:** This strategy simulates the way the human neural system works. It samples the pixels in each image and matches them to a known index of pixel patterns. The ability to recognize characters through abstraction is great for faxed documents and damaged text. Neural networks are ideal for specific types of problems, such as processing stock market data or finding trends in graphical patterns. Neural network is most efficient network.

IV. COMPARATIVE ANALYSIS OF CHARACTER RECOGNITION

Table 2: Comparative Analysis Of Character Recognition

S. No	Author	Method	Classifier	Accuracy (%)
1.	Anshul Gupta, Manisha Srivastava and Chitralekha Mahanta [6]	Support vector machine	Neural Network	98.86(training)
				62.93(test)
2.	Aiquan Yuan, Gang Bai, Lijing Jiao and Yajie Liu [7]	LetNet 5	Convolutional neural network	93.7(upper case)
				90.2(lower case)
3.	J.Pradeep, E.Srinivasan and S.Himavathi [5]	Hybrid feature extraction	Feed forward	95.96
			NN Radial base function	93.82
			NN Nearest Neighbour	91.88
4.	N.M Noor,M.Razaz and P.Mahley [9]	Geometry extraction	Geometric density	77.89
			Geometric feature	76.44
5.	Rajib Lochan Das, Binod Kumar Prasad and Goutam Sanyal [11]	Local and global feature extraction	HMM	98.26
6.	Rakesh Kumar Mandal and N R Manna[13]	Row wise segmentation	Single ANN	80
7.	Huiqin Lin,Wennuan Ou,Tonglin Zhu [14]	Direction element feature	Correcting Assignment Algo	99.03
8.	D.K.Patel,T.Som,Manoj Kumar Singh [15]	Ecludian distance	ANN	92.31
9.	Huihang Zaho,Dejian Zhou,Zhaohua Wu [16]	Surface mount technology	BP neural network	98.6
10.	Peng Xu[18]	Particle swarm method	PSO-BP neural network	86.8(capital)
				85.3(small)
11.	P.M. Patil ,P.S. Dhabe, U.V. Kulkarni and T.R. sontakke [19]	Hyperline segment	FHLSSN	72.10
			MFHLSSN	72.55

In above table we show some previous work related to offline handwritten English character recognition. The work perform by researcher is good in many cases but there is always a scope to perform better.The brief review of character recognition which is done by researcher in past is given:

Anshul Gupta, Manisha Srivastava and Chitralekha Mahanta [6] focus especially on offline recognition of handwritten english words by first detecting individual

characters. The main approaches for offline handwritten word recognition can be divided into two classes, holistic and segmentation based. They first segment the words into individual characters and then represent these characters by features that have good discriminative abilities. They also explore different neural network classifiers to find the best classifier for the CR system. They combine different CR techniques in parallel so that recognition accuracy of the system can be improved.

The organization of the paper is as follows: Firstly deals with segmentation of words into individual characters where a heuristic algorithm is used to first oversegment the word followed by verification using neural network.
i.)Segmentation using a heuristic algorithm
ii.) Manual marking of segmentation points
iii.) Training of the Artificial Neural Network (ANN)
iv.) Testing phase of the segmentation technique

Feature extraction of handwritten characters is discussed in this paper. After this describes selection procedure of a suitable classifier. This is done by testing multilayer perceptron (MLP), radial basis function (RBF) and support vector machine (SVM) and selecting the one that has the maximum accuracy. Then post processing is discussed where different character recognition techniques are combined in parallel by using a variation of the Borda count. In this paper support vector machine (SVM) as classifier and achieved 98.86% classification accuracy on the training data set and 62.93% on the test data set.

Aiquan Yuan, Gang Bai, Lijing Jiao, Yajie Liu [7] applies convolution neural network technique for handwritten english character recognition. They use a modified LeNet-5 CNN model, with special settings of the number of neurons in each layer and the connecting way between some layers. Outputs of the CNN are set with error-correcting codes, thus the CNN has the ability to reject recognition results. For training of the CNN, an error-samples-based reinforcement learning strategy is developed. Experiments are evaluated on UNIPEN lowercase and uppercase datasets, with recognition rates of 93.7% for uppercase and 90.2% for lowercase, respectively. This paper focuses mainly on offline HECR on UNIPEN dataset [8], with 26 characters for uppercase and lowercase, respectively. Section II shows the common *LeNet-5* model of CNNs and the modifications we make on it. A training strategy based on reinforcement of *error-samples* are described in this paper. In this paper when we use LetNet-5 mathod using convolution neural network as classifier get accuracy 90.20%.

J.Pradeep E.Srinivasan and S.Himavathi [5] uses feature extraction technique for offline english handwritten method. In this paper, an off-line handwritten English character recognition system using hybrid feature extraction technique and neural network classifiers are proposed. A hybrid feature extraction method combines the diagonal and directional based features. In this paper, a hybrid feature extraction scheme by combining two zonal based approaches, namely, diagonal and directional is proposed. The features obtained are used to train neural network based classifiers such as feed forward network and radial basis function network. A comparison is carried out with the nearest neighbour classifier. The hybrid features are used to train to Neural Network based classifiers and the results obtained are presented. The best recognition system is identified and the experimental results are presented and discussed. A Hybrid feature extraction based off-line handwritten character recognition system with different classifiers namely, Feed forward NN, radial basis function NN and nearest neighbour network for recognizing handwritten English alphabets is proposed. A hybrid feature extraction technique, combining two different approaches namely, diagonal based feature extraction and directional based feature extraction is used. The different classifiers have been trained with 200 sets of 26 alphabets and tested extensively. Experimental results show that the feed forward neural network is distinctly superior to the other classifiers in recognizing the handwritten English alphabets. The Feed forward classifier is found to exhibit an average recognition accuracy of 95.96% and a worst case accuracy of 88%. Here we use feed forward neural network, radial basis function neural network and nearest neighbour neural network classifiers. Among these three classifiers feed forward neural network classifiers show better accuracy than others. So we prefer feed forward neural network classifier.

The problem of handwritten character recognition is made more difficult by the fact that scanned, isolated characters are rotated or slanted in either direction and by a variable amount. This paper [9] provides a possible solution to the problem of character rotation in the offline recognition of isolated handwritten English alphabets. The geometric properties will be combined with other feature extraction techniques to produce input vectors from which to classify images. This technique was incorporated into two classifiers, namely a density classifier, and a feature classifier [10]. Fuzzy Logic inference engines are used for classification as they have the ability to represent uncertainty and imprecision. Feature extraction can take place after the character image has been pre-processed. Geometric features based on multi-stroke recognizer techniques are extracted to collect the global features of the input image. These features are able to provide a classifier with helpful data about the character image. The geometric features may be included with local features, such as pixel density or line and point distribution, as the input vector to a classifier. Geometric features are extracted and then used to calculate polygons. To select features that best identify the character shape some ratios of perimeters and areas of the polygons are used. Characters such as O and Q fall into the circle shapes and the first feature ratio, P2 ch /Ach is used to distinguish this group from other character shapes as this ratio has the lowest values compared to other shapes. Characters I, l and i fall into the line group and are identified by comparing the height of the enclosing rectangle Her with its width Wer. The Her/Wer ratio for the line group has values greater than other shape groups. The

rest of the features are selected using similar methods of analysing ratio values to identify unique traits for each shape group.

There are two classifiers are used density classifier and feature classifier. The pixel density classifier measures the black pixel distribution of the input image. Subdividing the character image, the ratio of black to white pixels is calculated, the results being inserted into a vector. The amount of line direction, both horizontal and vertical, in each stripe is also determined and this data, along with the horizontal vertical symmetry ratio are inserted into the vector. The feature classifier is based upon the system in. It takes normalised, thinned input images and the features measured are the number and positions ofjunctions, crosses, and endpoints. The data is inserted into a vector together with the parameters from geometric feature extraction. Geometry density classifier give 77.89% accuracy. Geometry Feature classifier give 76.44% accuracy

Rajib Lochan Das, Binod Kumar Prasad and Goutam Sanyal [11] use local and global feature extraction based on HMM. Recognition rate of handwritten character is still limited around 90 percent due to the presence of large variation of shape, scale and format in hand written characters. A sophisticated hand written character recognition system demands a better feature extraction technique that would take care of such variation of hand writing. In this paper, a recognition model based on multiple Hidden Markov Models (HMMs) followed by few novel feature extraction techniques for a single character to tackle its different writing formats. We also propose a post-processing block at the final stage to enhance the recognition rate further. Using thier proposed recognition system they have achieved a good average recognition rate of 98.26 percent [11]. With the usage of HMM models for the pattern recognition or character recognition, a HMM model keeps information for a character when the model is trained properly and the trained model can be used to recognize an unknown character. The advantage with HMM based systems is that they are segmentation free that is no pre- segmentation of word/line images into small units such as sub-words or characters is required [12]. Features are extracted in both global and local processing. Generally for each character, a single HMM model is considered and trained by feature vectors. But we observe that some handwritten characters (e.g. A, W) show two completely different formats . We use multiple HMM models for these characters whereas for other characters, we have taken single HMM model.Global features of an image sample describe its overall structure. In this paper three global featurtes are used. In this paper gradieant features, projection features and curvature features are used. In this paper 5 state chain HMM model is used for character recognition.

Rakesh Kumar Mandal and N. R. Manna [13] try to recognize hand written English alphabets using single layer ANN. This approach makes the ANN simple, easy to implement and understand. Row-wise segmentation technique was developed and used here to achieve optimum accuracy. This paper is an approach to develop a method to get the optimized results using the easily available resources. Row-wise segmentation helps to extract out some common features among distinct handwriting styles of different people. In this paper a concept of recognizing hand written character pattern has been developed and implemented called Row-wise segmentation technique. RST helps in minimizing errors in pattern recognition due to different handwriting styles to great extent. In this method input pattern matrix is segmented row-wise into different groups. Target pattern is also grouped where each group is the numeric equivalent of the chronological position of each English alphabet. Each input segment is fully interconnected with each target group. Number of target groups is equal to the number of rows in the input matrix.

In general, the overall program has been divided into two parts, training and testing. Training requires the net to read segmented input patterns and testing requires the net to read any test character pattern, to read the produced target samples and to count the majority of samples and to find out the numeric equivalent of the sample to identify the character. The standard characters are written on a piece of A4 size paper with uniform square sized boxes. Each character is written in one box. Blue/Black ballpoint pen is used to write the characters on the paper. 26 uppercase letters of English alphabets are taken for the training purpose. The characters are captured using a scanner. The captured images are used to generate input matrices. Each input matrix corresponding to an alphabet is segmented row-wise, each row forming an input pattern, which were fed to the perceptron neural network for the training. RST is used for setting the targets. Weights have been calculated using RST. Few English alphabets written by different people are taken for testing. Scilab has been used to test the results and the accuracy is measured. From this we get 80% average accuracy.

Huiqin Lin, Wennuan Ou and Tonglin Zhu [14] create an alogorithm for handwritten character recognition in correcting assingnment system. Handwritten character recognition is the key technique in correcting assignment system as well as development of aided instruction software. Considering the disparity in distribution of the pixel, they propose a distribution based algorithm for handwritten character recognition. Based on the theory of Image Segmentation, the centroid of a character can be found. Around this centroid, the image is divided into equal angle regions clockwise and the direction feature of the character distribution can be obtained. The hand-writing restraint will be flexible since the method named Deflection Correction is adopted, as well as the matching error will be reduced. In the principle of high-accuracy matching, we use the minimal matching database to approach the real-time character match. The algorithm provides us a satisfactory recognition rate, especially on numbers and English characters, and the results of recognition can be classified effectively which are defined as Learning Functions. Without necessity of thinning and

finding the starting location and direction of the character, this method directly computes the regularity of distribution in each region of a character image [14]. It overcomes the complex process of extracting skeleton which traditional method has to adopt. With the advantage of speediness, accuracy and robustness, the system has an excellent vista. In this paper we use a matching algorithm. This matching algoriym show step by step processing of character recognition. By follow these step charcter recognition task is occur. Different step follow different task according to the algorithm require.

D.K. Patel,T. Som and Manoj Kumar Singh [15] presents a novel method of handwriting character recognition which exploits a compression capability of discrete wavelet transform to enhance the accuracy of recognition at the pixel level, the learning capability of artificial neural network and computational capability of Euclidean distance metric. The problem of handwritten character recognition has been tackled with multiresolution technique using discrete wavelet transform and learning rule through the artificial neural network. Recognition accuracy is improved by Euclidean distance metric along with recognition score in case of misclassification. Features of the handwritten character images are extracted by discrete wavelet transform used with appropriate level of multiresolution. Handwritten characters are classified into 26 pattern classes based on appropriate properties i.e. shape. During preprocessing each character is captured within a rectangular box and then resized to a threshold size. Weight matrix of each class is computed using the learning rule of artificial neural network, and then the unknown input pattern vector is fused with the weight matrices of all the classes to generate the recognition scores. Maximum score corresponds to the recognized input character. Learning rule provides a good recognition accuracy of 88.46%. In case of misclassification, the Euclidean distance metric improves the recognition accuracy to 92.31 % and then its product with recognition score further improves the recognition accuracy to 99.23 %.

Huihuang Zhao, Dejian Zhou and Zhaohua Wu [16], we present an approach to recognizing characters in surface mount technology (SMT) product. The recognition rate of SMT product character is not very good, because of unexpected noise in SMT the product character image. An improved SMT product character recognition method is proposed which can improve the recognition rate. At first, the character image is changed into a gray image. Some appropriate image processing algorithms are used to eliminate the noise. Then, single character image is obtained after character segmentation and character normalization. Finally, a three-layer back propagation (BP) neural network module is constructed. In order to improve the convergence rate of the network and avoid oscillation and divergence, the BP algorithm with momentum item is used. As a result, the SMT product character recognition system is developed, and its implementation method and steps are introduced with practical examples. Experimental results indicate that the proposed character recognition can obtain satisfactory character-recognition rate and the recognition rate reached over by 98.6%.

The back propagation neural algorithm is a supervised form of learning [17]. An input vector is presented at the input layer and allowed to propagate forward to generate an output. The mean square error (MSE) between the actual and desired outputs is calculated and propagated back through the layers, to enable the connection weights at each stage to be adjusted in proportion to the error term so as to minimize it. The BP algorithm for training a net is an iterative procedure. It is applied repeatedly for a set of input-output pairs, called the training set, until the output error reaches a desirably small value.

Peng Xu [18], a new algorithm PSO- BP was studied, giving full play to both of the particle swarm algorithm of global optimization ability and BP algorithm's local search advantage, and compared identification of 140 pixels of English letters together with BP algorithm. Experimental results show that particle swarm algorithm used for the optimization of the neural network has a faster convergence speed, and simpler algorithm. The BP algorithm of multilayer network is by far the most widely application in neural network, and it has strong nonlinear approximation ability, adaptive ability, and selflearning ability. Basic BP algorithm based on the error of the gradient descent algorithm has slow convergence, and easily fall into the local minimum for most practical application. To cope with the disadvantages of the BP algorithm, the particle swarm optimization based on the BP neural network is adopted in this paper.. And we designed a kind of English letters recognition training system. The improved BP algorithm greatly reduced the number of network training, effectively prevented the training of network from the local minimum point. And we obtained better effect. In this paper to cope with standard offline digital recognition, we studied the PSO-BP neural network classifier establishment and the identification process. optimized BP algorithm using PSO improved the global search ability of neural network, accelerated the network convergence speed, guaranteeed the network generalization ability. Because both of PSO-BP network and the single BP neural network algorithm in network is based on the back propagation network algorithm, it made no big difference in digital recognition rate. This optimization algorithm can be used to to identify handwritten letters with its high speed and precision, reliable prediction, and it has a good value in theory and practical application. In this paper we get 86.8% accuracy for capital letters and 85.3 for small letters.

P.M. Patil, P.S. Dhabe and U.V. Kulkarni [19] membership function of fuzzy hyperline segment neural network (FHLSNN) proposed by U.V.Kulkarni and T.R.Sontakke is modified to maintain convexity. The modified membership function is found superior than the function defined by them, which gives relatively lower values to the patterns which are falling close to the hyperline segment (HLS) but far from two end points of HLS. The performance of modified fuzzy hyperline segment neural network (MFHLSNN) is tested with the two splits of FISHER IRIS

data and is found superior than FHLSNN. The modified neural network is also found superior than the general fuzzy min-max neural network (GFMM). The parameter θ, which is the maximum length of HLS, controls the total number of hyperline segments created. For less value of θ both FHLSNN and MFHLSNN creates more hyperline segments than large value of θ. Generalization power of an algorithm depends on the convexity of the membership function.

VI. CONCLUSION

The character recognition methods have developed amazingly in the last decade. A variety of techniques have emerged, influenced by developments in related fields such as image recognition and face recognition. In this paper we provide review of various techniques used in offline handwritten character recognition. These techniques provide better accuracy by use of different classifier.This review provide information about different classifier used in character recognition techniques. This comprehensive discussion will provide insight into the concepts involved, and perhaps provoke further advances in the area. The promise for the future is significantly higher performance for almost every character recognition technology area.

REFERENCES

[1] Aiquan Yuan, Gang Bai, Po Young, Yanni Guo, Xintig Zhao,"*Handwritten english word recognition based on convolution neural networks*", 2012 International conference on trontiers in handwriting recognition.

[2] Pranab k Charles,V. Harsh, M. Swath, CH. Deepth,"*A review on the various techniques used for opticle character recognition*", International journal of engineering research and application, vol.2, pp.659-662, jan-Feb 2012.

[3] R. Plamondon and S.N. Srihari,"*Online and offline handwritten character recognition: A comprehensive survey",* IEEE transaction on pattern analysis and machine intelligence, vol22, no1, pp-63-84,2000.

[4] Nafiz Arica and Fatos T. Yarman," *An overview ocharacter recognition focused on offline handwriting*", IEEE transaction on system. Man and Cybeznetic, vol.31.no.2. May 2011.

[5] J.Pradeep, E.Srinivasan, S.Himawathi, "*Perfomance analysis of hybrid feature extraction technique for recognizing english handwritten character",*978-1-4673-4805-8/12/2012IEEE.

[6] Anshul Gupta, Manish Srivastva, Chitraleka Mahanta,"*Offline handwritten character recognition using neural network*", 2011 International conference on computer application and industrial electronics.

[7] Aiquan yuan, Gang Bai, Lijing Jiao, Yajie Liu,*"Offline handwritten english character recognition based on convolution neural network*, 2012, 10th IAPR International workshop on document analysis system."

[8] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet,"*Unipen project of on line data exchange and recognizer benchmarks"*. In Pattern Recognition 1994, vol.2-conference B: Computer vision & image processing, Proceeding of the 12th IAPR International. Conference on, volume 2, pages 29_33. IEEE, 1994.

[9] N.M. Noor, M.Razaz and P.Manely Cooke,*"Global geometry extraction for fuzzy logic based handwritten character recognition"* ,ICPR,2004.

[10] P.Manley Cooke and M. Razaz,"*Amultiple classifier for handwritten character recognition*," Proc of 16th Int Conf. on system engineering.Vol 2,PP 451-455, UK, 2003.

[11] Rajib Lochan Das, Binod kumar Prasad and goutam Sanyal,"*HMM based offline handwritten writer independent english character recognition using global and local feature extraction*,"International Journal of Computer Application,Volume 46-No 10,May 2012.

[12] P.Natrajan, S.saleem, R.Prasad,E.Macrostie and K. Subramaniam," *Multilingual offline handwriting recognition using hidden markov model: A script independent approach,* "Springer Book,ISSN 0302-9743,Vol 4768,March 2008.

[13] Rakesh Kumar Mandal, N R Manna," *Handwritten english character recognition using rowwise segmentation techniques*," ISDMISC 2011 By IJCA.

[14] Huiqin Lin, Wennuan Ou,Tanglin Zhu,"*The research of algorithm for handwritten charcter recognition in correcting assignment system,".*

[15] D. K. Patel, T. Som, Manoj Kumar Singh, " *Multiresolution technique to handwritten english charcter recognition using learning rule and Euclidean distance matric,".*

[16] Huihung Zhao, Dejian Zhou and Zhaohua Wu," *SMT product charcter recognition based on BP neural network,"* 2010 Sixth international conference on natural computation(ICNC 2010).

[17] Hasin H. C., Et al," *An adaptive training algorithm for back propagation neural network,*" IEEE transaction of system,Man and cybernetic.Vol-25,PP-512-514,March 1995.

[18] Peng Xu," *Research of standard and handwritten english letter recognition system based on the PSO-BP neural network*," IEEE,2011.

[19] P.M. Patil, P.S. Dhabe, U.V. Kulkarni and T. R. Sontakke,"*Recognition of handwritten charcter using modified fuzzy hyperline segment nueral network*,".

[20] Ashoka H. N., Manjaiah D. H., rabindranath Bera," *Feture Extraction Techniques for neural network based pattern recognition",* International Journal of Computer Science & Engineering 4(3), PP. 331-339,2012.

[21] Saurabh Shrivastva and Manu Pratap Singh*," Performance evaluation of feed forward neural network with soft computing techniques of hand written english alphabets*," Applied Soft Computing, Vol.11, PP.1156-1182, 2011.

[22] Serra," *Morphological filtering: An Overview*," Signal Process, Vol.38. No.1, PP.3-11, 1994.

[23] Alessandro Vinciarelli," *Noisy text categorization*", IEEE Transaction on pattern analysis and machine intelligence", Vol.27, No.12, PP.1882-1889,

December 2005.

[24] Salvador Espan-boquera, Maria Jose Castro-Bleda," *Improving offline handwritten text recognition with hybrid HMM/Annmodels*", IEEE transaction on machine learning and intelligence. Vol.33, No.4, PP.767-779, April 2011.

[25] A. K. Jain," *global directed evaluation of binarization methods*," IEEE transaction on pattern analysis and machine intelligence, Vol.17, PP1191-1201, December 1995.

[26] L. Lam, S. W. Lee and C.Y. Suan," *Thinning methodologies: A comprehensive survey*", IEEE Transaction on pattern analysis and machine intelligence, Vol.14, PP.869-885, Set 1992.