ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

A New Approach to Handwritten Character Recognition

Ch Suresh Kumar Raju, K Mayuri, A Swapna

Assistant Professor, Dept. of IT, IARE, Dundigal, Hyderabad, India

Assistant Professor, Dept. of CSE, IARE, Dundigal, Hyderabad, India

Assistant Professor, Dept. of IT, IARE, Dundigal, Hyderabad, India

ABSTRACT: Character recognition is a process by which computer recognizes handwritten characters and turns them into a format which a user can understand. Computer based pattern recognition is a process that involves several sub-processes. 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 creates paperless environment. The novelty of this approach is to achieve better accuracy, reduced computational time for recognition of handwritten characters. The proposed method extracts the geometric features of the character contour. These features are based on the basic line types that forms the character skeleton. The system gives a feature vector as its output. The feature vectors so generated from a training set, were then used to train a pattern recognition engine based on Neural Networks so that the system can be benchmarked. The algorithm proposed concentrates on the same. It extracts different line types that forms a particular character. It also concentrates on the positional features of the same. The feature extraction technique explained was tested using a Neural Network which was trained with the feature vectors obtained from the proposed method.

KEYWORDS: character recognition, character segmentation, feature extraction, preprocessing, neural network.

I. INTRODUCTION

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as image formats. Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. It contributes immensely to the advancement of an automation process and can improve the interface between man and machine in numerous applications. Several research works have been focusing on new techniques and methods that would reduce the processing time while providing higher recognition accuracy. The writing is usually captured optically by a scanner and the completed writing is available as an image. The neural networks have been successfully used to yield comparably high recognition accuracy levels .Several applications including mail sorting, bank processing, document reading and postal address recognition require off-line handwriting recognition systems. As a result, the off-line handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy.

The purpose of handwritten character recognition is to take handwritten English characters as an input, process the character, train the neural network algorithm, to recognize the character and modify the character to be a beautified version of the input. It is aimed at developing the software which will be helpful in recognizing the characters of English language. It is restricted to English characters only. It can be further developed to recognize the characters of different languages. One of the primary means by which computers are endowed while human like abilities is through the use of a neural network. Neural networks are particularly useful for solving problems that cannot be expressed as series of steps, such as recognizing characters, classifying them into groups, series prediction and data mining. Character recognition is perhaps most common use of neural networks. The neural network is presented with a target vector and also a vector which contains the pattern information. This could be an image and hand written data. The

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

neural network then attempts to determine if the input data matches a pattern that the neural network has memorized. The neural network trained for classification is designed to take input sample and classify them into groups. These groups may be fuzzy, without clarity defined boundaries. It is concerned detecting free handwritten characters.

Based on the research it found that Genetic algorithm has best discriminating features to recognize handwritten bangle digits but not in English characters. A genetic algorithm is also used for handwritten character recognition but it gives a poor detection of characters. This method cannot recognize every kind of English alphabets (different style of writing such as cursive, etc.,).It recognizes only the capital letters, we can train the letters. But it consumes large amount of time. The accuracy level is very less. Here is we give spacing between the letters it fails to recognize the exact character. The modern way of writing is compared with historical way of writing which completely gives an incorrect output. Handwritten character recognition using genetic algorithm involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting.

Hand written character recognition is comparatively difficult, as different people have different handwriting styles. And, as of today, OCR engines are primarily focused on machine printed text and ICR for hand "printed" (written in capital letters) text. There is no OCR engine that supports handwriting recognition as of today. To identify handwritten characters with the use of RST neural network, LAMSTAR neural network and Support Vector Machine. For the RST and LAMSTAR, a suitable neural network has to be constructed and be trained properly. For the SVM, multiple Binary Classifiers are required for Multi-Class Classification which we need for this problem. The program should be able to extract the characters one by one and map the target output for training purpose. After automatic processing of the image, the training dataset has to be used to train “classification engine” (LAMSTAR NN or SVM) for recognition purpose.

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 character 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.

II. PROPOSED METHODOLOGY

Our goal is to detect the character which is given as input, in whatever style the input text might be. In this project, we develop a model for handwritten character recognition. We also present the algorithms for recognizing the characters which is given as the input and give the correct output for the user. In this recognition process there won’t be any wrong recognition of characters. Instead it recognizes the input and gives the correct output for the user, with high accuracy and in a less time while compared to existing once.

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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

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 recognized character. The selection of the combination of feature-classifier contributes to the performance of the system. Several research works has been focusing toward evolving such methods to reduce the processing time and providing higher recognition.

Steps for back propagation algorithm

The BP algorithm consists of the following steps:

1. Define a training-sample for the network.
2. Compare the gotten output value with the required ones, and calculate the error for every output neuron.
3. Calculate the required output for every single neuron. Consider the incremental factor, which shows us how much every neuron weight has to be changed, so that they will be perfect in values. This shows the local error.
4. Modify the weight in from of every neuron in the way to minimize the local error.
5. Give a level of blame to every neuron, this ay giving higher responsibility for those neurons with greater weights before them.
6. Repeat the method from step 3 for the neurons of the previous layer, using the “blame” as factor.

Character recognition follows the existing modules to develop the stages of the character recognition.
They are

1. Drawing Text
2. Pre-processing
3. Segmentation
4. Feature extraction
5. Classification
6. Recognition

Let’s discuss each and every module briefly; the above modules can be represented by the block diagram.

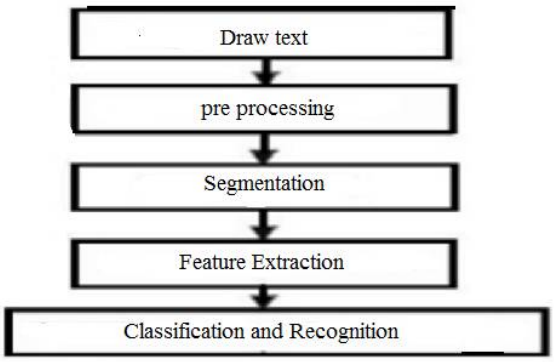

Fig 2.1: Block diagram of character recognition

1 Draw Text

In this module, the input is given through drawing a text, i.e., from the dataset particular character image is selected and we can check those text to particular character. The character is processed by the following modules.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

2 Preprocessing

The uploaded image may have some noise. This noise may be due to some unnecessary information available in the image. Various steps involved in the preprocessing are

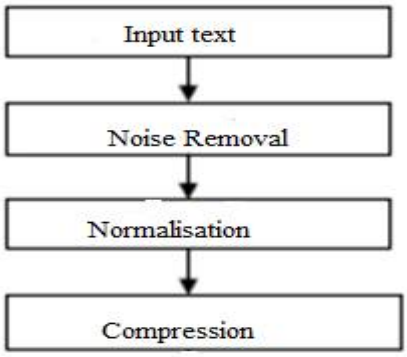

Fig 2.2: Preprocessing steps

The main objective of pre-processing is to de noise and enhance the quality of image.

2.1 Input Text

Here the input is given through the characters, each character can be given by the user and the character can be processed future steps as shown in the above figure.

2.2 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.

2.3 Normalization

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.

2.4 Compression

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

3 Segmentation

In the segmentation stage, an image consisting of a sequence of characters is decomposed into sub-images of individual characters. 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.

The segmentation stage takes in a page image and separates the different logical parts, like text from graphics, lines of a paragraph, and characters (or parts thereof) of a word. After the preprocessing stage, most HCR systems isolate the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

individual characters or strokes before recognizing them. Segmenting a page of text can be broken down into two levels: page decomposition and word segmentation, When working with pages that contain different object types like graphics, headings, mathematical formulas, and text blocks, page decomposition separates the different page elements, producing text blocks, lines, and sub-words. While page decomposition might identify sets of logical components of a page, word segmentation separates the characters of a sub-word.

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.

4 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.

5 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 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.

Main principles of back propagation neural networks

The main principles of BP neural networks can be concluded as following.

1. Input the character image into the BP neural networks.
2. Calculate the Error Function by comparing the recognizing result of the neural networks and the known character result so as to adjust the value of the connective parameters between layers and make the output of the network more approximate to the known result.
3. Train the networks with a series of known images with the purpose of optimizing the parameters of the networks.

Experimental results show that a simple BP neural network is not efficient enough in character recognition with a recognition rate of less than 90%. The direct reason is the loss of spatial compounding information. In order to improve the performance of the arithmetic parallel BP neural network paper is selected. The parallel BP neural network consists of two simple BP neural networks: BP neural networks A and BP neural networks B. In the input layer of neural networks A, the character image is input in a row-first way so that the row spatial information of the character is preserved. And in the input layer of neural networks B, the character image is input in a column-first way so that the column spatial information is preserved. Therefore, the proposed parallel BP neural networks can preserve the spatial compounding information and can effectively improve the recognition rate.

It describes a Back Propagation technique for feature extraction applicable to segmentation based word recognition systems. The proposed systems extract geometric features of the character contour. These features are based on basic line types that forms the character skeleton the system gives a feature vector as its output. The feature vectors so

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

generated from a training set were then used to train a character recognition engine based on neural networks so that the system can be bench marked. It provides an easy user interface to input the object image. User will be able to upload the image without any difficulty. System will preprocess the given input to suppress the background. It will detect the text regions present in the image. It will retrieve text present in the image and displays them to the user.

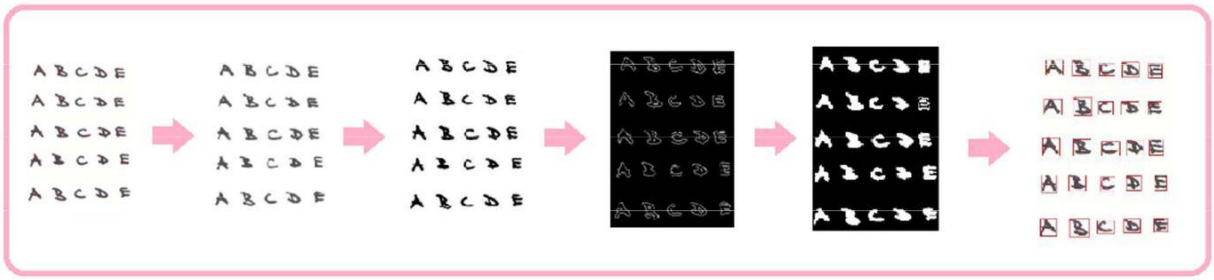

Fig 2.3: Automatic image processing steps

The description of above figure 2.3 as follows.
1- Read the character on Microsoft visual studio workspace.

2- Convert to grayscale.

3- Convert to binary image.

4- Edge detection.

5- Morphology.

At the step image dilation and Image filling is performed.

6- Blob analysis.

At this step, all the objects on the image and all the properties of each object are found.

7- Plot the object location

At this step, the location of each object is plotted.

After these steps, the characters can be extracted from the text and thus we would be able to extract the most important features of each character.

Later feature extraction is done here Neural networks can be used, if we have a suitable dataset for training and learning purposes. Datasets are one of the most important things when constructing a neural network. Without proper dataset, training will be useless. In order to get a proper data set: First we have to scan the text. After the text is scanned, we define processing algorithm, which will extract important attributes from the text and map them into a database or better to say dataset.

Extracted attributes will have numerical values and will be usually stored in arrays. With these values, neural networks and SVM can be trained and we can get a good end results. The problem of well-defined datasets lies also in carefully

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

chosen algorithm attributes. 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.

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. Attributes are important and have a crucial impact on end results. In the following, the steps required to extract the most important features of each character are described:

The sub-images have to be cropped sharp to the border of the character in order to standardize the sub-images. The image standardization is done by finding the maximum row and column with 1s and with the peak point, increase and decrease the counter until meeting the white space, or the line with all 0s. This technique is shown in figure below where a character “C” is being cropped and resized.

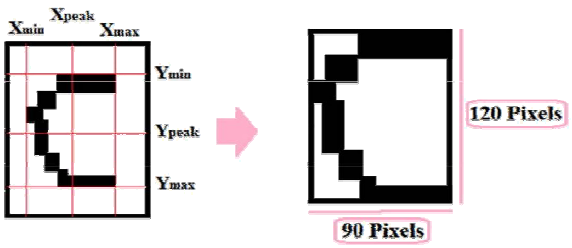

Fig 2.4 Cropped and Resized Picture

The image pre-processing is then followed by the image resize again to meet the network input requirement, 5×7 (***or 6×8***) matrices, where the value of 1 will be assign to all pixel where all 10×10 (***15×15)*** box are filled with 1s, as shown below

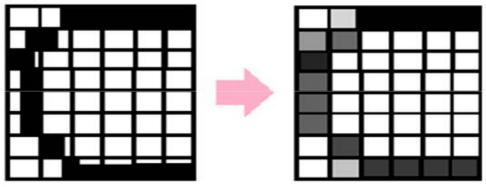

Fig 2.5: Image resize again to meet the network input requirement

Finally, the 5×7 (***6×8***) matrices is concatenated into a stream so that it can be feed into network with 35 (***48***) input neurons. The input of the network is actually the negative image of the figure, where the input range is 0 to 1, with 0 equal to black and 1 indicate white, while the value in between show the intensity of the relevant pixel. By this,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

we are able to extract the character and pass to another stage for future "classification" or "training" purpose of the neural network or SVM.

Network set-up

The Back propagation (BP) learning algorithm was used to solve the problem. The goal of this algorithm is to minimize the error energy at the output layer. In this method a training set of input vectors is applied vector-by-vector to the input of the network and is forward-propagated to the output. Weights are then adjusted by the BP algorithm. Subsequently, these steps are repeated for all training sets. The algorithm stops when adequate convergence is reached.

III. EXPERIMENTAL RESULTS

Training algorithm

To train the network to recognize the English Alphabet characters, the corresponding 5×7 (6×8) grids are applied in the form of 1×35 (1×48) vectors to the input of the network. Then the weights are calculated using the equations provided in the text book for the BP NN. The initial learning rate was experimentally set to 1.5 which is divided by a factor of 2 every 100 iterations and is reset to its initial value after every 400 iterations and the momentum rate is set to 0.95**.**

Testing algorithm

For testing, the weights that were calculated during the training are used. The testing inputs are given in the form of a 1×35 (1×48) vectors for the corresponding 5×7 (6×8) grids. Characters are considered recognized if all the outputs of the network were no more than 0.01 off their respective desired values.

The following graph shows the accuracy rate compared to existing systems (genetic algorithm) to proposed

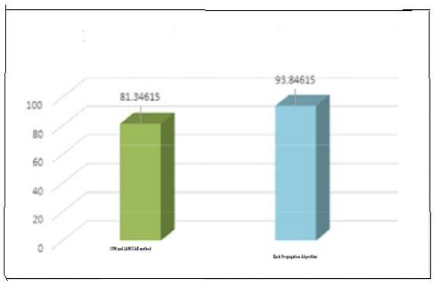

Fig 3.1: Comparison graph

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.

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. Fig 3.1 shows the accuracy rate comparison.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The following graph shows the accuracy rate between Genetic algorithm An SVM and LAMSTAR method And Back propagation method.

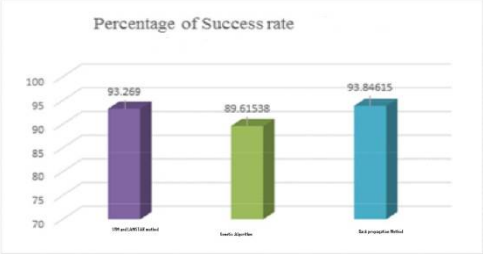

Fig 3.2: Percentage of success rate

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%. Fig 3.2 shows percentage of success rate.

Processing applications as searching for content, editing, and 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.

IV.CONCLUSION

The method of training neural network with extracted features from sample input of each character has detection accuracy to a great extent. The proposed methodology has produced good results. The system is evaluated for a set of sample input in efficient manner

The major approaches used in the field of handwritten character recognition during the last decade have been reviewed. Different pre-processing, segmentation, feature extraction, classification techniques are also discussed. Though, various methods for treating the problem of hand written English letters have developed in last two decades, still a lot of research is needed so that a viable software solution can be made available. The existing OCR for handwritten has very low accuracy. We need an efficient solution to solve this problem so that overall performance can be increased.

Results obtained in the proposed work indicate that structural classification facilitates result improvement. Implementation of the handwritten character recognition system with neural networks gives promising results over template matching alone in both respects; recognition rate and execution time. More over implementation of hybrid network in case of both character and compound character gives improved accuracy over single network.

FUTURE WORK

It helps us to determine the efficient and more accurate results for recognition of English handwritten characters and digits. Offline handwritten English alphabet characters recognition system using new type of feature extraction.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Many techniques for recognition of Offline English Handwritten Characters have been suggested. But still an efficient OCR for the recognition of hand written letters does not exist. Few steps have been taken for Hand written and Hand printed (which is a constrained hand writing) English letter recognition.

Various challenges are identified which may provide more lively interest to the researchers. These challenges are: difficulty to identify the diverse human writing styles, different angles of letters, different shapes and size of letter, pure input quality, low accuracy rate in recognition etc. Hence, a lot of research work is to be done to solve these problems.

REFERENCES

[1] Gauri Katiyar and Shabana Mehfuz, 'A hybrid recognition system for off-line handwritten characters', *Springer Plus*, Vol:5, Issue No:357, 2016
[2] Rakesh Kumar Mandal, N R Manna, 'Handwritten english character recognition using row wise segmentation techniques', ISDMISC 2011 By IJCA
[3] 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.
[4] Anshul Gupta, Manish Srivastva, Chitraleka Mahanta, 'Offline handwritten character recognition using neural network', 2011 International conference on computer application and industrial electronics.
[5] Peng Xu, 'Research of standard and handwritten english letter recognition system based on the PSOBP neural network', IEEE, 2011.
[6] P.M. Patil, P.S. Dhabe, U.V. Kulkarni and T. R. Sontakke, 'Recognition of handwritten character using modified fuzzy hyperline segment neural network'.
[7] D. K. Patel, T. Som, Manoj Kumar Singh, 'Multi resolution technique to handwritten english character recognition using learning rule and Euclidean distance metric'.
[8] 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.
[9] 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."
[10] J.Pradeep, E.Srinivasan, S.Himawathi, 'Perfomance analysis of hybrid feature extraction technique for recognizing english handwritten character', 978-14673-4805-8/12/2012IEEE.

BIOGRAPHY

1. **Ch Suresh Kumar Raju** is currently working as an Assistant professor, Department of IT, Institute of Aeronautical Engineering College, Hyderabad. His current research interests include Data mining, Algorithms, Cloud Computing.

2. **K. Mayuri** is currently working as an Assistant professor, Department of CSE, Institute of Aeronautical Engineering College, Hyderabad. Her current research interests include Data Mining, Network Security, Mobility Management and Cloud Computing.

3. **A Swapna** is currently working as an Assistant professor, Department of IT, Institute of Aeronautical Engineering College, Hyderabad. Her current research interests include Data mining, Algorithms, Cloud Computing.