ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

English Cursive Handwritten Character Recognition using OCR (Optical Character Recognition)

Dr.M.P.Chitra [1], Shobana S [2], Shiny A[3], Madhumitha G[4]

Professor, Dept. of ECE, Panimalar Institute of Technology, Chennai, Tamil Nadu, India[1]

UG Student, Dept. of ECE, Panimalar Institute of Technology, Chennai, Tamil Nadu, India[2]

UG Student, Dept. of ECE, Panimalar Institute of Technology, Chennai, Tamil Nadu, India[3]

UG Student, Dept. of ECE, Panimalar Institute of Technology, Chennai, Tamil Nadu, India[4]

ABSTRACT: This paper displays an exhaustive survey of disconnected character acknowledgement framework. In this proposed strategy the information picture is in checked organisation and it will be changed over into the proportional character in printed configuration by utilizing optical character recognition acknowledgement (OCR). The checked picture is sustained in to the computer and it is perceived utilizing the Hidden Markov Model and is changed over into a similar word in printed characters. The KNN(kth nearest neighbour)order calculation is created and has been utilized for this characterization process.

KEYWORDS: OCR-Optical Character Recognition, HMM- Hidden Markov Model, KNN- Kth nearest neighbour

I. INTRODUCTION

The English cursive characters ought to be perceived by utilizing the optical character Recognition (OCR). Optical character acknowledgement, normally shortened to OCR, is the mechanical or electronic change of filtered pictures of written by hand, typewritten or printed content into machine-encoded content. It is broadly utilized as a type of information passage from some kind of unique paper information source, regardless of whether documents, deals receipts, mail or any number of printed records. It is a typical strategy for digitizing printed messages with the goal that they can be electronically looked, put away more minimally, showed on-line, and utilized in machine procedures, for example, machine interpretation in picture to content. OCR is a field in Pattern recognition, text recognition, artificial intelligence and computer vision. Also, the Hidden Mark the relationship between the adjoining character and word. It is additionally utilized for the shape acknowledgement Optical character acknowledgement, normally shortened to OCR, is the mechanical or.

Also, it is utilized to decide the hidden characters from the discernible character. Handwritten Character Recognition (HCR) is the procedure of change of manually written content into machine intelligible structure. The serious issue in Handwritten Character Recognition(HCR) framework is the variety of the handwriting styles, which can be totally extraordinary for various journalists. The goal of hand written character recognition framework is to implement user friendly computer assisted character recognition ,to digitalize and interpret the manually written content into machine intelligible content. . The method is based on horizontal and vertical centerline of the character. The characters are segmented into horizontal and vertical lines and position of the character is obtained according to those lines.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

The features are calculated using sliding window techniques. The characters are classified using multilayer perceptron (MLP). The correction rate obtained was 99.28 % for the 19437 Amazigh printed characters and 96.32% for the 20150 Amazigh handwritten characters. The discrete wavelet transform(DWT) based handwritten character recognition system along with Euclidean distance metrics. The classification of the testing vector is depending upon the Euclidean distance. Minimum distance decides the class of the test vector. This system gives good accuracy of 90%. A grid approach for recognition of an offline handwritten character using grid approach is proposed. Extracted features are train by neural network as classifier of the character in classification stage. The recognition system of experimental results shows that this technique is effective and reliable. The overall procedure results in recognition rate are 96.9%.[1],[2],[3].

A system with feed forward neural network having ability of a machine to interpret handwritten characters from sources like paper document, photograph etc. to digital computerized form is the aim of HCR systems. The neurons of output layer have a feedback connection from their output line. Experimental result shows that an efficient recognition. However, the proposed system is not a complete one. Some other techniques may be combined with this approach to increase the efficiency of the system. The work can be extended to recognize characters or numerals of some other languages also[4]. A neural network along with surf feature approach to solve complex character recognition issue. This approach has been evaluated using noise parameter. The evolution is performed by PSNR. The classification is done by using back propagation neural network. The success rate of this system is 98.77% [5].

II. TYPES OF RECOGNITION

Handwritten Character Recognition is isolated into two classifications,

- **Online character recognition:**

It is system in which recognition is performed when characters are under creation.

- **Offline character recognition:**

It is framework in which previously written by hand records are generated, scanned ,produced in computer and they are perceived.

There are four strategies for cursive manually written word acknowledgement.

Holistic Approach:

It is a strategy in which whole word is perceived without part them by extracting features of whole word.

Segmentation based Approach:

Characters are sectioned from word.

Reorganization based Segmentation approach:

Character characterization and division are performed at the same time by utilizing proper learning technique.

Mixed Approach:

This strategy comprises of blend of above strategies.

In this paper we succinct overview of accessible HCR for English language. HCR systems are examined with their quality and short comings. Diverse sorts of highlights are extricated and distinctive kinds of classifiers are utilized to characterize the information characters. The present examination is centered around examination of conceivable methods to build up a disconnected HCR framework for English language for both separate characters and cursive words.

III. DESIGN AND IMPLEMENTATION STEPS

An outline of the systems embraced in this examination work is displayed beneath. A framework is created to perceive cursive manually written English characters utilizing MATLAB, for a subset of the English Characters. The real advances engaged with penmanship acknowledgement framework and the means included are,

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

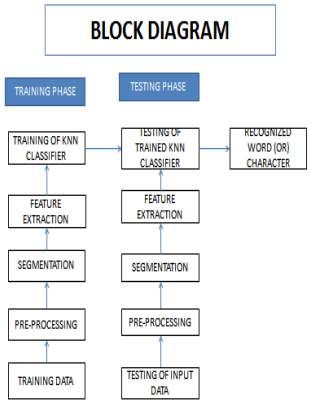
International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

1. Pre-Processing
2. Segmentation
3. Feature Extraction
4. Recognition
5. Classification



1. Pre-Processing:

Pre-Processing is concerned for the most part with the decrease of commotions and changeability in the information. The commotion from the picture is diminished by utilizing median filter. The number and sort of preprocessing calculations utilizes on the checked pictured rely upon numerous variables, for example, paper quality, goals of the filtered picture, the measure of skew in the picture and the design of the content. Different pre-processing tasks are performed before acknowledgement to improve the nature of the information picture.

- RGB TO GRAY CONVERSION:

In the pre-processing first stage is to change over the information into RGB picture into dark scale picture.

- BINARIZATION:

Binarization is the way toward changing over a gray scale picture (0 to 255 pixel esteems) into binary picture (oto1 pixel qualities) by choosing a limit an incentive in the middle of 0 to 255. The nearby thresholding technique is connected on the grayscale picture as

$$g(x,y) = \{1\ f(x,y) > Th,$$
$$0\ otherwise$$

Where g(x,y) is paired picture pixel, f(x,y) is dark picture pixel, This the neighbourhood limit esteem. The parallel picture is upgraded by utilizing morphological channels like erosion and dilation. Erosion is a morphological task which evacuates developing part while dilation fills the holes inside the article.

- SKEW IDENTIFICATION AND ADJUSTMENT:

While examining the picture, if the paper source record isn't adjusted appropriately, it might cause the parts to be titled. This could prompt wrong conduct of the OCR framework. To keep this, Radon Transform (skew identification and Rectification strategy) has been contrived which distinguish and expel the skew from the picture and later the limits of specific pictures are balanced.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

2. Segmentation:

It is an activity that looks to decay a picture of arrangement of characters into sub pictures of person images. Character segmentation is a key necessity that describes the utility of customary character Acknowledgement frameworks. It incorporates line, word, character segmentation. Distinctive techniques utilized can be arranged dependent on the kind of contented technique being pursued like acknowledgement based segmentation.

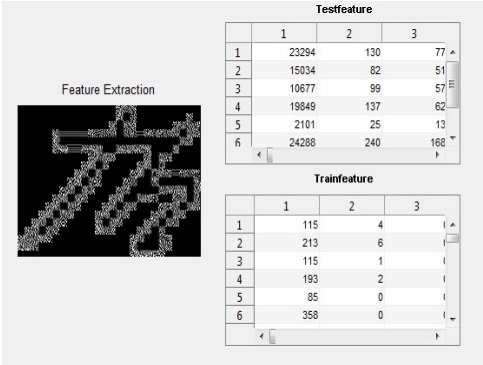
- LINE SEGMENTATION**:**

In a printed content, the content lines are nearly of same stature, gave that the content is written in a particular text dimension. On the off chance that the content is created by a sort machine, unquestionably the text dimension will be uniform all over the place. Between two content lines, there is a limited even band with either no pixel or not many pixels. Consequently, by checking break-focus through them and putting away them will be helpful for distinguishing the valleys in it, content line groups can be recovered.

- CHARACTER SEGMENTATION**:**

After the line division, consider every single line which is portioned before experiencing the procedure of character division, Each line is sectioned in its individual characters(secluded) for further activity.

Example of a feature extraction



ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

3. FEATURE EXTRACTION:

Feature extraction is the issue of removing the letters from the pre_prepared information, the data, which is most important for grouping purposes, in the feeling of limiting the inside class design inconstancy, while upgrading the between-class design changeability.

4.RECOGNITION:

The last objective of character recognition is to get the class codes of character designs. On fragmenting character examples or word to a class out of a pre-processing characterized class set. Example characterization has been the primary subject of example acknowledgment field and is frequently taken as an equivalent word for" Pattern Recognition(PR)".Generally,design acknowledgeme-nt strategies are named template and feature based approach.

5. CLASSIFICATION:

The arrangement procedure ought to be finished by utilizing the KNN order calculation. The K-closest neighbor's calculation is a technique for arranging objects dependent on nearest preparing precedents in the component space. The KNN order separates the information in to a test and preparing set. It ought to depend up on the separation. The accompanying picture ought to be demonstrates the KNN separate figuring calculation.

KNN Classification Algorithm:

KNN is a least difficult directed learning calculation used to characterize the information. The K-closest

KNN Classification Algorithm

Algorithm: Simple k nearest neighbors pseudo code

Procedure: Find the class labels

Input: k-number of nearest neighbors;D-the set of test sample:T-the set of training sample

Output;L-the label set of test sample

1. read the data file(Training Data)
2. read the data file(Testing Data)
3. L = {}
4. for each d in d and t in T do
5. Neighbors(d) = {}
6. If |Neighbors(d)| < k then
7. Neighbors (d) = closest(d,t) U Neighbors(d)
8. End if
9. If |Neighbors(d)| = k then
10. Break
11. L = test class of (Neighbors(d)) U L
12. End for

Notes:

Neighbors (d) return the k nearest of d

Closest (d, t) return to the closest elements of t in d

Test class(s) returns the class label of S

neighborcalculation (KNN) is a method for classifying object by majority of votes. The value of k decides the search space. If k=1 means the object is simply assign the class which is nearest. The advantage of the Knn is its classification accuracy is better when feature spaces are small but which feature space is large then its accuracy get decrease.k-Nearest Neighbor can be calculated by using Euclidian distance. It gives efficiency and productivity. It is a distance between two points in Euclidian space. It computes the root of square differences between co-ordinates of pair of data points. Formula for Euclidian distance is givenbelow.

$$\text{Distxy} = \Sigma(xik - xjk)\text{^}2$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

IV. RESULT AND ANALYSIS

To delineate the precision of English manually written and diverse example content pictures of various text styles of various sizes have been tried under KNN calculation by utilizing MATLAB 2010a and afterward the exactness was estimated utilizing the examples.

$$\% \text{ of Accuracy} = \frac{\text{Number of characters Found correctly}}{Totalnumberofpatterns}$$

The proposed calculation utilizes a dark scale histogram, thresholding range and the Pinnacle Motion to-Commotion Proportion (PSNR). In view of our experience, the mean estimation of the dim scale histogram is firmly associated with pixel force. We utilize the PSNR quality measure proposed by [5] on the grounds that it can quantify the similitudes between the first picture and the binarized picture. A higher PSNR demonstrates greater likeness between the two pictures. The PSNR esteem is characterized as pursues:

$$PSNR = 10 * \log_{10}\left[\frac{Max^2}{\frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2}{M*N}}\right]$$

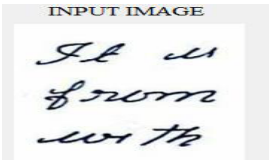
In the above condition, Max is a greatest incentive in the picture (for instance, in a dark scale picture, Max is 255) while m and n are the stature and weight of the picture, separately. I(i,j) is the first estimation of the picture, and K(i,j) is the incentive after a change. We ascertain PSNRs for each limit esteem, in additions of 5, in the extend between [1,256]. We ascertain the contrast between PSNR(t) and PSNR(original picture). In this calculation we ascertain PSNR by utilizing unique picture and every edge value[5].

TABLE 1:COMPARISON OF ACCURACY AND PSNR

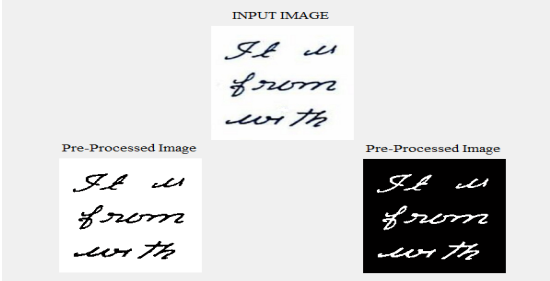
Parameters	Existing	Proposed
Accuracy	85.1%	90%
PSNR	10.41	20.5

By contrasting the current technique SVM classifier and our proposed KNN classifier, the KNN classifier gives more precision and PSNR. The accompanying pictures are demonstrating the info and yield of the cursive written by hand acknowledgment

Example of an input image



Example of a pre-processed image



ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

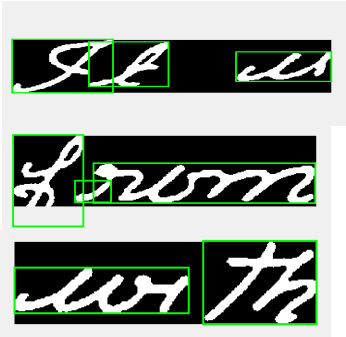
International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

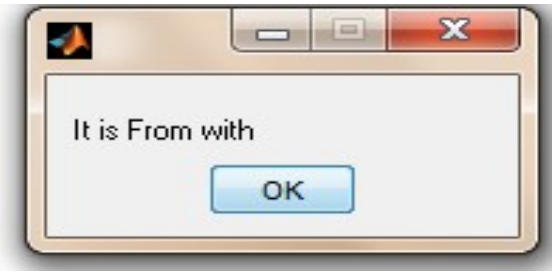
Vol. 8, Issue 3, March 2019

Example of bounding box



It demonstrates the transcribed info picture. The Fig 7 demonstrates the pre-handled picture. The Fig 8 indicates bouncing box of the portioned and splitted characters. The Fig 9 demonstrates the printed type of info cursive manually written picture**.**

Example output of cursive handwritten recognition



IV. CONCLUSION

The English cursive transcribed characters are perceived utilizing OCR. The mistakes which are caused because of clamor in the examined picture is expelled by applying middle channel in it. The prepared examples are utilized to lessen the complexities in the acknowledgment procedure. This methodology demonstrates a superior outcome with fast and exactness. The precision of our proposed work is above 90% by improving the KNN calculation.

REFERENCES

1. Youssef EsSaady, Ali Rachidi, Mostafa El Yassa, DrissMammass, 2011. “Amazigh Handwritten Character Recognition based on Horizontal and Vertical Centerline of Character”, International Journal of Advanced Science and Technology, 33:33-50.
2. Dileep Kumar Patel, TanmoySom, Sushil Kumar Yadav, Manoj Kumar Singh, 2012. “Handwritten Character Recognition Using Multiresolution Technique and Euclidean Distance Metric”, Journal of Signal and Information Processing, 3:208-214.
3. Prasad, P. Chaudhari, K.R. Sarode, 2014. “Offline Handwritten Character Recognition by using Grid approach”, International Journal of Application or Innovation in Engineering & Management (IJAIEM), 3(4):2319-4847
4. Apash Roy and N.R. Manna, 2014. “Handwritten Character Recognition with Feedback Neural Network”, International Journal of Computer Science & Engineering Technology (IJCSET), ISSN: 2229-3345 5:01.
5. ReetikaVerma, RupinderKaur, 2014. “An Efficient Technique for character recognition Using Neural Network & Surf Feature Extraction”, (IJCSIT) International Journal of Computer Science and Information Technologies, 5(2): 1995-1997, ISSN0975-9646.
6. Hassan El Bahi, ZouhirMahani and AbdelkarimZatni, “An offline handwritten character recognition system for image obtained by camera phone”, Mathematical Applications in Modern Science, ISBN: 978-1- 61804-258-3, pp:180-189.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

7. Pradeep, J., E. Srinivasan and S. Himavathi, 2012. ”Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten”, International Journal of Engineering (IJE) Transactions B: Applications, 25(2):99-106.
8. Vicente Bosch, Alejandro Hector Toselli, and Enrique Vidal, 2012. Statistical text line analysis in handwritten documents. In Frontiers in Handwriting Recognition (ICFHR), pp:201-206.
9. Sylvie Brunessaux, Patrick Giroux, Bruno Grilh`eres, Mathieu Manta,MaylisBodin, Khalid Choukri, Olivier Galibertand Juliette Kahn, 2014. The Maurdor Project: Improving Automatic Processing of Digital Documents. In Document Analysis Systems (DAS),pp:349-354.
10. Horst Bunke, SamyBengio and Alessandro Vinciarelli, 2004. Offline recognition of unconstrained handwritten texts using hmms and statistical language models. Pattern Analysis and Machine Intelligence, IEEE Transactions on,26(6):709-720.
11. N. Arica, F.T. Yarman-Vural, "One Dimensional Representation Of Two Dimensional Information For HMM Based Handwritten Recognition", Pattern Recognition Letters, vol. 21, pp. 583-592, 2000.
12. N. Arica, F.T. Yarman-Vural, "A New Scheme for Off-Line Handwritten Connected Digit Recognition",Proc. Int'l Conf. Pattern Recognition, pp. 1127-1131, 1998.