

# Chain Code Based Handwritten Cursive Character Recognition System with Better Segmentation Using Neural Network

Parikh Nirav Tushar<sup>1</sup>, Dr. Saurabh Upadhyay<sup>2</sup>

<sup>1</sup> M.E Student, Dept. of Computer Science & Engineering, S.P.B Patel College of Engineering, Mehsana, India

<sup>2</sup> Professor, Dept. of Computer Science & Engineering, S.P.B Patel College of Engineering, Mehsana, India

## ABSTRACT:

Character recognition plays an important role in many recent applications. Pattern recognition deals with categorization of input data. It is easy to recognize normal character but in cursive character we have to find out the boundary of a character that's why we have to apply better slant and segmentation techniques. A proper feature extraction method can increase the recognition ratio. In this paper, a chain code based feature extraction method is investigated for developing HCCR system. Chain code is working based on 4-neighborhood or 8-neighborhood methods. In this paper, 8-neighborhood method has been implemented which allows generation of eight different codes for each character. After feature extraction method, Classification techniques have been used for training and testing of Neural Network and other classifier.

**Keywords:** Pattern recognition, handwritten cursive character recognition (HCCR), segmentation with slant correction, feature extraction, chain code, neural network.

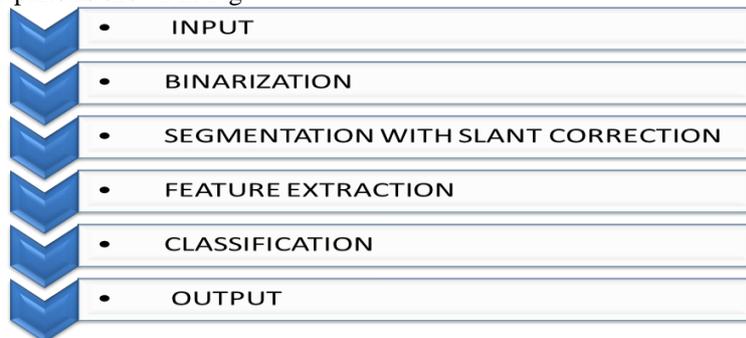
## I. INTRODUCTION

Pattern recognition is a field of study whose general goal is the classification of objects into a number of categories. Handwriting recognition has always been a challenging task in pattern recognition. Handwritten digit recognition is a system widely used in the United States. This system is developed for zip code or postal code recognition that can be employed in mail sorting. It is very useful for document digitalization.

For cursive character recognition first we have to do binarization. After that slant correction and segmentation will make character in proper manner. Which is useful for find out the starting and ending boundary of particular character. Next process of this mechanism is to design a dataset for feature extraction and another data set is for to train the classifier using Neural Network. In next process different feature extraction methods are applied on input data set and extract feature from it. This extracted feature applied on different classifier which was already trained through input data set which matches between feature and trained data.

## II. HCCR SYSTEM

This system is useful to find out and recognize handwritten cursive character. The steps of HCCR can be divided in four major parts as shown in Fig.



(Figure 1: HCCR System)

### 2.1 Binarization

Binarization (thresholding) refers to the conversion of a gray-scale image into a binary image. Image is converted into 1s and 0s form after binarization next step is slant correction and segmentation.

## 2.2 Slant correction and Segmentation

**Slant Correction:** Slant correction in cursive writing is a very difficult task. Alignment of character is differing from person to person. The slant is the strokes from the vertical direction, depending on the writing style. To create segmentation first we have to correct the angle of character.



Slant angle= 34 degree

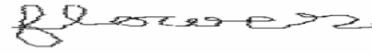


image after slant angle correction

**Segmentation:** In segmentation, the input image is segmented into individual character and then, each character is resized into  $m \times n$  pixels towards the extracting the features.

### Algorithm:

1. Calculate the size of image.
2. Frequency of black pixels in each row is counted in order to construct the row.
3. After row is segmented, individual character is segmented by finding boundary using connected components.
4. Repeat step 3 until all the character of row is segmented.
5. If all the row of image is segmented then stop else go to step 2.

**Line Segmentation:** Text line detection has been performed by scanning the input image horizontally after that Frequency of black pixels in each row is counted in order to construct the line segment.

**Word and Character Segmentation:** After a line has been detected, each line is scanned vertically for word segmentation. It divides line of characters into words.

Character Segmentation:



Connected strings of characters



Isolated Characters

## 2.3 Feature Extraction

Any given image can be decomposed into several features. Feature extraction technique is accurately retrieve features of characters. It is useful to find out the amount of black pixels from given data.

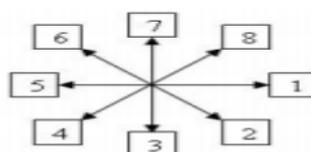
Useful properties might be:

- Aspect ratio
- Percentage of pixels over horizontal plan
- Percentage of pixels over vertical plan
- Number of strokes

- **Chain Code Generation**

Chain codes are used to represent the shape & boundary of connected characters. This representation is based on 4-connectivity or 8-connectivity of the segments. A chain code can be generated using boundary of an object in a clockwise direction and assigning a direction to the segments.

First, we pick a starting pixel location anywhere on the object boundary. Our aim is to find the next pixel in the boundary. There must be an adjoining boundary pixel at one of the eight locations surrounding the current boundary pixel. By looking at each of the eight adjoining pixels, we will find at least one that is also a boundary pixel.



(Figure 2: 8-neighborhood Chain Code)

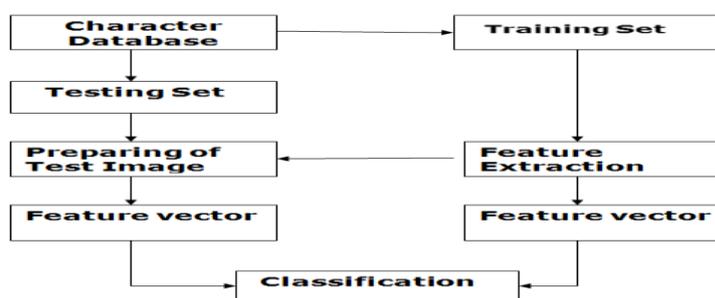
Algorithm for generating chain code considering 8- neighborhood is as follows:

- Step 1: Find out starting point which has nonzero values and store it in first
- Step 2: Initialize 0-7 total eight directions
- Step 3: Travels all 8 neighbors
- Step 4: Find first nonzero value
- Step 5: Add it in to chain code list
- Step 6: Move to next position
- Step 7: Check whether we reach to first point or not if not then go to step 3.

## 2.4 Classification Technique.

### • Neural Network

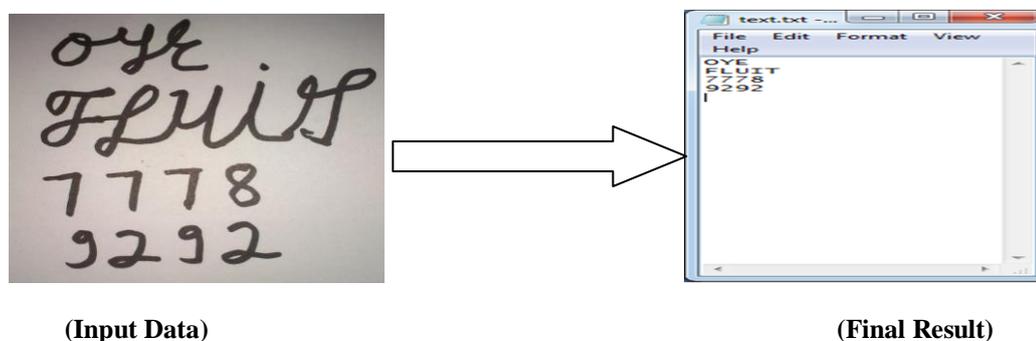
Artificial neural networks (ANN) provide the powerful simulation of the information processing and widely used in patten recognition application. The most commonly used neural network is a multilayer feed forward network which focus an input layer of nodes onto output layer through a number of hidden layers. In such networks, a back propagation algorithm is usually used as training algorithm for adjusting weights. The back propagation model or multi-layer perceptron is a neural network that utilizes a supervised learning technique. Typically there are one or more layers of hidden nodes between the input and output nodes. Besides, a single network can be trained to reproduce all the visual parameters as well as many networks can be trained so that each network estimates a single visual parameter. Many parameters, such as training data, transfer function, topology, learning algorithm, weights and others can be controlled in the neural network.



(Diagrammatic representation of how Neural Network works?)

Using this diagram we can easily identify how training and testing sets will work. First of all we have to train a data for Neural Network then use a testing data on it. After applying a feature extraction method to the data classification will apply and it compares testing data set and training data set After that it will generate a recognition rate.

## 3 Experimental Results and Output:



## III. CONCLUSION AND FUTURE WORK:

An accurate and an efficient off-line handwritten cursive character recognition system using a feature extraction method, namely, chain code is investigated. Using better feature extraction method like chain code we can achieve more accurate and efficient outcomes. To recognize a character we have to segment it properly therefore we have to apply better slant correction and segmentation logics. Characters must be written in proper manner.

This method gives recognition of 80% or more. Therefore applying more effective feature extraction will give a good recognition score. My next goal is to use other classification techniques like support vector machine (SVM) and hidden markov model (HMM) too.

## REFERENCES

- [1] Anita Pal<sup>1</sup> & Dayashankar Singh<sup>2</sup>, "Handwritten English Character Recognition Using Neural Network", International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 141-144
- [2] Tanmoy Som & Sumit Saha, "Handwritten character recognition by using Neural-network and Euclidean distance metric", Mathematics dept., Assam University, Silchar, INDIA.
- [3] Md.Mahbub Alam & Dr.M.Abul Kashem, "A Complete Bangla OCR System For Printed Characters", ISSN 2078-5828 (PRINT), ISSN 2218-5224 (ONLINE), VOLUME 01, ISSUE 01,2010 JCIT.
- [4] Tanmoy Som & Sumit Saha, "A new approach for slant angle correction and character segmentation of handwritten document", Department of mathematics Assam University, Silchar, INDIA.
- [5] H. Izakian, S. A. Monadjemi, B. Tork Ladani, and K. Zamanifar, "Multi-Font Farsi/Arabic Isolated Character Recognition Using Chain Codes" World Academy of Science, Engineering and Technology 43 2008.
- [6] H.Fujisawa, Y.Nakano and K.Kurino, "Segmentation methods for character recognition from segmentation to document structure analysis". Proceeding of the IEEE, vol.80, and pp.1079-1092. 1992.
- [7] H.Fujisawa, Y.Nakano and K.Kurino, "Segmentation methods for character recognition from segmentation to document structure analysis". Proceeding of the IEEE, vol.80, and pp.1079-1092. 1992.
- [8] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery", 1998.
- [9] H.Fujisawa, Y.Nakano and K.Kurino, "Segmentation methods for character recognition from segmentation to document structure analysis". Proceeding of the IEEE, vol.80, and pp.1079-1092. 1992.
- [10] Yi-Kai Chen and Jhing-Fa Wang, "Segmentation of Single-or Multiple-Touching Handwritten Numeral String Using Background and Foreground Analysis", IEEE PAMI vol.22, 1304-1317, 2000.
- [11] Pal, U. and B.B. Chaudhuri, "Indian script character recognition: A survey," Pattern Recognition", vol. 37, no.9, pp. 1887-1899, 2004.
- [12] Ravi K Sheth, N.C.Chauhan, Mahesh M Goyani," A Handwritten Character Recognition Systems using Correlation Coefficient", selected in International conference V V P Rajkot, 8-9 April 2011.
- [13] Dewi Nasien, Habibollah Haron, Siti Sophiayati Yuhani "The Heuristic Extraction Algorithms for Freeman Chain Code of Handwritten Character", International Journal of Experimental Algorithms, (IJE), Volume (1): Issue (1)