Handwritten Character Recognition for major Indian Scripts:A Survey

Sneha U.Bohra, Satish J.Alaspurkar
Department of Computer Science
G.H.Raisoni College of Engineering & Management
Amravati,India
snehab30@gmail.com, alaspurkar.satish@gmail.com

Abstract— Character Recognition is an active field of research today. It comprises of Pattern Recognition and Image Processing. Character Recognition has gained enormous attention due to its application in various fields. Character Recognition is broadly categorized into Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR). OCR system is most suitable for the applications like multi choice examinations, printed postal address resolution etc, while application of HCR is wider as compared to OCR. HCR is useful in cheque processing in banks; almost all kind of form processing systems, handwritten postal address resolution and many more. Even though, sufficient studies have performed in foreign scripts like Chinese, Japanese and Arabic characters as well as Indian Scripts like Devnagri, still some scripts remain less contributed. In this survey, we provide a detail study of classifiers as well as various techniques proposed by different researchers for handwritten character recognition for Indian Scripts specially Devnagri, Urdu and Malayalam. This survey will serve as the first and basic step for the researchers for recognizing the characters from various regional scripts in India. It will surely provide a way for the development of recognition tools for Indian scripts. In coming days, character recognition system might serve as a key factor to create paperless environment by digitizing and processing existing paper documents.

Keywords-Character Recognition; Image Processing; Pattern Recognition; Handwritten Character Recognition (HCR); Optical Character Recognition (OCR)

I. INTRODUCTION

This section presents a brief introduction about the entire process of Character Recognition. It describes all the phases of the process followed by the Character Set of Devnagri, Gujrati,Urdu and Malayalam Script.

A. Process of Character Recognition

The entire process of character recognition consists of various phases as described in the following Fig. 1.These phases can be applied to the image that is obtained from either the scanned document or photograph.

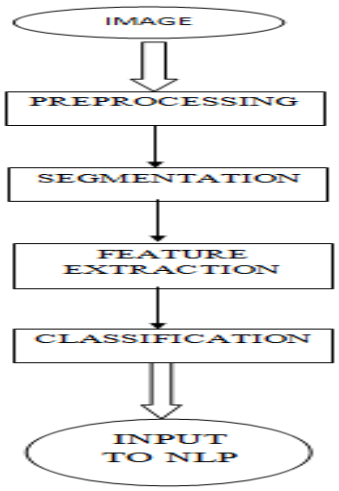

Figure 1: Block diagram of Character Recognition

- Preprocessing Phase
  This phase is applied on the image so as to acquire the image suitable for the segmentation phase. It involves various steps like removal of noise from the image, resizing the image, binarization of the image or removal of slant angle from the document etc.Various type of filtering methods like Gaussian filtering method, Min-max filtering method etc are applied for noise removal. The

process of binarization is the conversion of colored image into its black and white equivalent image. In order to speed up the process of recognition, the images are often resized.

- Segmentation Phase
  This phase is considered to be the most important phase as the final output of the entire process depends upon the output of this phase. This step decomposes the image in hierarchical way. It first segments the image into rows by applying row histogram, and then identifies words by applying column histogram and then finally the characters are recognized from the words.
- Feature Extraction Phase

  This step is the key phase for any pattern recognition application. This step intends in extracting the important features of the objects which can be used to distinguish it from the other objects. The aim of this phase is to minimize the within class variation and maximize the between class variation so that the object can be uniquely recognized.

  Various 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. The feature vector so obtained as an output of this phase is used to train the system.
- Classification
  The feature vector is given as an input to this important phase. This step classifies the vector on the basis of various classifiers. Most commonly used trained classifiers are Artificial Neural Network or Support Vector Machine. The classifiers compare the given vector with the stored pattern and give the best match as an output.

II. CHARACTERISTICS OF VARIOUS INDIAN SCRIPTS

India is a multi-lingual country containing nearly 22 official languages. These 22 scheduled languages are written using just 12 scripts. For example, Hindi, Marathi, Sanskrit, Konkani, Nepali, Dogri and Mathili are written using Devnagri scripts. .Sindhi is written using Devnagri script in India and Urdu script in Pakistan.Assamese, Manipuri and Bengali is written using Bengali script.Punjabi is written using Gurumukhi script, whereas all other languages have their own scripts. The important property of Indian scripts is that they do not have the concept of lower and upper case letters and almost all scripts except Urdu is written from left to right[1]. Following section describes the characteristics of various scripts.

A. Characteristics of Devnagri script

Devnagri script used to write Hindi, Konkani, Marathi, Nepali, Sanskrit, Bodo, Dogri and Mathili. Like most of the other Indian languages, this script is also derived from Ancient Brahmi script. It is phonetic in nature and hence writing maps sounds of alphabets to specific shapes. It is written from left to right. The horizontal line connecting the top of the characters in a word is usually used recognize the script. However, in some words, all the characters are not connected. This script is made up of 13 vowels and 33 consonants. These characters are called basic character. Modified Characters are formed when a vowel follows a consonant forming modified shapes. Depending on the vowel used, this modified shape is placed either at the left, right or at the bottom of the consonant which is often called Matras. Similarly, a consonant or vowel following a consonant forms Compound characters. Compound characters can be combinations of two consonants as well as a consonant and a vowel. Figure 2 shows vowels and consonants in Devanagiri scripts[7].

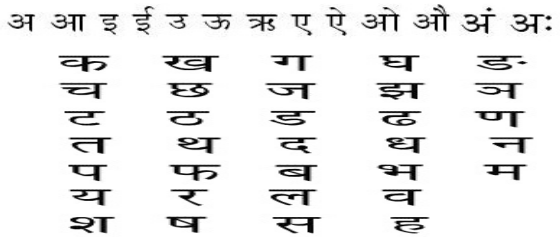

Figure 2: Vowels and Consonants in Devnagri Script.

B. Characteristics of Urdu Script

Urdu is basically an official language of Pakistan and one of the major sixteen languages in India. This language is complex to recognize as compared to above two mentioned scripts because it is constructed in the form of curves and dots. The basic character set and diacritic marks are as shown in the figure 3.It consists of 40 basic characters. This script is also written from right to left. Important characteristics of the Urdu script are that each letter has multiple forms depending on its position in the word. Another feature of Urdu is that Urdu words are written without short vowels or diacritic symbol[3].

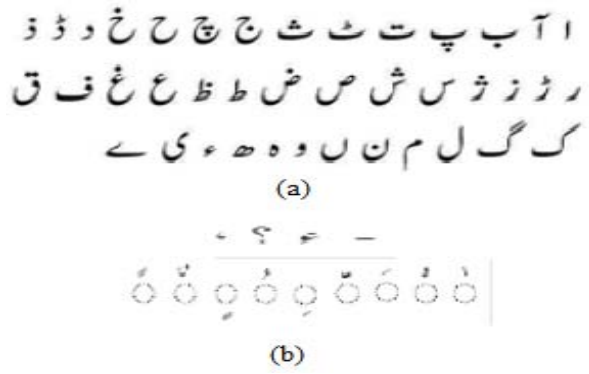

Figure 3 (a) Basic Character set and (b) punctuation and diacritic marks.

C. Characteristics of Malayalam script

Malayalam is the official language for the State of Kerala, which is derived from the Grantham script, descendant of Ancient Brahmi. The character set consists of 51 letters which includes 15 vowels and 36 consonants. The complete character set of Malayalam is depicted in Fig 1.5.The Malayalam script exhibits no inherent symmetry and thus making the recognition task very tedious[5].

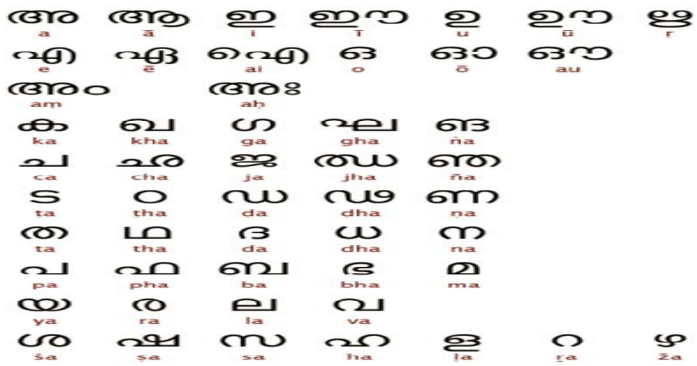

Figure 4 : Vowels and Consonants of Malayalam script.

III. DESCRIPTION OF VARIOUS CLASSIFIERS

This section describes various classifiers that classify the feature that are extracted by the Feature Extraction step and on the basis on various criteria classifies the character and recognizes it in unique way. Various classifiers are as follows:-

- Taxicab Metric: Simplest measure is the absolute difference of two patterns. Taxicab distance between two points is the sum of absolute differences of their coordinates. Let's say A is the template feature image which is to be compared with test image Y. Assume that image is of M X N. Taxicab measure between two images A and B is given as follow:

$$M_d = \sum_{i=1}^{M} \sum_{j=1}^{N} |A(i,j) - B(i,j)|$$

Taxicab metric is also known as rectilinear distance, L1 distance, city block distance, Manhattan distance etc.

- Euclidian distance / L2 Norm: Euclidean distance is widely used distance measure, which is also known as L2 norm or nearest neighbor classifier. Euclidean distance is ordinary distance between two points which could be measured using ruler .Euclidean distance between two vectors of size N is given by:

$$E_d = \sqrt{\sum_{i=1}^{N} (A_i - B_i)^2}$$

In same way, Euclidean distance between two dimensional images can be given by:

$$E_d = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (A(i,j) - B(i,j))^2}$$

This equation can be generalized for any dimensions and any norm. L norm of one dimensional vector can be given as:

$$ED_L = \sqrt[L]{\sum_{i=1}^{N} (A_i - B_i)^L}$$

When L is 2, it becomes Euclidean measure and hence Euclidean distance is also known as L2-norm.

- Chess board distance: Chess board distance between two vectors is the greatest of their differences along any coordinate dimension. It is also known as L metric or chebyshev distance. Chess board distance between two vectors X and Y with n coordinates,

$$D_{CB} = \max_i (x_i - y_i)$$

- Cross correlation: Cross correlation is similarity measure between two signals in signal processing. It also has the applications in pattern recognition to compare two patterns or to find distance or similarity between two patterns. The correlation coefficient is given as,

$$R = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(A_{ij} - \bar{A}).(B_{ij} - \bar{B})}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(A_{ij} - \bar{A})^2\right)\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(B_{ij} - \bar{B})^2\right)}}$$

Where R indicates the Correlation Coefficient between two matrices A and B, where A and B are matrices or vectors of the same size. A' and B' are mean of A and B respectively. A correlation coefficient is a numerical, descriptive measure of the strength of the linear relationship between two variables. Values for the correlation coefficient range between -1 and +1; with a correlation coefficient of +1 indicating that the two variables have a perfect, upward-sloping (+) linear relationship and a correlation coefficient of -1 showing that the two variables are perfectly related in a downward-sloping, (-) linear sense. A correlation coefficient of 0 demonstrates that the variables have no relationship, and are independent.

- Artificial Neural network (ANN): Artificial neural network is widely accepted classifier for diverse patterns. ANN works on phenomenon of biological neurons and learns to classify unseen data. Multilayer neural networks have been widely used in pattern recognition applications. Various paradigms have been used. The different network models are specified by Network topology, Node characteristics, Training rules.

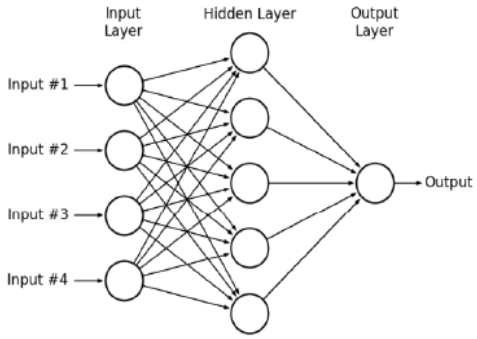

Figure 5: Architecture of ANN

Architecture of neural network depends on nature and complexity of applications. However, multilayer neural network with proper choice of parameter is capable enough to classify almost any pattern. The back propagation model or multisupervised learning technique. Typically there are one or more layers of hidden nodes between input and output nodes.

There are many parameters that control the performance of neural network, like Number of layers, Number of neurons in each layer, Transfer function used between two layers, Learning algorithm, Number of epochs etc.

- Support vector machine (SVM): Support Vector Machine is supervised learning tool which is used for classification and regression. The basic SVM takes a set of input data and predicts, for each given input. Given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category or the other. More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

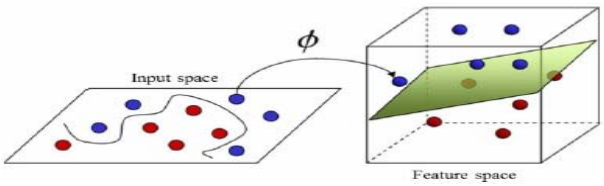

Figure 6: Feature transformation using SVM

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite dimensional space be mapped into a much higher dimensional space, presumably making the separation easier in the space[1].

IV. RECOGNITION OF INDIAN SCRIPT

This section describes a various techniques proposed by different researchers for recognition of Devnagri, Urdu and Malayalam Script.

- Devnagri Script:-

In this paper, they have proposed system that is concerned with the recognition of Devnagri numerals and basic characters only and data used in the present work has been collected from different individuals. Data was collected for both numeral recognition and character recognition. A flat bed scanner was used for digitization. Digitized images so obtained was in gray tone with 300 dpi and stored as TIF format. They used a histogram based global binarizing algorithm to convert them to two-tone (0 and 1) images ,where '1' represents object point and '0'represents background point.

Feature Extraction:-Histograms of direction chain code of the contour points of the characters was used as feature for recognition. They used 64 dimensional features for recognition purpose. The feature extraction techniques are described below.

- 64 Dimensional Feature Extraction

Given a two-tone image, first find the contour points of the image by the following algorithm. For all object points in the image, consider a 3 x 3 window surrounded to the object point. If any one of the four neighboring points (as shown in Fig.6 (a)) is a background point then this object point (P) is considered as contour point. Otherwise it is a non-contour point.

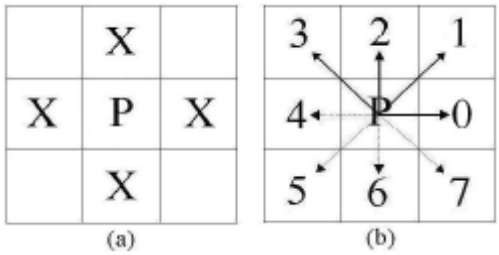

Figure. 6 (a) For a point P and its four neighbors are shown by 'X'
(b) For a point P the direction codes for its eight neighboring points

The bounding box (minimum rectangle containing the character shown in Fig. 7(b)) of an input character is then divided into 7 x 7 blocks (as shown in Fig.7(c)). In each of these blocks, the direction chain code for each contour point is noted and the frequency of the direction codes is computed. Here we use chain code of four directions only [directions 0 (horizontal), 1 (45 degree slanted), 2(vertical) and 3 (135 degree slanted)]. Refer Fig.6 (b) for illustration of four chain code directions. We assume chain code of direction 0 and 4, 1 and 5, 2 and 6, 3 and 7, are same. Thus, in each block we get an array of four integer values representing the frequencies and those frequency values are used as feature. Histogram of the values of these four direction codes in each block of a Devnagri numeral is shown in Fig.7(e) . Thus, for 7 x 7 blocks we get 7 x 7 x 4= 196 features. To reduce the feature dimension, after the histogram calculation in 7 x 7 blocks, the blocks are down sampled with a Gaussian filter into 4x4 blocks. As a result we have 4 x 4 x 4 = 64 features for recognition.Histogram of all the direction obtained after down sampling is shown in Fig.7(f).

Example of feature extraction process on a Devnagari character is shown in Fig.8.To normalize the features,maximum value of the histograms from all the blocks was computed.Each of the above features are divided by this maximum value to get the feature values between 0 and 1.

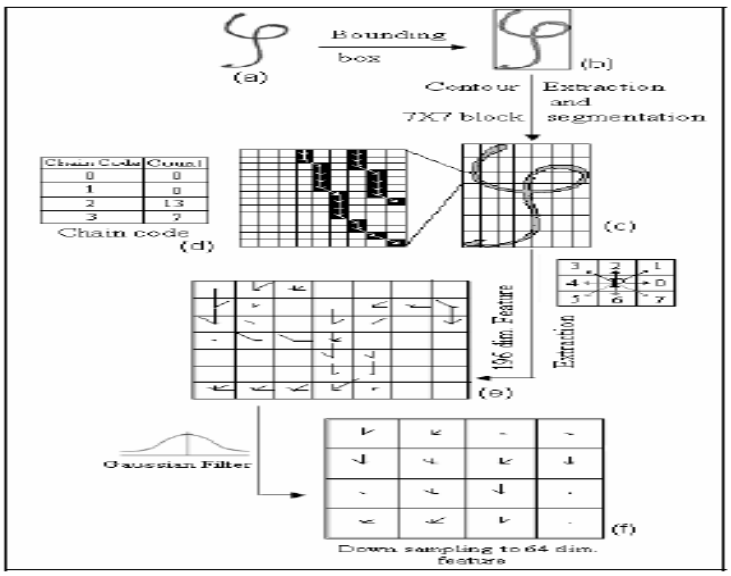

Figure 7: Pictorial representation of the 64 dimensional feature extraction process for a sample Devnagri numeral. (a) Two tone image of a Devnagri numeral 'five', (b) Bounding box of the numeral. (c) Contour of the numeral shown in black color and the bounding box is segmented into 7 X 7 blocks. (d) Chain code of a block shown in zoomed version. (e) 196 dimensional Chain code features of each block. (f) 64 dimensional features obtained after down sampling using a Gaussian filter.

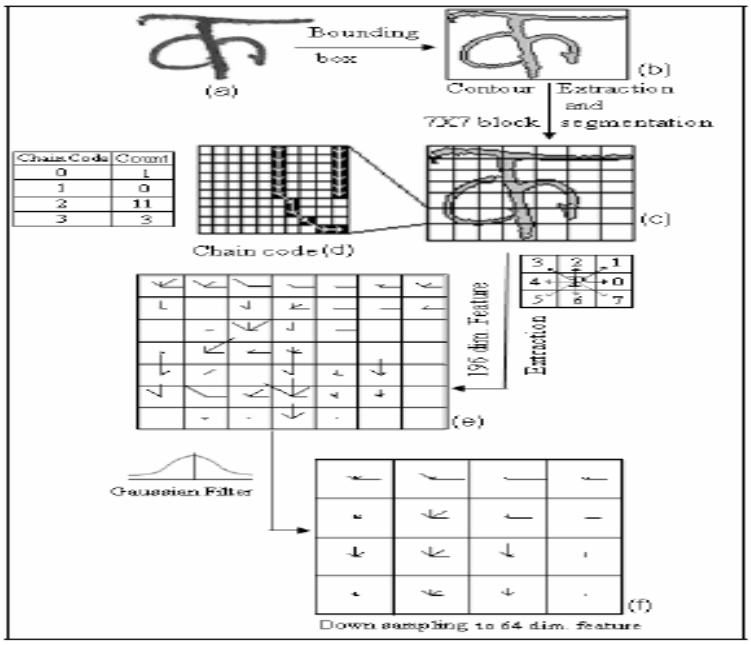

Figure 8: Pictorial representation of the 64 dimensional feature extraction process for a sample Devnagri character. (a) Two tone image of a Devnagri character, (b) Bounding box of the character. (c) Contour of the character shown in black color and the bounding box is segmented into 7 X 7 blocks. (d) Chain code of a block shown in zoomed version. (e) 196 dimensional Chain code features of each block. (f) 64 dimensional features obtained after down sampling using a Gaussian filter.

- Recognition Classifier

Many classifiers have been used by the researchers for handwritten character recognition. They used quadratic classifier for recognition purpose because it was observed that this classifier gives better results than other classifiers like Bayes classifier, subspace method etc. Descriptions of the quadratic classifier used for recognition purpose is as given below. A Modified Quadratic Discriminant function (MQDF) is used by the quadratic classifier which is defined by,

$$g(X) = (N + N_0 - n - 1)\ln\left[1 + \frac{1}{N_o\sigma^2}\left[\| X - M \|^2 - \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_i + \frac{N_o}{N}\sigma^2}\{\Phi_i^T (X - M)\}^2\right]\right]$$

$$+ \sum_{i=1}^{k} \ln(\lambda_i + \frac{N_o}{N}\sigma^2)$$

Where, X is the feature vector of an input character, M is a mean vector of samples, is the ith eigen vector of the sample covariance matrix, λi is the ith eigen value of the sample covariance matrix, n is the feature size, σ2 is the average variance of all classes, N is the average sample of all classes, and N0 is selected experimentally and we consider N0 = 3N/7 for 64 dimensional feature. They did not use all eigen values and their respective eigen vectors for the classification.They sorted the eigen values in descending order and took first 20 eigen values and their respective eigen vectors for classification. Rejection in the system was done if for a character the difference of 1stand 2nd value of g(X) is smaller than a threshold.

- Result Obtained:-From the experiments based on the above methodology, it was noted that the overall recognition accuracy of the proposed scheme for numerals is 98.86% and for characters is 80.36%[8].

- Urdu Script:-

In this paper, Moment Invariants (MI) was used to evaluate seven distributed parameters for handwritten isolated Urdu characters. In any character recognition system, the characters are proposed to extract features that uniquely represent properties of the character . The MIs are well known to be invariant under translation, rotation, scaling and reflection. They are measures of the pixel distribution around the center of gravity of the character and allow capturing the global character shape information. In the presented work, the moment invariant were evaluated using central moment of the image function f(x,y) up to third order. Urdu Characters can be grouped into single component and multi component characters as shown in Figure 9.

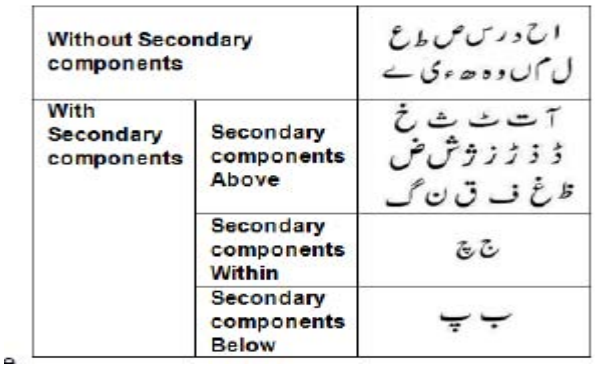

Figure 9:Component based Urdu Characters

- Methodology

First Step is to verify whether character consists of single component or more than one component. If character is single component then image is normalized into 60 X 60 and divided into three horizontal zones for features extraction as shown in figure 10 (b). From each zone 7 Moment Invariant features and from whole image 7 Moment Invariant features were computed, in this manner 28 Moment Invariant features were determined.

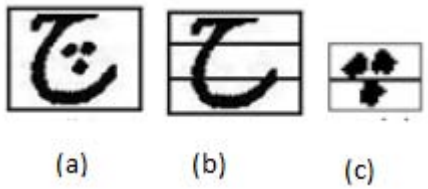

Figure 10 :(a) Character Cheem (b)Primary Component of Cheem ,
(c) Secondary Component of Cheem

Afterwards SVM was used for classification and character is put into appropriate class and finally recognized the single stork component character. If Character belongs to multi component character then first primary and secondary component were separated and stored on different locations. The image of primary stroke was normalized into 60 X 60 by maintaining height and width ratio and further divided into 3 zones. The secondary component was normalized into 22 X 22 and divided into 2 horizontal zones as shown in the above figure. After separating character into primary and secondary two alternative were considered, first is counting number of secondary strokes and their types like with one dot below, with three dots below, three dot above and one toe above. But it may give incorrect results in case of handwritten character recognition due to variation in writing secondary components. For example characters Pay should consists of a primary component (i.e.) and secondary components (i.e. three dots below). But in case of handwritten characters the structure of secondary components varies from writer to writer.

Counting number of secondary strokes and their types may give unsatisfactory results in case of handwritten characters. That why Moment Invariant of secondary component are considered as a feature instead of number of secondary components and types. The segmented secondary components were normalized into 22 X 22. Then normalized image is divided into two horizontal zones for features extraction

as shown in figure 10 (c). From each of two zone 7 Moment Invariant features and from whole image 7 Moment Invariant features were computed, in this manner 21 Moment Invariant features were determined. Later using SVM the type of secondary component is recognized which could be one of the component shown in figure 11.

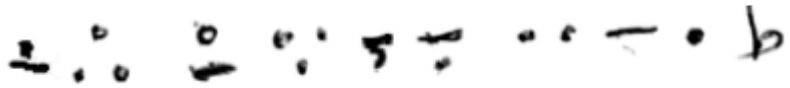

Figure 11: Possible Secondary Components

The entire flowchart of character recognition is as shown in the figure 12.

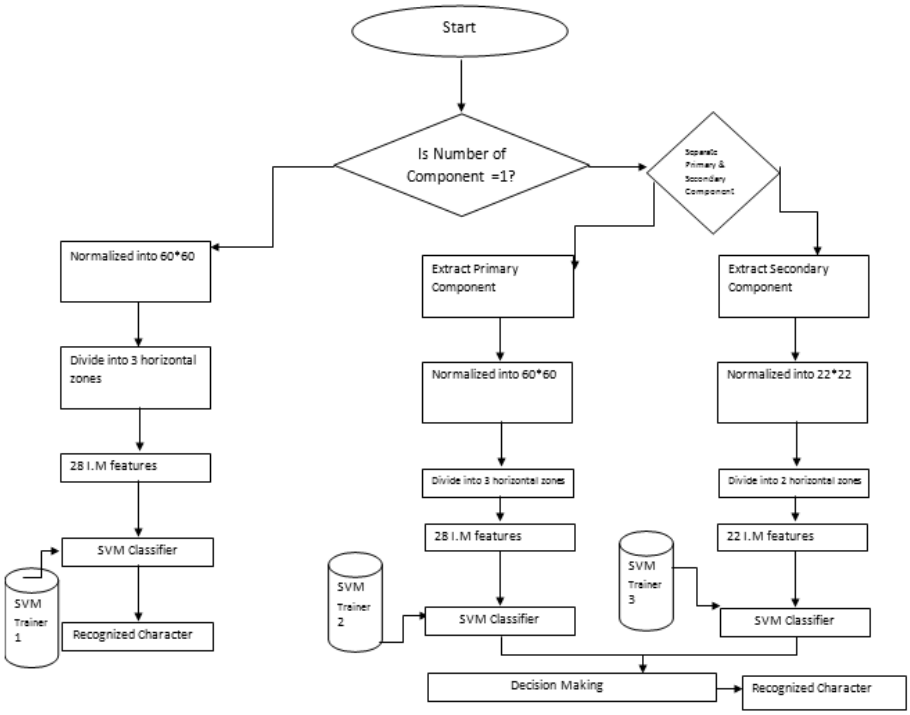

Figure 12:Entire Flowchart of Character Recognisation of Isolated Urdu Character.

- Result Obtained:- An overall accuracy up to 93.59% for all offline handwritten isolated Urdu characters has been achieved by this method[3].

- Malayalam Script:-

This method employs recognition of isolated and combinational handwritten characters in a noiseless environment. The basic principle is to identify specific terminologies in each character and extend the same to a set of characters in order to achieve accurate results with very low complexity algorithms. The separation of letters is shown in figure 4 which uses intensity variations for segregating the line and character from the scanned image.This work separate the entire character set in to three different classes. Ra type characters, Pa type characters and Special symbols. This classification is based on the shape and appearance of the character. This shape feature is extracted to recognize the letter.

- Methodology:-

Character separation:-In order to apply the algorithm for recognition of handwritten characters, segregation of scanned images is of prime importance. Rather than adopting the normal projection profile methodology, an alternative technique to identify foreground and background colors of the scanned image was incorporated. This is used to authenticate the letters written in colors other than the identified background color. The H-L notation has been adopted to represent background points and valid character path. The steps followed to segregate and to separate the individual letters from the scanned input image can be postulated as follows.

Step 1. Process the image from the top left, one segment at a time from right extreme to left extreme.

Step 2. A pixel with H intensity is authenticated as a valid point in a character.

Step 3. Horizontal, vertical and diagonal comparisons are devised to identify the constituents of the character

Step 4. The isolated character may be restructured to a window with special emphasis on boundary points.

Step 5. The dynamic widow is then processed and various checks are applied to identify the character.

Step 6. Once identification is completed, the character is inserted in the correct position in the sequence identified on comparison with the position coordinates.

Step 7. The processed window is isolated and the rest of the document is scanned for the next character.

Step 8. On encountering the next significant intensity, the above steps are applied and process is repeated until the entire document is processed.

In the case of combinational characters, we propose a way to identify the isolated as well as the complex connected script of Malayalam language in a noiseless environment. The flow chart of the system to recognize the characters is illustrated in figure 4.8. Here in our study we start with the assumption that to find an isolated character. On successful the corresponding character is recognized and reported. In case of a connected character, we take the recognition process a step higher and will try to segregate the character into its corresponding counterpart and analyze each segment individually.

Initially we study the image and our first step will be to separate the characters assuming the connected character as a single character using any of the character separation algorithms and enclosing it in a matrix which we will be analyzing in the later part of our study. If check for an isolated character fails then we will analyze the pattern in the pattern analysis phase where we will classify the connected characters into three modes.

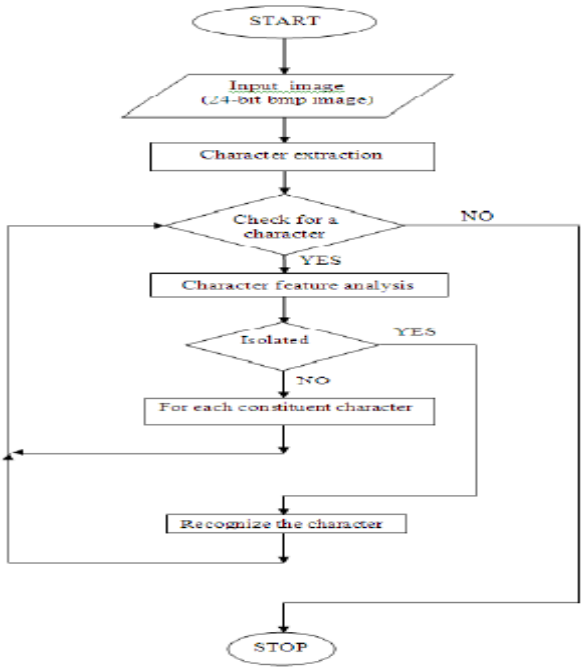

Figure 13:Flowchart for Combinational HCR System

Feature extraction technique:-Once the segregation is accomplished, the feature extraction process is initialized. The length and breadth of each character can be calculated by manipulating the HLH intensity values of the segregated image, which in turn, is stored in a dynamic window matrix. Inferences are arrived at on the basis of the sequence pattern procured on horizontally processing the dynamic matrix. Furthermore, the pattern with highest probability is identified. The matrix is then processed for vertical as well as straight line patterns.

Consider the character depicted as in figure 14. The intensity pattern HLH can be observed in the letter and hence infer that two vertical pillars exist on processing the image horizontally.

This work separates the entire character set in to three different classes. Ra type characters, Pa type characters and Special symbols. This classification is based on the shape and appearance of the character. This shape feature is extracted to recognize the letter.

Figure 14: Horizontal HLH Pattern

- Recognition strategy

On analyzing different input samples, it is observed that when the image is processed horizontally, a HLH occurrence as a subset of the sequence pattern is observed. Hence the character is inferred to be "ra" type. After completing the horizontal processing, the character is vertically processed as many times as many recurrences were observed. A HL pattern near the top leads to the inference that there is a probable letter path. The characters are classified as "Pa" type on identifying a vertical or a horizontal line and require special sequence identification checks to recognize the different "ra" type patterns within it. This includes characters which has the characteristics of either "Pa" and "Pa" type characters or both. Special characters require horizontal, vertical and a few diagonal sequence checks to identify the character. Algorithm for recognizing a character is shown below.

1) Identify the characteristic window

2) Apply horizontal check and find the high probabilistic recurring times of HLH patterns

3) Identify and classify the character into "ra", "pa" or special type characters.

4) The view port window is segmented corresponding to recurrences and each segment is further investigated.

5) Vertical processing is applied to each segment to determine the specific intensity sequences.

6) Diagonal cross-sectional analysis may be applied for special characters.

We have go in for the recognition based on the connected characters into the above 3 categories. The isolated character recognition specified in the algorithm is based on the HLH intensity patterns. After extracting the character into a matrix, if it is a normal isolated character it is recognized. In the case of vertical connected characters the letter gets horizontally partitioned based of most probabilistic occurrence of the high intensity .The division of character into various sized small elements are carried out and each small part is further analyzed to find if it makes up to an isolated letter. When we succeed in getting both the combination of letters in a pre-expected manner and both are found to be liable the particular connected character is written in the modern style applicable for it. In the case of a horizontal recurrence vertical partitioning takes place and the particular letter sequence is identified. In special recurrence characters we will use the HLH intensity patterns to understand the characteristics and special vertical checks and horizontal checks are applied on the character as a whole and on the parts and the correct letter sequence gets identified.

The algorithm for the system is shown below. Here check for isolated characters are performed and horizontal and vertical checks are done. Based on this the character is placed in any of the three categories.

Step 1 : Extract the character into a matrix.

Step 2 : Check for isolated character occurrence.

Step 3 : If true display the correct character and go to step 10.

Step 4 : Check the length by width ratio of the matrix.

Step 5 : If length < width then goes to step 7.

Step 6 : Horizontal recurrence analysis is carried out .If successful recognition go to step10.

Step 7 : Vertical recurrence analysis is carried out. If successful recognition go to step 10.

Step 8 : Special recurrence algorithms are undertaken.

Step 9 : The character stands unidentified.

Step 10 : Stop.

In case of connected character we try to separate them into one of the following 3 categories. The horizontal recurrence, vertical recurrence or special recurrence. In horizontal recurrence two characters are combined horizontally, in vertical recurrence vertically and in special recurrence the characters may have no inherent characters of the combined characters but still a combination of the original letters.

- Result Obtained:- The overall efficiency of 92% has been achieved[6].

V. CONCLUSION

This survey can serve as a base for various other researchers so that they can adopt Character Recognition of various Indian Scripts as their domain of research. This study described various classifiers for the Indian scripts .It is also observed that Gujrati script still lacks behind in Handwritten Character Recognition.

ACKNOWLEDGMENT

The success of this paper needed co-operation & guidance of number of people. I, therefore consider it as my prime duty to thank all those who had helped me through this venture. It gives me immense pleasure to express gratitude to Guide Prof. S. J. Alaspurkar, who provided me constructive and positive feedback during the preparation of this paper. I express my sincere thanks to all other staff members of Computer Science and Engineering Department for their kind co-operation. Lastly,I express my ineptness to all those who directly or indirectly helped me in making the completion of this paper possible.

REFERENCES

[1] Mansi Shah & Gordhan B Jethava ,"A Literature Review On Hand Written Character recognition" Indian Streams Research Journal.Vol -3 ISSUE –2, March.2013,ISSN:-2230-7850.

[2] Ajmire P.E,R.V Dharaskar,V M Thakare ," A Comparative Study of Handwritten Marathi Character Recognition " International Journal of Computer Applications (IJCA) 2012.

[3] Imran Khan Pathan,Abdulbari Ahmed Ali & Ramteke R.J.,"Recognition of Offline Handwritten IsolatedUrdu Character " Advances in Computational Research ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, 2012, pp.-117-121 .

[4] Rajbala Tokas & Aruna Bhadu, "A comparative analysis of feature extraction techniques for HCR",International Journal of Advanced Technology & Engineering Research (IJATER) ,2012.

[5] Jomy JohnPramod K. V, Kannan Balakrishnan, "Handwritten Character Recognition of South Indian Scripts:A Review" National Conference on Indian Language Computing, Kochi, Feb 19-20, 2011.

[6] Abdul Rahiman & Rajasree M S, "An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations" International Journal of Computer Theory and Engineering, Vol. 3, No. 3, June 2011.

[7] U. Pal T. Wakabayashi2, F. Kimura,"Comparative Study of Devnagari HCR using Different Feature and Classifiers", 2009 10th International Conference on Document Analysis and Recognition.

[8] N. Sharma, U. Pal, F. Kimura, and S. Pal," Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier", ICVGIP 2006, LNCS 4338, pp. 805 – 816.