



A Review on Handwritten Character Recognition

Deepashree Keshav Mehendale

Assistant Professor, Dept. of Computer Science, Dr. D. Y. Patil ACS College, Pimpri, Pune, India

ABSTRACT: Handwritten Character recognition also known as HCR is a challenging and an emerging field which has gain a wide importance due to its use in various fields. Character recognition is one of the most fascinating areas of pattern recognition and artificial intelligence. There are different methodologies used for character recognition. Document verification, digital library, reading bank deposit slips, reading postal addresses, extracting information from cheque, data entry, applications for credit cards, health insurance, loans, tax forms etc. are application areas of digital document processing. This paper gives a review of research work carried out for recognition of hand written characters. In Handwritten text there is no constraint on the writing style. Handwritten letters are difficult to recognize due to diverse human handwriting style, variation in angle, size and shape of letters. This paper discusses the working principle and the various approaches of handwritten character recognition.

KEYWORDS: HCR (Handwritten character recognition), pattern recognition, artificial intelligence working principle, methodologies

I. INTRODUCTION

Handwriting recognition is the ability of a computer to receive and interpret handwritten input from sources such as paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. Alternatively, the movements of the pen tip may be sensed "on line", for example by a pen-based computer screen surface. In offline character recognition system, document is first generated, digitized, stored in computer and then it is processed. While in case of online character recognition system, character is processed while it was under creation. External factors like pressure speed of writing have an influence in case of online system. Based on that system can be classified as OCR or HCR respectively. Accuracy of HCR is still limited to 90 percent due to presence of large variation in shape, scale, style, orientation etc.

The first important step in any handwritten recognition system is preprocessing followed by segmentation and feature extraction. Preprocessing includes the steps that are required to shape the input image into a form suitable for segmentation. In the segmentation, the input image is segmented into individual characters and then, each character is resized into $m \times n$ pixels So as to extract the features. The Selection of appropriate feature extraction method is probably the single most important factor in achieving high recognition performance. Several methods of feature extraction for character recognition have been reported in the literature. The widely used feature extraction methods are Template matching, Image transforms Graph description, Contour profiles, Zoning, Geometric moment invariants, Fourier descriptors, Gradient feature and Gabor features. An artificial neural Network as the backend is used for performing classification and recognition tasks. In the off-line recognition system, the neural networks have emerged as the fast and reliable tools for classification towards achieving high recognition accuracy.

II. WORKING

Working Principle Any character recognition system goes under following steps, i.e. Image acquisition, Preprocessing, Segmentation, Feature extraction, classification and post processing.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Block Diagram of Character Recognition System

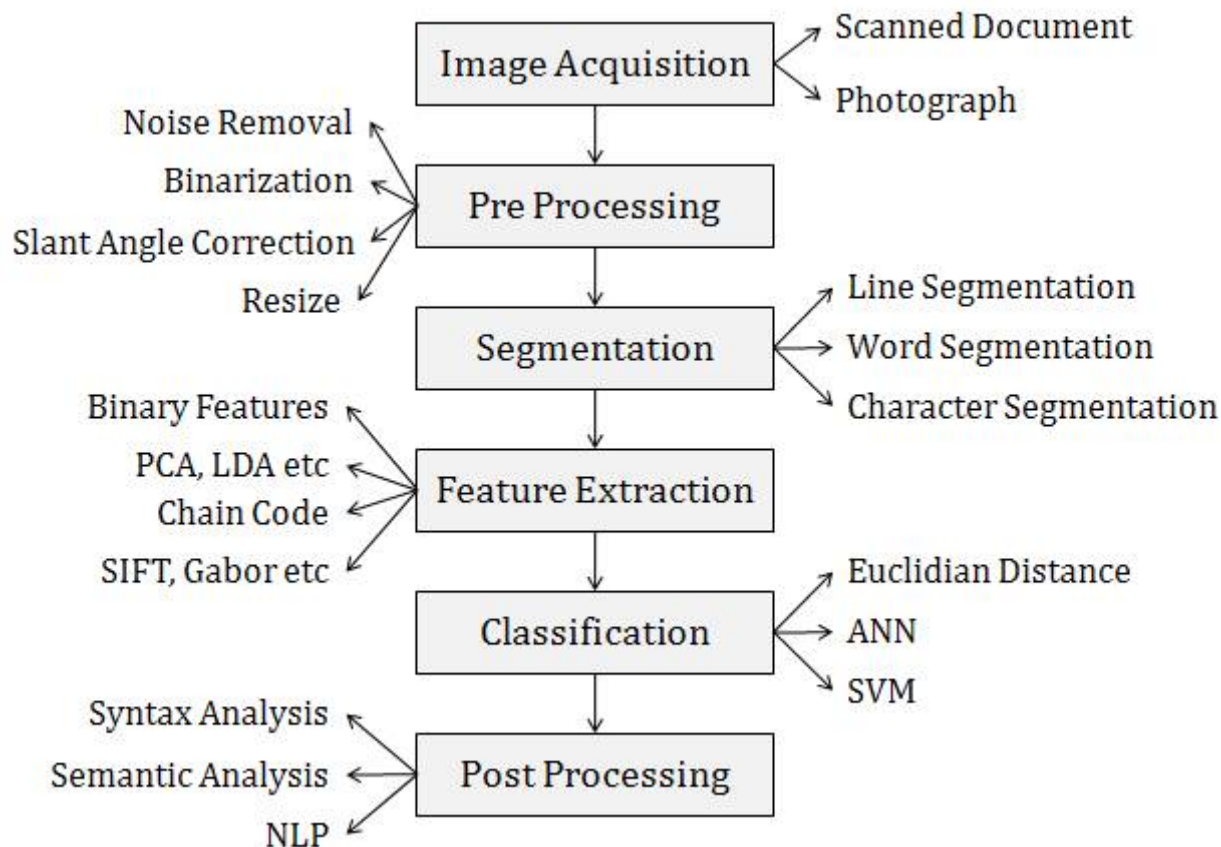


Image acquisition: Images for HCR system might be acquired by scanning handwritten document or by capturing photograph of document or by directly writing in computer using stylus. This is also known as digitization process.

Preprocessing: Preprocessing involves series of operations performed to enhance to make it suitable for segmentation. Preprocessing step involves noise removal generated during document generation. Proper filter like mean filter, min-max filter, Gaussian filter etc. may be applied to remove noise from document.

Binarization: process converts gray scale or colored image to black and white image. Binary morphological operations like opening, closing, thinning, hole filling etc. may be applied to enhance visibility and structural information of character. If document is scanned then it may not be perfectly horizontally aligned, so we need to align it by performing slant angle correction. Input document may be resized if it is too large in size to reduce dimensions to improve speed of processing. However reducing dimension below certain level may remove some useful features too.

Segmentation: Generally document is processed in hierarchical way. At first level lines are segmented using row histogram. From each row, words are extracted using column histogram and finally characters are extracted from words. Accuracy of final result is highly dependent on accuracy of segmentation.

Feature Extraction: Feature extraction is the base of pattern recognition application. Feature extraction techniques like Chain Code (CC), Scale Invariant Feature Extraction (SIFT), zoning, Gradient based features, Histogram might be applied to extract the features of individual characters. These features are used to train the system.

Classification: When input image is presented to HCR system, its features are extracted and given as an input to the trained classifier like artificial neural network or support vector machine. Classifiers compare the input feature with stored pattern and find out the best matching class for input.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Post processing: This step is not compulsory; sometimes it helps to improve the accuracy of recognition. Syntax analysis, semantic analysis kind of higher level concepts might be applied to check the context of recognized character.

III. LITERATURE REVIEW

In research article Irregular Pattern Recognition Using Artificial Neural Network [1] two tool boxes are used Matlab tool and Neural Network tool box. Here the focus is on using Artificial Neural network technique for character recognition. This technique is able of recognizing characters with irregular pattern. As Artificial neural networks are self-learning and self-organized network systems in the pattern recognition the neural network system is trained to identify the most similar pattern from the database that are similar to the input image. DWT Discrete Wavelet Transform is used for feature extraction of handwritten characters. For the purpose of classification and recognition artificial neural network is used. It is the fastest algorithm for pattern recognition problems and the memory requirements for this algorithm are relatively small in comparison to the other algorithms. This method achieves 98.7% of accuracy.

In Handwritten Character Recognition Using Method Filters [2] three method filters for recognition viz: - intersection count method, shape enclosed method and centroid based method have been discussed. In intersection count method the number of intersection points made by intersecting edges in a character are calculated, then based on the number of intersection points the characters are recognized. In Shape enclosed method the shapes enclosed in the character are found and depending on the shape, character is identified. In the centroid based method the grids are taken as points the grid blocks with black pixels are considered as points. The centroid is found based on the formulae depending on the centroid calculated this value is compared with the centroid value in the database and then the character is recognized. These three methodologies are combined with matrix matching method. All these methods are used sequentially and as the process continues, the number of possible matches keeps decreasing steeply and this makes the process fast and efficient. The number of comparison decreases as compared to the matrix matching method because the combination of the four methods act as a filter. By using this method the number of comparisons done are reduced. Thus this method is more simple and efficient method of character recognition.

In the research paper Diagonal Feature Extraction Based Handwritten Character system using Neural Network [3] the focus is on diagonal based feature extraction every character image of size 90 x 60 pixels is divided into 54 equal zones, each of size 10x10 pixels. The features are extracted from each zone pixels by moving along the diagonals of its respective 10X10 pixels. Each zone has 19 diagonal lines and the foreground pixels present along each diagonal line is summed to get a single sub feature and thus 19 sub-features are obtained from the each zone. These 19 sub-features values are averaged to form a single feature value and placed in the corresponding zone. This procedure is sequentially repeated for the all the zones. There could be some zones whose diagonals are empty of foreground pixels. The feature value corresponding to these zones are zero. Finally, 54 features are extracted for each character. In addition, 9 and 6 features are obtained by averaging the values placed in zones row wise and column wise, respectively. As a result; every character is represented by 69 features, that is, 54 +15 features it is identified that the diagonal method of feature extraction yields the highest recognition accuracy of 98% for 54 features and 99% for 69 features.

In Printed and Handwritten Character & Number Recognition of Devanagari Script using Gradient Features [4] research work two operators Sobel and Robert are discussed. The gradient is calculated in 8,12,16,32 directions. The Sobel operator uses two templates to compute the gradient components in horizontal and vertical directions respectively.

$$g_v(i,j) = f(i-1,j+1)+2f(i,j+1) + f(i+1,j+1) - f(i-1,j-1) -2f(i,j-1) -f(i+1,j-1)$$
$$g_h(i,j) = f(i-1,j-1)+2f(i-1,j) + f(i-1,j+1) - f(i+1,j-1) -2f(i+1,j) -f(i+1,j+1)$$

The gradient strength and directions are calculated as follows

$$G(i,j) = \sqrt{g_v^2(i,j) + g_h^2(i,j)}$$
$$\Theta = \arctan g_v(i,j) / g_h(i,j)$$

The Robert templates are as follows:-

$$g_v(i,j) = f(i+1,j) - f(i,j+1)$$
$$g_h(i,j) = f(i+1,j+1) - f(i,j)$$

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Result using Sobel and Robert operators

L	8	12	16	32
Handwritten Dataset (Character) using Sobel operator	94 %	94.76%	96%	97%
Handwritten Dataset (Character) using Robert operator	94.45 %	95.06%	95.67%	96.09%

Sobel operators gives a bit higher accuracy compared to Robert operator for handwritten character recognition.

Handwritten Character Recognition Using Twelve Directional Feature Input and Neural Network [5] research work focuses on calculating twelve directional features inputs depending upon gradient. A back propagation feed-forward neural network is used to recognize the handwritten characters. After training the network with back-propagation learning algorithm, high recognition accuracy can be achieved. The direction has been divided into 12 regions with each region covering angle of 30 degree, hence direction value of any pixel may have only 12 values assigned from 1 to 12. This approach increases the information content and gives better recognition rate with reduced recognition time. First the gradient is calculated and then these gradient values are mapped with the 12 direction values with angle of 30 degree. This method provides high accuracy and requires less training time for handwritten character recognition.

In the research work Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid Feature Extraction [6] hybrid feature extraction for offline handwritten character recognition has been proposed. A hybrid of structural, statistical and correlation features is used for character recognition. In the first step, the type and location of some elementary strokes in the character are identified. The strokes comprises of horizontal, vertical, positive slant and negative slant lines—as we observe that the structure of any character can be approximated with the help of a combination of simple straight line strokes. The strokes are identified by correlating different segments of the character with the chosen elementary shapes. These normalized correlation values at different segments of the character give correlation features. For making feature extraction more robust, we add in the second step certain structural/statistical features to the correlation features. The added structural/statistical features are based on projections, profiles, invariant moments, endpoints and junction points. This enhanced, powerful combination of features results in a 157-variable feature vector for each character, which we find adequate enough to uniquely represent and identify each character. The extracted feature vector is used during the training phase for building a support vector machine (SVM) classifier. The trained SVM classifier is subsequently used during the testing phase for classifying unknown characters the obtained results show higher efficiency regarding classifier accuracy, memory size and training time.

Hand Written English Character Recognition using Column-wise Segmentation of Image Matrix (CSIM) [7] in this research the image is converted into a binary matrix of size 80 * 80. This matrix is compressed to 10 * 10 matrix by using compression algorithm. The 10 x 10 matrix is segmented column-wise into 10 segments of size 10 each. All the columns of a particular character are mapped into identical patterns used to recognize that particular character. A counter is used to find out the number of identical patterns produced by the test character. Majority of the identical pattern present in a test character decides the existence of a particular character.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

IV. CONCLUSION

In the research work carried out various methods of character recognition have been discussed. The table below gives the brief description of all the methods discussed

Sr. No	Author	Preprocessing	Segmentation	Feature Extraction	Classification	Features
1	Sujitshree, Venugopal	Noise reduction, Normalization and Binarization and	Edge based and Region based segmentation	Discrete wavelet transform	Artificial Neural Network	Able to recognize characters with irregular pattern with 98.7% of accuracy.
2	Nikita Sarkar, Harsh Kishore	Noise reduction	Segmentation based on method used of feature extraction.(finding intersection points, shape enclosed)	Intersection count, Centroid and Shape enclosed.	Filter Methodology	Efficient both with respect to time as well as space and leads to greater accuracy than applying just one complex method.
3	J Pradeep, E Srinivasan, S Himavathi.	Noise removal, Binarization	The preprocessed input image is segmented into isolated characters by assigning a number to each character using a labelling process.	Diagonal Feature Extraction	Feed forward back propagation neural network	98% for 54 features and 99% for 69 Features.
4	Anilkumar Holambe, Dr Ravinder Thool, Dr S.M Jagade	rectification of distorted images, improving the quality of images	Edge determination	Gradient feature Sobel and Robert operator	SVM	Sobel -97% Robert-96.09%
5	Dayashankar Singh, Sanjay Singh, Dr Maitreyee Dutta	Noise removing, Skeletonization and Normalization	Thresholding	Gradient Feature Extraction	12 Directional Features feed forward neural network	high accuracy and less training time for handwritten character recognition.
6	Muhammad Naeem Ayyaz Imran Javed and Waqar Mahmood.	Noise Removal, Filtering	Thresholding, Boundary tracing	Correlation function based features	SVM based classification	Able to capture local and global variations in handwritten character styles. higher efficiency with respect to speed, memory, and classification accuracy.
7	Rakesh Kumar Mandal, N R Manna.	Compression of matrix	Column wise Segmentation	Extraction based on patterns in columns of characters.	Neural Network	This method helps to obtain a sharp eye in the identification of the pattern



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

There are various methods of character recognition. Choosing the best methodology is a challenge. HCR is a challenging task as the handwritten characters varies in shape, size for every individual

REFERENCES

1. Sujitshree, C Venugopal- Irregular Pattern Recognition Using Artificial Neural Network, International Journal of Advanced research in Computer Science and Management Studies Volume 3 Issue July 2015.
2. Nikita Sarkar, Harsh Kishore - Handwritten Character Recognition Using Method Filters-International Journal of Research in Engineering and Technology pISSN: 2321-7308
3. J Pradeep, E Srinivasan, S Himavathi.- Diagonal Feature Extraction Based Handwritten Character system using Neural Network. International Journal of Computer Application Volume 8 No October 2010.
4. Anilkumar Holambe, Dr Ravinder C Thool, Dr S.M Jagade - Printed and Handwritten Character & Number Recognition of Devanagari Script using Gradient Features. International Journal of Computer Application Volume 2 No- June 2010.
5. Dayashankar Singh, Sanjay Kr Singh, Dr Maitreyee Dutta. - Hand Written Character Recognition Using Twelve Directional Feature Input and Neural Network, International Journal of Computer Application, Volume 1- No 3
6. Muhammad Naeem Ayyaz, Imran Javed and Waqar Mahmood - Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid Feature Extraction, Pak J Engg & Appl Sci Vol 10 Jan 2012
7. Rakesh Kumar Mandal, N R Manna.- Hand Written English Character Recognition using Column-wise Segmentation of Image Matrix (CSIM),
8. S. K. Thilagavathy and Dr. R. Indra Gandhi.- Recognition of distorted character using edge detection algorithm, International Journal of Innovative Research in Computer and Communication Engineering, Volume 1 Issue 4 June 2013.
9. Pratik Madhukar Manwatkar, Dr. Kavita R. Singh. Text recognition from images. International Journal of Advanced Research in Computer Science and Engineering, Volume 4 Issue 11 November 2014.
10. Sandeep Dangi, Ashish Oberoi, Nishi Goel- Performance Comparison between different feature extraction techniques with SVM using Gurumukhi script. Int Journal of Engineering and Research and Applications Vol 4 Issue July 2014
11. Arya A, Anil Kumar -A Review on Geometrical Analysis in Character Recognition. IOSR Journal of Computer Engineering Volume 1, Issue 2 Mar- Apr 2015.
12. Bharatratna P Gaikwad- Critical Review on Handwritten character recognition by using several classifiers. International Journal of Advanced Research in Computer Science and Engineering Volume 5 Issue 8, August 2015.

BIOGRAPHY

Mrs. Deepashree Mehendale is a Assistant Professor in Computer Science Department, Dr. D.Y. Patil ACS college, Pimpri, Pune. She received M.C.S., degree in the year 2002 from Pune University and also qualified the UGC-NET Exam in 2009, from Pune India. Her research interests are Image Processing, Cryptography and Information Security.