A Simple Feature Extraction Method for Analysis of Hand Written Characters

S.VIJAYPRASATH
Associate Professor, Department of Electronics and Communication Engineering, PSNA College of Engineering & Technology, Dindigul, Tamilnadu- 624001, India
vijayprasathme@psnacet.edu.in

Abstract. The printed and handwritten alphabets as character recognition by feature extraction are described during this proposed work. A predefined set of 108 samples each containing twenty-six alphabets taken by paper handwritten are taken as set for training. The system portrayed is a good alternative solution for HCR & plays well with detection of handwritten characters. Here, the method experiences binarization and pre-processing for the segmentation stage. Individual characters after segmentation undergo feature extraction part. The artificial neural network is trained using the extracted features. A simple ANN is used as a classifier utilizing the average epochs and MSE for detecting the HCR recognition without a false alarm.

1. Introduction
HCR method is a wonder that can be implemented into a large-scale postal sorting system. We need to undertake a factor that there may be a demand in the future for 10000 postal drafts to be displaced to diverse locations within a state. Sorting those types of requirements within a single day requires more manpower and a lot of support. In that case, HCR will provide support by understanding individual handwriting written on a postal address that will be dispatched to different locations. Sandeep proposed a brand new Forty PTFE method which is utilized for HCR recognition [9]. When we focus upon HCR in the literature instead of concentrating over a single pixel at a time there is a probability that neighbourhood pixels with different weights are analysed with direction-based orthogonality detection for autonomous movement of segments [2]. Another HCR method made use of DBFE for the HCR system that makes use of alphabets the foremost important think about accomplishing high recognition performance is perchance supported the selection of the suitable feature extraction method [4].

The other HCR method use splitting part by 4 as a computation LFF, CCH besides with shadow points to extract the vector. With a mind of building an output, they have synchronized tremendous sample data of around 4800, the HCR rate observed is 92.80% [11]. An interesting HCR method has proposed a ZBF based approach to bring the language diverse nature using left cursive kannada vowels as pattern for a vector. They created a route map before KNN is used and then other features like density weights were proposed for features to be given as an input to a MATLAB SVM Tool kit to recognize HCR in different situations [6]. Authors have also had pinpointed a proposed Modified Back Propagation Algorithm for the popularity of scanned image conversion to OCR.

The authors pinpointed the plus and minus of assorted OCR methods. This algorithm was utilized in neural network. This method computed error rate efficiently. This has significant improvement in ANN. This method provided 100% accuracy in OCR [7]. Another interesting & the base for all classification of HCR method using 6 step algorithms for Capital letters from the digital picture obtained was realized NNT method available. Both Sobel and Pitwet detectors with CE use of extraction are discussed here [5]. The authors have proposed that a supervised learning observatory technique may be utilized by the rear propagation model or multilayer perceptron by considering the simulation parameters like scaled conjugate training, MSE, Histogram & NN training inputs so as to scale back errors between the originality and received figures output. A new BPA algorithm was used and it focused on minimizing the educational time [8].
2. Proposed methodology

The key feature of this proposed method lies in the feature extraction that provides a considerable efficiency compared with existing methods. Statistics on various handwritten styles of different people are considered and the features among them are displaced. We have performed a simple analysis of running letters.

![Proposed Flow diagram](image)

**Fig. 1 Proposed Flow diagram**

2.1. Image Acquisition

Digital Picture from the postal address or a mailbox with handwriting is the source of this acquisition methodology. Input format is accepted as JPEG, BMT etc. Acquisition is through devices like scanner, digital camera, etc.

2.2. Preprocessing

Whatever nature or high definition scanned picture is given as input this has to be converted into the machine understanding for improving the raw data. Defined rules involved for pre-processing an image are as follows

- Noise removal
- Binarization
- Edge detection
- Dilation

2.3. Segmentation

The categorization of HCR is disintegrated into individual characters. In our methodology, bounding box segmentation is used. The character is confined only towards the BBR & supporting dimensions and so this method is likely appropriate to add labels to every handwritten characters. BBR based segmented characters are normalized. The Butterworth high pass filtering and skeletonization are performed. This method uses a skeletal residue to remove foreground pixels and preserves the original handwritten characters.

2.4. Feature Extraction

The initial foremost step involved is to select the universe of Discourse (UOD). The reason for catching UOD is to fix a matrix that fits to draw points to create a skeleton for the alphabet present in the camera image obtained through acquisition to preserve their independency.

2.4.1. Zoning Starters & Intersections.

Next to UOD zoning is applied to get handwritten character skeleton. This extraction method can be compared with a line following technique (LFT) to define a Starter. Points in LFT begins with a single neighborhood that are the origin of an alphabet in an OCR. A skeleton is a most interesting feature that supports starters to identify the surrounding pixels in order to depict the intersection points. All the new line segments will get terminated when IS found with the pixel
2.4.2. Character Traversal (CT).

Once all (9*9) windows obtained after UOD then this is followed by CT. There are three steps involved in achieving the CT for alphabet Skelton processing.

- Step 1: Split the windows of equal size (Z1 to Z9) and search for the starters and IS
- Step 2: Draw the Line segment (LS) of scanned alphabet joining starters and IS
- Step 3: Group all the immediate LS together to analyze the path of traversal.

2.4.3. Distinctive Line Segments.

Once CT has been completed there are certain rules to be applied to find out DLS based upon four directions and their normalized diagonal pixels. The direction type and its corresponding line for drawing a FE vector is depicted in table 1. Feature Vector (FV) is one of the predominant factors that is usually analyzed through the data obtained from the line segments. This FV is independent for irrespective zones. Utilizing the role of image matrix obtained image is zone partitioned (9 / 3).

| Direction Type | Line          |
|----------------|---------------|
| 1 or 5         | Vertical      |
| 3 or 7         | Horizontal    |
| 2 or 6         | Right Diagonal|
| 4 or 8         | Left Diagonal |

3. Simulation results

After getting the image as input the foremost acquisition step is carried out and the results are shown in figure 4 and 5. The General model of training an ANN is depicted in figure 6. A comparative plot with different epochs (167) and Mean Square error is portrayed as plots in below figures (7). Performing a rigorous FV with NN the outputs are displayed in below figures (9).
These two images after extracting the FV are then trained to find out the exact iteration time. Here it is obtained at 151.

**Training results:**
- Input nodes: 108
- Output nodes: 26
- Training epochs: 1000
4. Conclusion
One of the easiest HCR recognition methods is proposed here using the Zonal & Geometry FE technique. Here the simulation outputs provide an accuracy of 98.9% for the individual character of diverse persons. This method can be still improved by reducing the delay in ML techniques by perfect training to pertain the efficiency in real-time applications.

5. References

[1] Bhattacharya U and Chaudhuri BB 2009 Handwritten Numeral Databases of Indian Scripts and Multistage Recognition of Mixed Numerals IEEE Transactions on Pattern Analysis and Machine Intelligence, 3 444- 457
[2] Satish Kumar 2016 Features Based on Neighbourhood Pixels Density - A Study and Comparison/ICTACT Journal of Image and Video Processing, 06(3) 1159 - 1166
[3] Vikas J and Vijay H 2010 A Review of Research on Devnagari Character RecognitionInternational Journal of Computer Applications. 12 8-15
[4] Pradeep J, Srinivasan E and Himavathi S 2011 Diagonal Based Feature Extraction For Handwritten Alphabets Recognition System Using Neural NetworkInternational Journal of Computer Science & Information Technology327-38
[5] Kauleshwar Prasad, Devvrat C,Ashmika Lakhotiya and Dheeren Umre 2013 Character Recognition Using Matlab’s Neural Network ToolboxInternational Journal of u- and e- Service, Science and Technology. 613 - 20
[6] Dhanda B V, Gururaj Mukrambi and Mallikarjun Hangarge 2012 Handwritten Kannada Vowels and English Character Recognition SystemInternational Journal of Image Processing and Vision Sciences1(1) 12 - 17
[7] Jitendra Shrivastav, Ravindra Kumar Gupta and Shailendra Singh 2015A Modified Back Propagation Algorithm for Optical Character RecognitionInternational Journal of Advanced Computer Technology2(6)
[8] Mansi Shah, Gordhan B and Jethava 2013A Literature Review On Hand Written Character Recognition’, Indian Streams ResearchIndian Streams Research3 1- 13
[9] Sandeepr Saha, Nabarag Paul, Sayam Kumar Das and Sandip Kundu 2013 Optical Character Recognition using 40-point Feature Extraction and Artificial Neural NetworkInternational Journal of Advanced Research in Computer Science and Software Engineering, 3495 - 502.
[10] Pal U and Chaudhuri BB 2004 Indian script character recognition: a surveyPattern Recognition37(9) 1887 - 1899
[11] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, Dipak Kumar Basu and Mahantapas Kundu 2008 Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character RecognitionIEEE Region 10 Colloquium and the Third International ConferenceICIS pp 8 – 10.