Handwritten Character Recognition using Radial Histogram

Lipi Shah[1], Ripal Patel[2], Shreyal Patel[3], Jay Maniar[4]
Electronics & Telecommunication[1, 2, 3, 4], BVM Engineering College[1, 2, 3, 4]
Email: shahlipi9@gmail.com[1]

Abstract- This paper includes a proposed technique for the feature extraction using radial histogram and classification using Euclidean Distance Classifier for Gujarati handwritten Character. It includes simple pre-processing steps like binarization & normalization. Once these steps are applied on handwritten character image then we find radial histogram of character image in 72 directions at 5 degree interval. It provides us 72 feature vectors by counting number of black pixels in each direction. For classification purpose we used Euclidean distance classifier. Use of this feature extraction and classification technique is quite easy but it gives less accuracy. For this we got 26.86% of accuracy for handwritten character.

Index Terms- Optical character recognition, feature extraction, radial histogram, Euclidean distance.
.

1. INTRODUCTION

Gujarati language is official language for Gujarat state. It is an indo-aryan language. It is derived from a language called old Gujarati which is the ancestor language of the modern Gujarati and Rajasthani languages. There are 12 vowels and 34 consonants in Gujarati language.

In this paper we have proposed method for feature extraction and classification of handwritten character recognition.

2. REVIEW WORK

According to paper of [2], in which behavior of different model of neural network used in OCR is described. They have considered different parameters like hidden layer, size of hidden layer and epochs. They have used the back propagation neural network. The implementation of the fully connected back propagation network gave reasonable results toward recognizing characters.

The paper[12], includes a handwritten recognition algorithm based on structural characteristics, histograms and profiles, is presented. The well-known horizontal and vertical histograms are used, in combination with the newly introduced radial histogram, out-in radial and in-out radial profiles for representing 32x32 matrices of characters, as 280-dimension vectors. The recognition process has been supported by a lexical component based on dynamic acyclic FSAs (Finite-State-Automata).

As in [13], worked on Gujarati script document and Hough transform technique is used for the skew detection and correction. In this paper, the technique described a number of sub process that are to be applied while using Hough transform technique for skew detection.

As in [21], concept includes combination 6 different feature extraction techniques and then followed by MLP network Classifier for handwritten digit. The combination of the feature extraction methods is motivated by the observation that different feature extraction algorithms have a better discriminative power for some types of digits.

In paper [27] an OCR for Handwritten Devnagari Characters have presented. Basic symbols are recognized by neural classifier. They have used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features.

3. FEATURE EXTRACTION & CLASSIFICATION

Feature extraction is a heart of system. After pre-processing and segmentation of character from document image features of characters are extracted. In this stage individual characters are represented as unique feature to maximize the recognition rate with least amount of element.

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Classification stage is the main decision making stage of the system and uses the feature vector in previous stage (feature extraction) to identify the text segment according to preset rules.

In this stage the basic task to design a decision rule that would be easy to compute and thus will maximize the certainity of the misclassification relative to the power of the feature extraction scheme employed. The block diagram of entire system is shown in figure 1.

4. PRPOSED METHODOLOGY

In this paper we have concentrated on structural characteristics for feature extraction. For this we have approached a new methodology of radial histogram. Radial histogram is computed from 72 directions at 5 degree interval. Number of black pixels are counted for each of 72 directions and hence getting total 72 unique feature vectors.

After getting 72 feature vectors using Euclidean distance classifier the characters are classified accordingly.

In this Classifier technique first of all the Feature Extraction Technique is applied on the images i.e. trainee and test images and then a feature is considered for the comparison. Now, the distance is calculated between the two same features of two images by subtraction method. This distance is called the Euclidian Distance. The flow of radial histogram technique is shown in figure 2.

Algorithm:

- Apply all pre processing steps.
- Normalize image into 32x32 scale.
- We define the value of the radial histogram *Hr* at an angle φ as the sum of black pixels on a rad that starts from the center of the character matrix (i.e., the element in the 16th row and the 16th column) and ends up at the border of the matrix, forming an angle φ with the horizontal axis. The radial histogram values are calculated with a step of 5 degrees (i.e., 72 features):

$$H_r = \sum_{i=1}^{16} f(\|16 - i \sin\emptyset\|, \|16 + i \cos\emptyset\|),$$

$$\emptyset = 5 * k, k \in [0,72]$$ [12]

- Then we plot radial histogram using these 72 feature vectors which will be used for further processing in recognition.
- After getting feature vectors they are compared to pre-defined feature vectors for each of character using Euclidean distance i.e. subtraction of two images feature vector.
- The image for which we get minimum distance is the correct match for test image character.
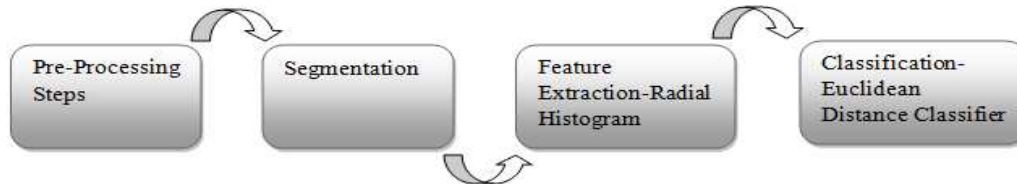

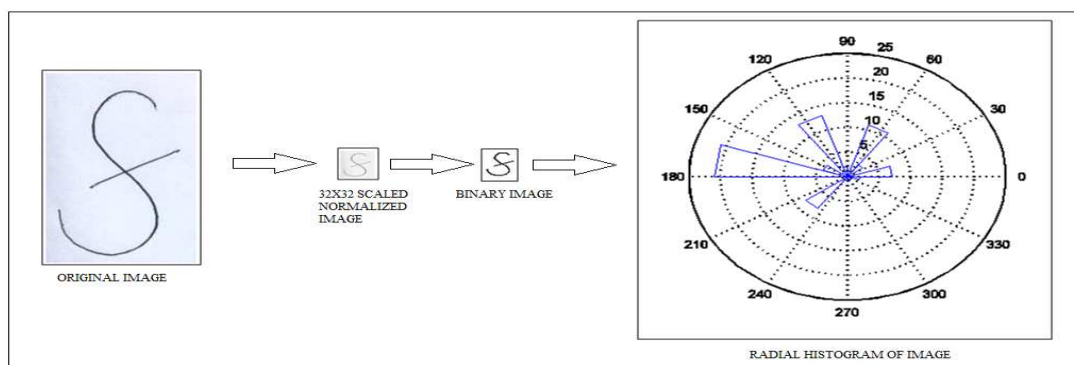Fig. 1. Block Diagram of System
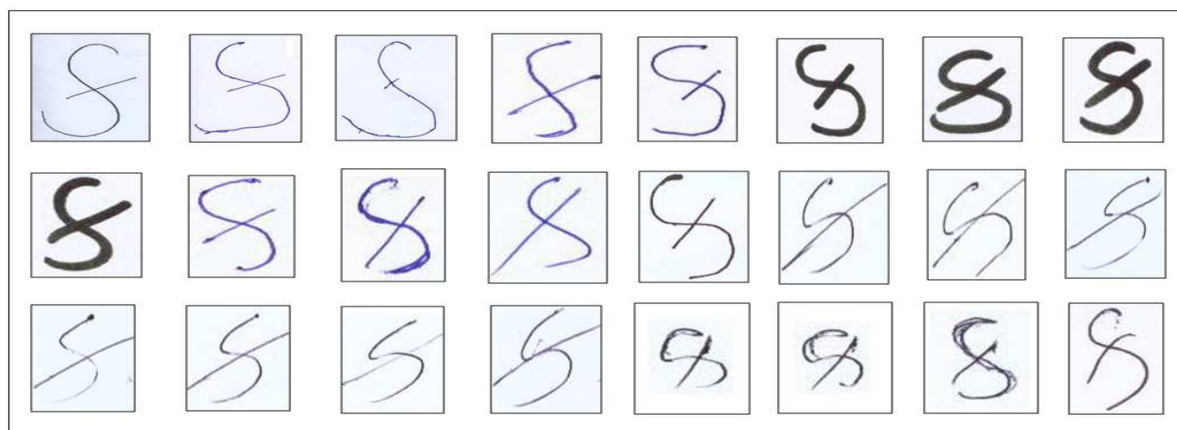

Fig. 2. Flow of Proposed Approach

Fig. 3. Database of Handwritten Character

Table 1. Results for Correctly Recognized Characters.

Characters	True Positive	True Negative	Total Characters	Characters	True Positive	True Negative	Total Characters
Ka	85	235	320	Ya	86	234	320
Kha	90	230	320	La	84	236	50
Ga	124	196	320	Va	56	264	320
Gha	84	236	320	Sha	23	27	50
Ca	70	250	320	Sha	86	234	320
Cha	75	245	320	Sa	75	245	320
Ja	86	234	320	Ha	12	38	50
Jha	69	251	320	Ala	72	248	320
Ta	110	210	320	Ksha	74	246	320
Tha	87	233	320	Gna	62	257	320
Da	82	238	320	A	50	110	160
Ddha	65	255	320	Aa	45	115	160
Ana	15	35	50	E	40	120	160
Ta	79	241	320	Ee	56	104	160
Tha	85	235	320	U	35	125	160
Da	126	194	320	Uu	32	128	160
Dha	76	244	320	Ae	42	118	160
Na	95	225	320	Aei	46	114	160
Pa	72	248	320	O	49	111	160
Pha	73	247	320	Ou	55	105	160
Ba	120	200	320	Am	42	118	160
Bha	109	211	320	Aha	45	115	160
Ma	85	235	320	Total	3149	8571	11720
Total accuracy for correct recognition is 26.86%							

5. RESULTS AND DISCUSSION

This paper includes the feature extraction method for Gujarati handwritten character. Using radial histogram approach for handwritten character we get accuracy of 26.86%.

The database of handwritten characters which shows the variability for same character for different person's handwriting is shown in figure 3 and Table 1 shows the results we get.

6. CONCLUSIONS

In this paper we have proposed a method for feature extraction and classification. Using radial histogram we find feature vectors and using this feature vectors and Euclidean distance classifier the characters are classified.

This method is easy to implement and in use but it gives less accuracy. For the similar looking characters as well variation in handwriting for person to person system fails to recognize character correctly, hence reducing the accuracy of system. To get more accurate result we can add more feature vectors by combining more methods.

Centre for providing us the huge Gujarati Handwritten character database.

REFERENCES

[1] A Sehad(2007a), L Mezai(2007b), M T Laskri(2007c), M Cheriet(2007d) , "Skew angle estimation of printed document using linear regression, wavelet transform and anisotropic diffusion" , Signal Processing and Its Applications.

[2] Chirag I Patel(2011a), Ripal Patel(2011b), Palak Patel (2011c),"Handwritten Character Recognition using Neural Network", International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011.

[3] Chirag Patel, Ripal Patel and Palak Patel ."Object Detection and Segmentation using Local and Global Property", International Journal of Emerging Technologies in Sciences and Engineering (IJETSE) Vol. 02, Issue. 02, pp 02-10 ISSN 1923-9181.

[4] Chirag. I. Patel(2012a) and Ripal Patel(2012b), "Contour Based Object Tracking," International Journal of Computer and Electrical Engineering vol. 4, no. 4, pp. 525-528, 2012.

[5] Chirag. I. Patel(2012a) and Ripal Patel(2012b), "Object Counting in Video Sequences," International Journal of Computer and Electrical Engineering vol. 4, no. 4, pp. 522-524, 2012.

[6] Chirag. I. Patel(2013a) and Ripal Patel(2013b), "Illumination Invariant Moving Object Detection," International Journal of Computer and Electrical Engineering vol. 5, no. 1, pp. 73-75, 2013.

[7] Chirag. I. Patel(2013a) and Ripal Patel(2013b), "Counting Cars in Traffic Using Cascade Haar with KLP," International Journal of Computer and Electrical Engineering vol. 5, no. 4, pp. 435-437, 2013.

[8] Chirag. I. Patel(2011a) and Ripal Patel(2011b), "Gaussian mixture model based Moving object detection from video sequence", in Proceedings of the ICWET 2011(Feb).

[9] Chirag. I. Patel(2011a) and Ripal Patel(2011b)and Alefiya Nalawala(2011c). Finding Object categories in Image collections. in Proceedings of the International Conference on Signal, Systems and Automation ICSSA 2011(JAN).

[10] Chirag. I. Patel(2013a) and Ripal Patel(2013b), "Robust Face Recognition Using Distance Matrice," International Journal of Computer and Electrical Engineering vol. 5, no. 4, pp. 401-404, 2013.

[11] Chirag I Patel(2012a) and Sanjay Garg(2012b), 2012, Article: Robust Face Detection using Fusion of Haar and Daubechies Orthogonal Wavelet Template. International Journal of Computer Applications 46(6):38-44, May 2012.

[12] E. Kavallieratou()2003a, K. Sgarbas(2003b), N. Fakotakis(2003c) and G. Kokkinakis(2003d), "Handwritten Word Recognition based on Structural Characteristics and Lexical Support", Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)

[13] Jay Maniar(2014a), Shreyal Patel(2014b), Lipi Shah(2014c), Ripal Patel(2014d), "Rotation Estimation of Gujarati Script Document using Hough Transform", Int. Journal of Engineering Research and Application ISSN : 2248-9622, Vol. 4, Issue 1(Version 3), January 2014, pp.33-36.

[14] Jayashree Prasad, "Data set of handwritten Gujraticharacters", http://tdildc.in/index.php?optio=com_updownload&task=viewdownloadtool&view=download&toolid=971&lang=en.

[15] Jayashree Prasad, Uday Kulkarni, 'Gujarati Character Recognition using weighted k-NN with mean chi square Distance Measure', International Journal of Machine Learning and Cybernetics ISSN 1868-8071, DOI 10.1007/s13042-013-0187-z.

[16] Jim Higgins(2005), Ed.D. "Introduction Linear Regression", 2005

[17] Lipi Shah(2014a), Ripal Patel(2014b), Shreyal Patel(2014c), Jay Maniar(2014d), "Skew Detection and Correction for Gujarati Printed and Handwritten Character using Linear Regression", International Journal of Advanced Research in Computer Science and Software Engineering 4(1), January - 2014, pp. 642-648.

[18] Nimesh Jivani(2013a)., Jignesh Padasumbiya(2013b), Thakkar, A, Patel(2013c), C. I.,. (2013, November). A new hybrid method for face recognition In Proceedings of (NUiCONE), 2013 (pp.1-5).

[19] P.Shivakumara(2003a), G. Hemantha Kumar(2003b), D. S Guru(2003c), P. Nagabhushan(2003d), "Skew Estimation of Binary Document Images Using Static and Dynamic Thresholds Useful for Document Image Mosaicing", National Workshop on IT Services and Applications (WITSA2003) Feb 27-28, 2003

[20] P.Shivakumara(2005a), G. Hemantha Kumar(2005b), D. S Guru(2005c), P. Nagabhushan(2005d), "A noval technique for estimation of skew in binary text documents image test based on linear regression analysis" , Sadhana Vol. 30, Part 1, February 2005

[21] Rafael M. O. Cruz(2010a), George D. C. Cavalcanti(2010b) and Tsang Ing Ren(2010c), "Handwritten Digit Recognition Using Multiple Feature Extraction Techniques and Classifier Ensemble", IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing

[22] Ray Smith(1994), „A Simple and Efficient Skew Detection Algorithm via Text Row Algorithm‟, December, 1994

[23] Ripal Patel(2011a) and Chirag I Patel(2011b) "Goal Detection from unsupervised Video Surveillance‟ in proceeding of ACITY-2011, Chennai, India, July 2011 published I Springer (LNCS) in communications in computer and information science (CCIS) series. ISSN:1865-0929

[24] Ripal Patel(2012a),Chirag I Patel(2012b) and Ankit Thakkar(2012c),‟ Aggregate Features Approach for Texture Analysis‟, in Proceedings of the Nuicone 2012, Ahmedabad.

[25] Ripal Patel(2012a), Nidhi Rathod(2012b) and Ami Shah(2012c), Comparative Analysis of Face Recognition Approaches: A Survey. International Journal of Computer Applications 57(17):50-69.

[26] Ripal Patel(2014a) , Nidhi Rathod (2014b), Ami Shah(2014c), Mayur Sevak(2014d), „Face Recognition using Eye Distance and PCA Approaches‟, International Journal of Computer Science and Information Technologies,vol 5,issue 1, January 2014

[27] Sandhya Arora(2008a), Debotosh Bhattacharjee(2008b), Mita Nasipuri(2008c), Dipak Kumar Basu(2008d), Mahantapas Kundu(2008e). "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", 2008 IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10.