International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 11, Issue 03 (March 2015), PP.16-20

A Novel Framework For Numerical Character Recognition With Zoning Distance Feature Extraction Approach

T V K P Prasad[1] , T Srinivasa Rao[2], V.Dilip Kumar[3]
[1-3]*Assistant Professor, Department of CSE, SRKR Engineering College, Bhimavaram, West Godavari, AP, INDIA.*

Abstract:- Advancements of Computer technology has made every organization to implement the automatic processing systems for its activities. One of the examples is the recognition of handwritten characters, which has always been a challenging task in image processing and pattern recognition. In this paper we propose Zone based features for recognition of the handwritten characters. In this zoning approach a digit image is divided into 8x8 zones and centre pixel is computed for each zone. This procedure is sequentially repeated for entire zone. Finally features are extracted for classification and recognition.

Keywords:- *Image processing, multi-zoning, pattern recognition, handwritten characters;*

I. INTRODUCTION

Character Recognition (CR) is a branch of pattern recognition, and also a branch of computer vision. CR has been extensively researched for more than four decades. With the advent of digital computers, many researchers and engineers have been engaged in this interesting topic. It is not only a newly developing topic due to man potential applications, such as bank check processing, postal mail sorting, automatic reading of tax forms and various handwritten and printed materials, but it is also a benchmark for testing and verifying new pattern recognition theories and algorithms. In recent years, many new classifiers and feature extraction algorithms have been proposed and tested on various CR databases and these techniques have been used in wide applications [1-4].

Numerous scientific papers and inventions in CR have been reported in the literature. It can be said that CR is one of the most important and active research fields in pattern recognition. Today, CR research is addressing a diversified number of sophisticated problems. Important research in CR includes degraded (heavy noise) omni-font text recognition, and analysis/recognition of complex documents (including texts, images, charts, tables and video documents). Handwritten numeral recognition, (as there are varieties of handwriting styles depending on an applicant's age, gender, education, ethnic background, etc., as well as the writer's mood while writing), is a relatively difficult research field in CR[1-4].

The focus of this paper is the recognition and verification of unconstrained handwritten numerals, which is a challenging research project as these numerals are written without any constraints, (e.g., they are not all written in separate boxes, nor all written neatly, nor all using a specific type of pen). In addition, as mentioned before, unconstrained handwritten numerals have varieties of writing styles due to different backgrounds of the writers. Technically speaking, OCR systems pursue a high recognition rate while seeking the highest reliability, which makes it practical for recognizing unconstrained handwritten numerals.

II. LITERATURE

Sandhya Arora et.al [5]present an OCR for Handwritten Devnagari Characters. Basic symbols are recognized by neural classifier. We have used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features. Shadow features are computed globally for character image while intersection features, chain code histogram features and line fitting features are computed by dividing the character image into different segments. Weighted majority voting technique is used for combining the classification decision obtained from four Multi Layer Perceptron(MLP) based classifier. On experimentation with a dataset of 4900 samples the overall recognition rate observed is 92.80% as we considered top five choices results. This method is compared with other recent methods for Handwritten Devnagari Character Recognition and it has been observed that this approach has better success rate than other methods

E.Kavallieratou et.al [6] presents a handwritten recognition algorithm based on structural characteristics, histograms and profiles. The wellknown horizontal and vertical histograms are used, in combination with the newly introduced radial histogram, outin radial and inout radial profiles for representing 32x32 matrices of characters, as 280dimension vectors. The proposed technique has been developed as the last

module of an integrated document analysis system, The recognition process has been supported by a lexical component based on dynamic acyclic FSAs (Finite-State-Automata).

Luiz S. Oliveira et.al [7] proposes a modular system to recognize handwritten numerical strings. It uses a segmentation-based recognition approach and a Recognition and Verification strategy. The approach combines the outputs from different levels such as segmentation, recognition, and postprocessing in a probabilistic model. A new verification scheme which contains two verifiers to deal with the problems of oversegmentation and undersegmentation is presented. A new feature set is also introduced to feed the oversegmentation verifier. A postprocessor based on a deterministic automaton is used and the global decision module makes an accept/reject decision. Finally, experimental results on two databases are presented: numerical amounts on Brazilian bank checks and NIST SD19. The latter aims at validating the concept of modular system and showing the robustness of the system using a well-known database.

Rafael M. O. Cruz et.al[8] proposed a handwritten digit recognition system which uses multiple feature extraction methods and classifier ensemble. 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. Six features sets were extracted, two proposed by the authors and four published in previous works. It is shown that combining these feature sets is sufficient to achieve high recognition rates. Several combination schemes were tested, showing good results. A scheme using neural networks as a combiner achieved a recognition rate of 99.68%, the highest one on the MNIST database.

Dharamveer Sharma et.al [9] propose a recognition method which is able to account for a variety of distortions due to eccentric handwriting. The recognition of handwritten numerals is a challenging task in the field of image processing and pattern recognition. It can be considered as one of the benchmarks in evaluating feature extraction methods and the performance of classifiers. The performance of character recognition system depends heavily on what kind of features are being used. The objective of this paper is to provide efficient and reliable techniques for recognition of handwritten numerals. In this paper we propose Zoning based feature extraction system which calculates the densities of object pixels in each zone. Firstly the whole image is divided into 4 4 zones. Further in order to gain more accuracy these zones are divided into 6 6 zones. The division of zones carried out up to 8 8 zones. Hence 116 features are extracted in all. Nearest neighbour classifier is used for subsequent classification and recognition purpose.

Le Cun, Y et.al [10] described two novel methods for achieving handwritten digit recognition. The first method is based on a neural network chip that performs line thinning and feature extraction using local template matching. The second method is implemented on a digital signal processor and makes extensive use of constrained automatic learning. Experimental results obtained using isolated handwritten digits taken from postal zip codes, a rather difficult data set, are reported and discussed.

Khofanzad, A. ; and Chung, C. [11] proposed new classification scheme for handwritten digit recognition. The method is based on combining the decisions of two multilayer perceptron (MLP) artificial neural network classifiers operating on two different feature types. The first feature set is defined on the pseudo Zernike moments of the image whereas the second feature type is derived from the shadow code of the image using a newly defined projection mask. A MLP network is employed to perform the combination task. The performance is tested on a data base of 15000 samples and the advantage of the combination approach is demonstrated.

Azeem, S.A. et.al [12] fills a void in the literature of online Arabic handwritten digits recognition as no systems are dedicated to this problem. The two main contributions of this paper are introducing a large online Arabic handwritten digits dataset and developing an efficient online Arabic handwritten digits recognition system. In the dataset, we collected 30,000 online Arabic digits from 300 writers. The developed system uses a combination of temporal and spatial features to recognize those digits. The system achieved 98.73% recognition rate. Comparison with a commercial product demonstrates the superiority of the proposed system.

Jain, A.K.et.al [13] investigate the application of deformable templates to recognition of handprinted digits. Two characters are matched by deforming the contour of one to fit the edge strengths of the other, and a dissimilarity measure is derived from the amount of deformation needed, the goodness of fit of the edges, and the interior overlap between the deformed shapes. Classification using the minimum dissimilarity results in recognition rates up to 99.25 percent on a 2,000 character subset of NIST Special Database 1. Additional experiments on an independent test data were done to demonstrate the robustness of this method. Multidimensional scaling is also applied to the 2,000×2,000 proximity matrix, using the dissimilarity measure as a distance, to embed the patterns as points in low-dimensional spaces. A nearest neighbor classifier is applied to the resulting pattern matrices. The classification accuracies obtained in the derived feature space demonstrate that there does exist a good low-dimensional representation space. Methods to reduce the computational requirements, the primary limiting factor of this method, are discussed.

Castellano, G et.al [14] developed a system for handwritten digits recognition using the Hough transform and a neural network. The input to the system consists of a 128×128 image containing one

handwritten digit. This input is processed using the Hough transform and fed into the neural network, which in turn performs the recognition task. In order to decrease the size of the input vector to the neural net and still preserve most of the information contained in the Hough space, this is sampled in a nonuniform way. It is also made translation and scale independent. An 80% mean recognition rate was obtained using a Kohonen's self organized feature map testing 720 samples of digits written by 18 different persons.

Rajashekararadhya, S.V et.al [15] presents Zone and Distance metric based feature extraction system. The character centroid is computed and the image is further divided in to n equal zones. Average distance from the character centroid to the each pixel present in the zone is computed. This procedure is repeated for all the zones present in the numeral image. Finally n such features are extracted for classification and recognition. Feed forward back propagation neural network is designed for subsequent classification and recognition purpose. We obtained 98 % and 96 % recognition rate for Kannada and Telugu numerals respectively.

Teow and Loe [16] proposed a handwritten digit recognition system based on a biological vision model. The features were empirically extracted by the model, which could linearly separate over a large training set (MNIST). The high recognition rate was reported, where the error rate was 0.59%.

Samit Kumar Pradhan et.al [17] shows a character recognition mechanism based on a syntactic PR approach that uses the trie data structure for efficient recognition. It uses approximate matching of the string for classification. During the preprocessing an input character image is transformed into a skeletonized image and discrete curves are found using a 3 x 3 pixel region. A trie, which we call as a sequence trie is used for a look up approach at a lower level to encode a discrete curve pattern of pixels. The sequence of such discrete curves from the input pattern is looked up in the sequence trie. The encoding of several such sequence numbers for the thinned character constructs a pattern string. Approximate string matching is used to compare the encoded pattern string from a template character with the pattern string obtained from the input character. We consider the approximate matching of the string instead of the exact matching to make the approach robust in the presence of noise. Another trie data structure (called pattern trie) is used for the efficient storage and retrieval for approximate matching of the string. We make use of the trie since it takes O(m) in worst case where m is the length of the longest string in the trie. For the approximate string matching we use look ahead with a branch and bound scheme in the trie. Here we apply our method on 43 Telugu characters from the basic Telugu characters for demonstration. The proposed approach has recognized all the test characters given here correctly; however more extensive testing on realistic data is required.

III. SYSTEM ARCHITECUTE AND ALGORITHM

For feature extraction we use the zone-based approach which is shown in Fig 1.This zone-based feature extraction method will provide good results even when certain pre-processing steps like smoothing, clean-smoothing, filtering and slant removing are not considered. This section explains the concept of the proposed feature extraction method used for extracting features for efficient classification and recognition. In first attempt we selects the input image, and it is normalized, i.e dividing into 8x8 blocks as shown in Fig 1. In the second attempt, we can compute the centre character pixel is and the character/numeral image is further divided into equal zones. In the third step we can measure the average distance from the character centre pixel to the pixels present in the zone column. This procedure repeated to all the columns and rows in the zones.

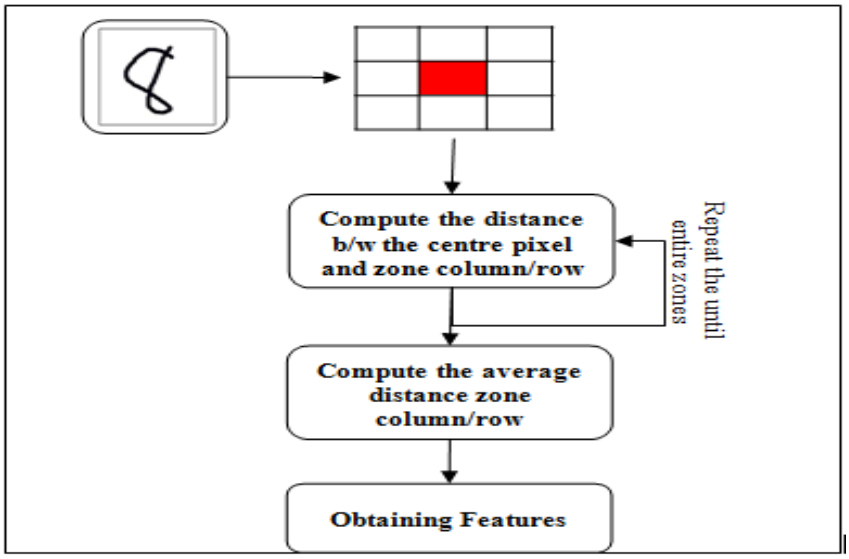

Fig 1 System architecture

Input: Input image **Output: Recognition**
1. Considers the input image and converting it into 8x8 blocks
2. Compute the input image centre pixel of each block and it is called a zone
3. Compute the distance b/w the centre pixel and zone column
4. Compute the distance b/w the centre pixel and zone row
5. Repeat the step 3 & 4 until entire zones
6. Measure the average distance of the zone column
7. Measure the average distance of the zone row
8. Finally such features will be obtained for recognition

IV. RESULTS

MNIST dataset is used for handwritten numeric characters are used to test and training. The data-set consists of 60,000 handwritten numeric characters for training which contains approximately 6,000 images of each class (0-9). The proposed zoning features are extracted and tested using cross Validation. The results for the proposed zoning approach classifier are shown in Fig 2 and statistics are shown in Table 1

Selection process	Validation	Train and Test
Zoning approach	96.12	96.54

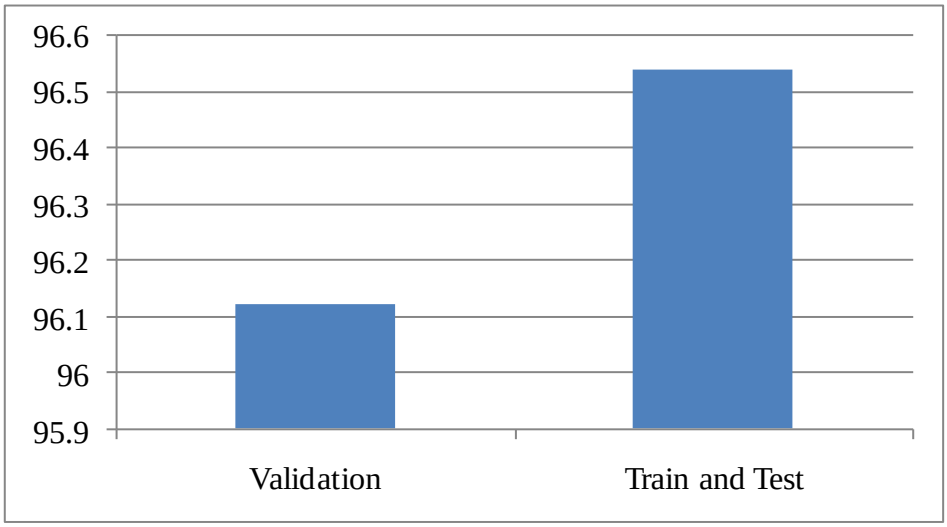

Fig 2 Results

V. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a framework for numerical character recognition with zoning distance feature extraction approach. In this work we have considered 6000 images from the MNIST database. Extensive experiments on different numeral databases and standard database like the, show that the proposed zone-based feature extraction system is robust. We have obtained a maximum recognition rate of 98.65 % for handwritten numerals using the proposed approach. In addition, we plan to extend our work to other Indian scripts like Telugu.

REFERENCES

[1]. A.F.R.Rahman, M.C.Fairhurst, "Recognition of handwritten Bengali Characters: A Novel Multistage Approach", Pattern Recognition, 35,997-1006, 2002.

[2]. R.Chandrashekaran, M.Chandrasekaran, Gift Siromaney, "Computer Recognition of Tamil, Malayalam and Devanagari characters", Journal of IETE, Vol.30, No.6, 1984.

[3]. U.Pal, N.Sharma, F.Kimura, "Recognition of Handwritten Kannada Numerals", 9th International Conference on Information Technology (ICIT'06), ICIT, pp. 133-136.

[4]. Dinesh Acharya U, N VSubba Reddy and Krishnamurthy, "Isolated handwritten Kannada numeral recognition using structural feature and K-means cluster", IISN-2007, pp-125 -129.

[5]. Arora, Sandhya, Debotosh Bhattacharjee, Mita Nasipuri, Dipak Kumar Basu, and Mahantapas Kundu. "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character

Recognition." In Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on, pp. 1-6. IEEE, 2008.

[6]. Kavallieratou, Ergina, K. Sgarbas, Nikos Fakotakis, and G. Kokkinakis. "Handwritten word recognition based on structural characteristics and lexical support." In Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on, pp. 562-566. IEEE, 2003.

[7]. Oliveira, Lui, Robert Sabourin, Flávio Bortolozzi, and Ching Y. Suen. "Automatic recognition of handwritten numerical strings: A recognition and verification strategy." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, no. 11 (2002): 1438-1454.

[8]. Cruz, Rafael MO, George DC Cavalcanti, and Tsang Ing Ren. "Handwritten digit recognition using multiple feature extraction techniques and classifier ensemble." In 17th International Conference on Systems, Signals and Image Processing. 2010.

[9]. Sharma, Dharamveer, and Deepika Gupta. "Isolated handwritten digit recognition using adaptive unsupervised incremental learning technique."International Journal of Computer Applications 7, no. 4 (2010): 27-33.

[10]. Le Cun, Y., L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard. "Handwritten Digit Recognition: Applications of Neural Net Chips and Automatic Learning." In Neurocomputing, pp. 303-318. Springer Berlin Heidelberg, 1990.

[11]. Khofanzad, A., and C. Chung. "Handwritten digit recognition using combination of neural network classifiers." In Image Analysis and Interpretation, 1998 IEEE Southwest Symposium on, pp. 168-173. IEEE, 1998.

[12]. Azeema, Sherif Abdel, Maha El Meseerya, and Hany Ahmeda. "On-line Arabic Handwritten Digits Recognition." (2012).

[13]. Jain, Anil K., and Douglas Zongker. "Representation and recognition of handwritten digits using deformable templates." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19, no. 12 (1997): 1386-1390.

[14]. Castellano, Gabriela, and Mark B. Sandler. "Handwritten digits recognition using Hough transform and neural networks." In Circuits and Systems, 1996. ISCAS'96., Connecting the World., 1996 IEEE International Symposium on, vol. 3, pp. 313-316. IEEE, 1996.

[15]. Rajashekararadhya, S. V., and P. Vanaja Ranjan. "Neural network based handwritten numeral recognition of Kannada and Telugu scripts." In TENCON 2008-2008 IEEE Region 10 Conference, pp. 1-5. IEEE, 2008.

[16]. L.N. Teow and K. F. Loe, Robust Vision-Based Feature and Classification Schemes for Off-Line Handwritten Digit Recognition, Pattern Recognition, Vol. 35, No.1, 2002, pp. 2355-2364.

[17]. Pradhan, Samit Kumar, and Atul Negi. "A syntactic PR approach to Telugu handwritten character recognition." In Proceeding of the workshop on Document Analysis and Recognition, pp. 147-153. ACM, 2012.