RULE BASED RECOGNITION OF PRINTED TELUGU NUMERALS

CH. N. MANISHA AND Y.K. SUNDARA KRISHNA
Krishna University, Machilipatnam, Andhra Pradesh, India
ch.n.manisha@gmail.com; yksk2010@gmail.com

E. SREENIVASA REDDY
Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
esreddy67@gmail.com

ABSTRACT: There are many people in India who speak Telugu language widely. It is an official language of Andhra Pradesh and Telangana States. In this paper, proposing some rules for recognizing printed Telugu numerals. The method recognized Telugu numerals by zoning method, end points method and cavity method. The rules used 2×2 zoning method, number of end points, position of endpoints in the zones and to identify whether the numeral contains cavity or not. This method more efficiently and easily recognized the printed numerals with these simple rules.

KEYWORDS: Telugu Numeral, Endpoint, Cavity, Zones, Rules

INTRODUCTION

Telugu language is an official language of Andhra Pradesh, Telangana, Andaman and Nicobar and Yanam [18]. It is one of the languages derived from Dravidian language[18]. The Telugu numerals act as a major role in Telugu language. So APSRTC buses are using Telugu numerals for the license plates. Most of the Telugu scripts contain the numerals in Telugu language only. The ten numerals that are from zero to nine numerals are shown in Table 1.

RELATED WORK

S.V. Rajashekararadya et. al. [1] proposed the zone based feature extraction algorithm for recognition of handwritten Kannada, Telugu, Tamil and Malayalam numerals. They used zone centroid and image centroids based distance metric feature extraction system. For the classification and recognition, they used the nearest neighbour and feed forward, back propagation neural network classifiers. They used equal zones. Finally 2n features they extracted for classification and recognition.

The printed Telugu numerals were recognized by U. Ravi Babu et. al. [2]. They used structural, skeleton and water reservoir features. In this, they used number of end points, Number of contours, Number of sides of the water reservoir of the image and the ratio of the length of the top line and bottom line are used as recognition methods for recognizing the Telugu numerals.

Kannada and English languages based numerals recognized in [3]. They used a bilingual approach to recognize the documents that contain the numerals of both languages. They divide the image into 64 zones. The pixel den-sity of each zone they prepared 64 features for recognizing the numerals.

Table 1. Telugu numerals

0	1	2	3	4	5	6	7	8	9
౦	౧	౨	౩	౪	౫	౬	౭	౮	౯

M. I. Razzak et. al. [4] described fuzzy rule based, HMM and Hybrid approaches for the recognition of Urdu and Arabic numerals from online and offline domains. In fuzzy rule based approach they provide rules based on structural features. In HMM based approach they built HMMs for each numeral. A similar conflict occurs between Indian numerals and old Arabic numerals. This conflict was resolved by them using hybrid approach which is the combination of fuzzy rules and HMM.

A zone-based feature extraction algorithm scheme for the recognition of off-line handwritten numerals of Kannada, Tamil, Telugu and Malayalam languages is described at [5]. In this the numeral image centroid is calculated and then the image is resized into 50×50 and this is divided into 25 equal zones. The average distance from image centroid in pixels present in the zone column is calculated. Similarly, they calculated row wise pixel distances. They extracted 500 features for classification and recognition. They used the nearest neighbour, feed forward, back propagation neural network and support vector machine classifiers are used for classification and recognition purposes.

G.G. Rajput et. al. [6] identified Printed and Handwritten Kannada numerals. They used crack codes and Fourier descriptors on 40×40 normalized binary image. They extracted feature vector of length 12. Crack Code they represent the line between objects and background pixels. The code is represented in the complex plane and 10 dimensional Fourier descriptors are computed and used as features. Each feature vector length is 12. For classification they used SVM and for an experiment they used 5-fold cross validation method.

G.G. Rajput in [7] proposed another method also to the identification system to print and handwritten Kannada numerals. They normalized to an image size is 40×40 pixels. They traced the chain code from the boundary of the numeral. These codes they represented by complex plane. They prepared feature vectors from 10 dimen-sional Fourier descriptors. They used an SVM classifier to recognize the numerals.

Telephone numbers on signboards recognized by Yamaguchi. T, et. al. [8] . They extracted candidate regions of digits from the image by using edge extraction, enhancement and labeling. For skew correction they used Hough trans-form and to slant correction they used the method of circumscribing digits with tilted rectangles.

Amin T. Alquadah et. al [9] stated English printed numerals based on shift and scale invariant. They used correlation factor between reference image and the testing image. Which is a correlation factor of reference image is achieved highest correlation factor that reference image digit is recognized as corresponding numeric.

Anilkumar N. Holambe et. Al [10] developed a method for recognition of printed and handwritten of Devanagari characters and numerals. They used Sobel and Robert operators for extracting Gradient feature of the devanagrai script. They compute gradient in 8,12,16,32 directions to extract the feature vectors. For classification they used directional vector separately.

Dhandra B.V. et. al. [11] recognized handwritten Kannada and Telugu digits based on zone features. They removed the noise with the help of morphological opening operation. They normalized the image into 32×32 size. The normalized image is divided into 64 zones. They computed each zone pixel density as a feature. They used KNN and SVM classifiers to recognize Kannada and Telugu numerals. This method is free from thinning and slant of digits.

Handwritten numeral recognition system of Devanagari script developed by Mahesh Jangid Kartar Singh et.al. [12]. For feature extraction they recursively subdivided the images into a balanced number of foreground pixels. They got highest accuracy with SVM classifier.

A method for Recognition of English handwritten numerals described in [13]. They used heuristic based and stroke based approaches are used at pre-classification. They extracted the features by sector based approach. In recognition of numerals they used back propagation neural network.

PROPOSED SYSTEM

The block diagram of proposed system shown in Figure 1. First we preprocess the Printed Telugu numeral document image. In preprocess we should apply skewing, binarization, thining slicing methods, identified endpoints and cavities. Preprocess method is explained clearly in section **Data Preprocessing**. In Section **Proposed Rules** we explain data preprocessing, proposed rules and how to apply these twelve rules to Telugu printed numeral document image.

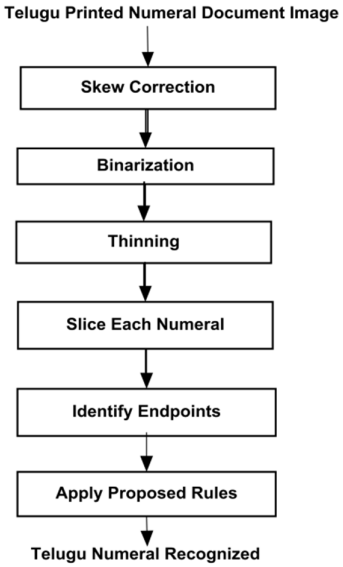

Figure 1. Block Diagram of Proposed System

Data Preprocessing

It is to propose twelve rules to recognize printed Telugu numeral images. It is very important a much necessary to preprocess the images, before applying these twelve rules.

In the preprocess the first step is skewing method. To the best of our knowledge, it is not necessary to apply this method to each and every individual numeral, because the proposed twelve rules are useful only to the printed numerals. So apply skewing correction method only to the entire document image. Next step is convert Numeral image into binary image by applying Otsu method[15]. After apply Thinning Method. Each numeral in Numeral image should be taken separetly as an individual image. For this segementation purpose horizontal ON pixels, vertical ON pixels are available. OFF pixels consider as space for each numeral in image. Next step identify end points to numerals. For the purpose of identification of endpoints and thinning method , predefined functions are available in some programming languages.

In this system it is not necessary to normalize the size of the numeral image. Figure 2 Shows the identification of end points in numeral image.

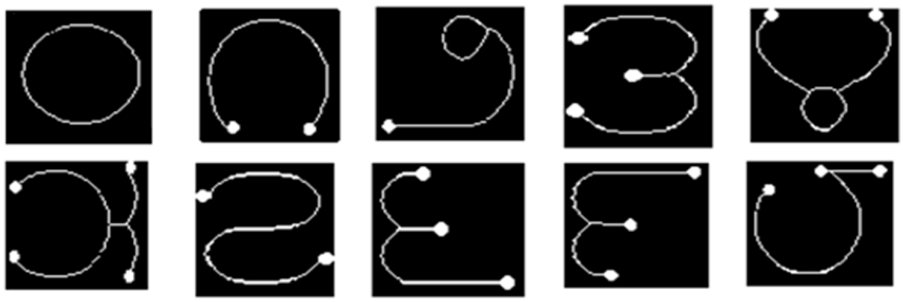

Figure 2. Identify Endpoints

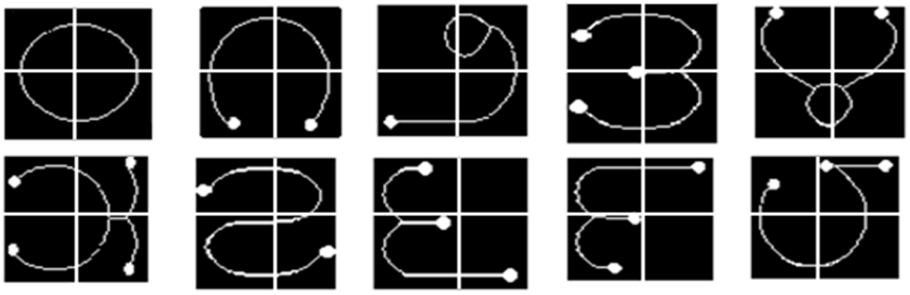

Figure 3. Identified Endpoints in Zones

Proposed Rules

After preprocessing, we have divided each numeral image into 2×2 zones. End points have already identified in preprocessing steps. It is equally important to identify the zones which contain endpoints and which donot contain endpoints. Figure 3 shows the above particulars.

It is necessary to check whether there are cavities or not in numeral images. But it is not necessary to count the total number of cavities.

These rules depend upon number of endpoints. Through there are endpoints to many rules, they depend on zones that do and donot contain endpoints. If there is conflict with numerals, then identify whether there is cavity in the numeral image or not. But it is not necessary to count the total number of cavities.

Let N is Number of End Points, C is Cavities Found or not. Z_1, Z_2, Z_3, Z_4 are First Zone, Second Zone, Third Zone and Fourth Zone respectively. Below twelve rules shows the recognition of Telugu printed numeral images.

R_1: **If** N is 0 **then** ZERO

R_2: **If** N is 1 **then** TWO

R_3: **If** N is 2 and [Z_1, Z_2] and C is true **then** FOUR

R_4: **If** N is 2 and [Z_1, Z_2] and C is False **then** EIGHT

R_5: **If** N is 2 and [Z_1, Z_4] **then** SEVEN

R_6: **If** N is 2 and [Z_3, Z_4] **then** ONE

R_7: **If** N is 3 and [Z_1, ~Z_2, Z_3, ~Z_4] **then** THREE

R_8: **If** N is 3 and [Z_1, Z_2, ~Z_3, ~Z_4] **then** EIGHT

R_9: **If** N is 3 and [Z_1, Z_4] and C is true **then** FIVE

R_{10}: **If** N is 3 and [Z_1, Z_4] and C is false **then** SIX

R_{11}: **If** N is 3 and [Z_2, Z_3] **then** NINE

R_{12}: **If** ~R_1 and ~R_2 and … and ~R_{11} **then** FIVE

To recognize ZERO and TWO, it is enough to identify only total number of endpoints. From these two rules zones and cavities identification is not necessary.

Total number of endpoints is the same to ONE, FOUR, SEVEN and EIGHT. So identify endpoints in different zones. But conflict occurs between FOUR and EIGHT. Therefore apply cavities identifcation method in order to solve this conflict problem. Same process apply for THREE, FIVE, SIX and EIGHT when number of endpoints is three. If all the rules are failed then the result become FIVE.

EXPERIMENTAL RESULTS

In this section the proposed twelve rules experimentally applied on nearly 2400 printed Telugu numerals. These printed Telugu numerals are taken from different types of fonts and different sizes of printed Telugu numerals. We achieved 97.92% accurate results. According to our survival knowledge we understand that number FIVE is printed as two different ways. One type of printing has three endpoints and one cavity and second type of printing has four endpoints and no cavity. Therefore, we consider FIVE as a special numeral. The experimental result shown in Table 2.

Table 2. Experimental Results

Numeral	No. of Tested Samples	No. of Recognized Samples	Recognition Rate
౦	240	240	100%
౧	240	240	100%
౨	240	240	100%
౩	240	210	87.50%
౪	240	240	100%
౫	240	240	100%
౬	240	230	95.83%
౭	240	240	100%
౮	240	230	95.83%
౯	240	240	100%
Overall Recognition			97.92%

COMPARISON OF METHODS

Comparison of different methods for recognition Telugu numerals with this method as shown in Table 3.

Table 3. Comparison of Different Methods

S.No.	Title of the work	Authors	Recognition Process	Accuracy Rate
1.	Handwritten Multiscript Numeral Recognition using Artificial Neural Networks in [21]	Stuti Asthana, Farha Haneef, Rakesh K Bhujade	Multilayer feed-forward back-propagation algorithm using two hidden layer.	96.53%
2.	Kannada, Telugu and Devanagari Handwritten Numeral Recognition with Probabilistic Neural Network: A Script Independent Approach in [22]	B.V.Dhandra, R.G.Benne, Mallikarjun Hangarge	• Directional density estimation • Water reservoirs Feature • Maximum profile distances Feature • Fill hole density Feature • Probabilistic neural network classification	97.20%
3.	An Approach for Telugu Numeral Recognition by Moment Invariants in Wavelet Transform Domain in [20]	M. Radhika Mani, R. Kavitha Lakshmi	• Boundary based shape representation scheme. • Haar wavelet transformed sub bands • Boundary Moment Invariants	-

S.No.	Title of the work	Authors	Recognition Process	Accuracy Rate
4.	Tri-scripts handwritten numeral recognition: a novel approach in [19]	Benne R.G., Dhandra B.V., Mallikarjun Hangarge	• The directional density estimation features • water reservoir principle • maximum profile distance features • fill-hole density features • KNN Classification	98.40%
5.	Handwritten Numeral/Mixed numerals recognition of south-Indian scripts: The zonebased feature extraction method in [5]	S. V. Rajashekararadhya, P. Vanaja Ranjan	• Compute distance between image centroid to each pixel as row wise and column wise in each Zone individually • Nearest neighbor classification • Feed-forward back propagation neural network classification • Support vector machine classification	98.60%
6.	Efficient Zone based Feature Extraction algorithm for handwritten numeral recognition of four popular south Indian scripts in [1]	S.V. Rajashekararadhya, , Dr. P. Vanaja Ranjan	• Compute distance between image centroid and each zone. • Compute distance between zone centroid and each pixel in the zone • Nearest neighbour classifier • Feed-forward back propagation neural network classification • Support vector machine classification	99%
7.	Printed Telugu Numeral Recognition based on Structural, Skeleton and Water Reservoir Features in [2]	U. Ravi Babu, Y. V. V. Satyanarayana, S. Marthu Perumal	• Number of contours • Number of endpoints • Number of Reservoirs • Side of Reservoirs • Ratio of the Top line to Bottom line	100%

CONCLUSION

Every time we need not identify the endpoints which do or donot locate in the zones and also not necessary to identify the cavities. According to certain rules we can only identify the position of endpoints and the cavity. For this we greatly reduce the time complexity and space complexity also. However, it is necessary to identify the total number of cavities.

The proposed rules are very simple and we can easily identify the printed Telugu numerals. However, for handwritten numerals, these rules are not enough. It is necessary to add some more process to identify the handwritten Telugu numerals. These rules are used to recognize offline printed Telugu numerals very efficiently. This algorithm is also free from normalized size of the numeral image.

REFERENCES

S.V. Rajashekararadhya, and P. Vanaja Ranjan. "Efficient Zone Based Feature Extraction algorithm for handwritten numeral recognition of four popular south Indian Scripts", *Journal of Theoretical and Applied Information Technology*, Vol. 4 Issue 12, pp. 1171-1181, 2008.

U. Ravi Babu, Y. V. V. Satyanarayana, and S. Marthu Perumal, "Printed Telugu Numeral Recognition based on Structural, Skeleton and Water Reservoir Features", *International Journal of Computers and Technology*, Vol. 10 No 7, pp. 1815-1824, August 2013.

B.V.Dhandra, Gururaj Mukarambi, and Mallikarjun Hangarge, "Kannada and English Numeral Recognition System", *International Journal of Computer Applications*, Vol. 26 No.9, pp. 17-22, July 2011.

M. I. Razzak, S.A.Hussain, A. Belaid, and M. Sher, "Multi-font Numerals Recognition for Urdu Script based Languages", *International Journal of Recent Trends in Engineering*, Vol. 2, No. 3, pp. 70-72, November 2009.

S.V. Rajashekararadhya, and P. Vanaja Ranjan. "Handwritten Numeral/Mixed Numerals Recognition of South-Indian Scripts: The Zone-Based Feature Extraction Method", *Journal of Theoretical & Applied Information Technology*, Vol. 7 No. 1, pp. 63-79, September 2009.

G.G. Rajput, Rajeswari Horakeri and Sidramappa Chandrakant. "Printed and Handwritten Kannada Numeral Recognition using Crack Codes and Fourier Descriptors Plate", *IJCA, Special issue on RTIPPR*, Vol. 1(1), pp. 53-58, 2010.

G.G. Rajput, Rajeswari Horakeri, and Sidramappa Chandrakant, "Printed and Handwritten Mixed Kannada Numerals Recognition using SVM", *International Journal on Computer Science and Engineering.* Vol.2 No.5, pp. 1622-1626, 2010.

Yamaguchi, T.; Nakano, Y.; Maruyama, M.; Miyao, H. & Hananoi, T., "Digit Classification on Signboards for Telephone Number Recognition", *in ICDAR, IEEE Computer Society*, pp. 359-363, 2003.

Amin T. Alqudah, Hussein R. Al-zoubi and Mahmood A. Al-Khassaweneh, "Shift and Scale Invariant Recognition of Printed Numerals", Vol. 21 No. 1, *Abhath Al-Yarmouk: Basic Sci. & Eng.*, pp. 41-49, 2012.

Anilkumar N. Holambe, Ravinder.C.Thool, and S.M.Jagade, "Printed and Handwritten Character & Number Recognition of Devanagari Script using Gradient Features", *International Journal of Computer Applications*, Vol. 2 No.9, pp. 38-41, 2010.

Dhandra B.V., Gururaj Mukarambi, and Mallikarjun Hangarge, "A Script independent approach for handwritten Bilingual Kannada and Telugu digits recognition", *International Journal of Machine Intelligence*, Vol. 3 issue 3, pp. 155-159, 2011.

Mahesh Jangid Kartar Singh, Renu Dhir, and Rajneesh Rani, "Performance Comparison of Devanagari Handwritten Numerals Recognition", *International Journal of Computer Applications*, Vol. 22 No. 1, pp. 1-6, 2011.

M. Hanmandlu, M. Vamsi Krishna, and P. Chaitanya Krishna, "Some Approaches to the recognition of handwritten numerals", *National Conference on Document Analysis and Recognition*, India, 2001.

Seong –Whan Lee. “Off-Line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network”, *IEEE Transactions on Pattern analysis and Machine learning*. Vol. 18 No. 6, pp. 648-652, 1996.

N. Anupama, Ch. Rupa and E. Sreenivasa Reddy, “Character Segmentation for Telugu Image Document using Multiple Histogram Projections”, *Global Journal of Computer Science and Technology Graphics & Vision*, Vol.13 Issue 5, pp. 11-15, 2013.

Alasdair Mcandrew, *Introduction to Digital Image Processing*, India, Cengage learning, 2004.

Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 2nd Ed., Pearson Education, 2004.

Telugu language. (2014, October 15). In *Wikipedia, The Free Encyclopedia*. Retrieved 14:49, October 15, 2014, from http://en.wikipedia.org/w/index.php?title=Telugu_language&oldid=629689576

Benne R.G., Dhandra B.V. and Mallikarjun Hangarge, "Tri-scripts handwritten numeral recognition: a novel approach", *Advances in Computational Research*, Vol.1, Issue 2, pp-47-51, 2009.

M. Radhika Mani, and R. Kavitha Lakshmi, "An Approach for Telugu Numeral Recognition by Moment Invariants in Wavelet Transform Domain", International Journal of Innovative Research in Computer and Communication Engineering, Vol.1, Issue 8, pp. 1676-1682, October 2013.

Stuti Asthana, Farha Haneef, and Rakesh K Bhujade, "Handwritten Multiscript Numeral Recognition using Artificial Neural Networks", *International Journal of Soft Computing and Engineering*, Vol.1, Issue 1, pp. 1-5, March 2011.

B.V.Dhandra, R.G.Benne, and Mallikarjun Hangarge, "Kannada, Telugu and Devanagari Handwritten Numeral Recognition with Probabilistic Neural Network: A Script Independent Approach", International Journal of Computer Applications, Vol. 26 No. 9, pp. 11-16, July 2011.