

# A Study on Recognition Methods of Telugu Numerals and Characters

Ch. N. Manisha<sup>1</sup>, E. Sreenivasa Reddy<sup>2</sup>, Y.K. Sundara Krishna<sup>3</sup>

<sup>1</sup>Research Scholar, Krishna University, Machilipatnam, Andhra Pradesh, India.

<sup>1</sup>ch.n.manisha@gmail.com

<sup>2</sup>Professor, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

<sup>2</sup>esreddy67@gmail.com

<sup>3</sup>Professor, Krishna University, Machilipatnam, Andhra Pradesh, India.

<sup>3</sup>yksk2010@gmail.com

**Abstract--** Recognition of Characters in digital data is a very big challenging issue for the researchers. Recognition of printed and handwritten of Telugu digital documents is a more complicated issue. The digital format of hardcopy of a document is called document image. The handwritten Telugu characters are recognized through offline and online. This paper analyzes the various methods to recognize the numerals and characters of Telugu language.

**Keywords---** Telugu, Characters, Numerals, Printed, Handwritten, Online, Offline, Recognition

## I. INTRODUCTION

Present days for everything the people are depend on electronic gadgets. On the development of new technologies all the data stores in the form of digital format. The data in the different languages are stores in digital format. That is the printed and handwritten data stores in the form of digitally.

For this reason two types of digital document images are available. That is printed characters of digital document images and handwritten characters of digital document images for different types of languages. Printed data are storing in the format of digital document images. The recognition of handwritten data is two types. That is online handwritten recognition and offline handwritten recognition.

Online handwritten recognition is depends on recognition of movement and strokes of the stylus on electronic gadgets. Offline handwritten recognition is recognition of characters in the document images.

There is no proper recognition systems are available for Telugu language. Because of its unconnected and connected components. It has different type of styles because some of the characters are have single component structure and some of the characters have multiple unconnected components structures.

The scanned copy of hard copy of a document or a page of book is called digital document image. Researchers mainly focus on recognition of handwritten characters of various languages. Because Among the printed and handwritten character recognition, handwritten character recognition is very toughest task. Another reason to give the importance to handwritten character recognition is many electronic gadgets not support regional languages.

However either online handwritten recognition and offline recognition data pre-processing is very important step. Because any type of data contains some noise. To proper recognition it is necessary to clean the noise on the digital data. Section-II, Section-III, Section-IV, Section-V and Section-VI describes Data Pre-processing, Telugu numeral recognition, Telugu character recognition and Comparative analysis and conclusion respectively.

## II. DATA PRE-PROCESSING

Before recognizing any object pre-processing the data is a very important task. To recognition of Telugu characters segmentation is some more difficult and important task. The similarity, connected and unconnected components of the Telugu characters segmentation is a challenging issue to the researchers.



Fig. 1: Printed Telugu Numerals



Fig. 2: Online Handwritten Telugu Numerals

The segmentation of printed Telugu characters well explained in [1]. The authors segmented the text line of characters with the help of fringe map. Using Projection Profiles, Connected components, spatial vertical relationships and nearest neighbour method to segmented overlapped text and characters of Telugu language in [2]. In [3] the authors segmented Telugu text into lines, words and words by using multiple histogram projections and morphological operators.

### III. TELUGU NUMERALS RECOGNITION

Telugu numerals play an important role in Telugu language. In traditional documents and license plates of APSRTC buses are using Telugu numbers. Various recognition methods are developing to recognize both print and handwritten of Telugu numerals. Section-A and section-B describes printed and handwritten of Telugu numerals.

#### A. Printed Telugu Numerals Recognition

To recognize printed Telugu numerals different methods implemented. Telugu numerals have a specific structure. Each numeral has connected structure. Based on structural, skeleton and water reservoir methods Telugu numerals are recognized in [4]. Fig.1 shows printed Telugu numerals.

#### B. Handwritten Telugu Numerals Recognition

Various methods were implemented to recognize handwritten of Telugu numerals. Zone based recognition method was implemented for four South-Indian language numerals in [11]. Using moment invariants to recognized

Telugu numerals in [8]. Now researchers are implementing bilingual and multilingual numeral recognition systems. Dhandra B.V. et. al. [13] recognized bilingual numeral recognition system. They recognized Kannada and Telugu numerals. In [10] a method implemented to recognize Kannada, Telugu and Devanagari numerals. Multiscript numeral recognition was implemented in [14]. Fig.2 shows Online handwritten Telugu numerals.

### IV. TELUGU CHARACTER RECOGNITION

To recognize Telugu characters various methods are implementing. Section-A and Section-B describes Printed Telugu character recognition and Handwritten Telugu character recognition respectively.

#### A. Printed Telugu Character recognition

Printed characters have a clear structure. Fig.3 shows printed Telugu characters [23]. C. Vasantha Lakshmi et.al. [5] recognized printed Telugu characters with the help of edge histograms. The authors deal with the recognition of similar characters of printed Telugu characters.

#### B. Handwritten Telugu Character Recognition

In the Recognition process handwritten of Telugu characters is very complicated task. Because it contains large number of characters set. So that collection of writing styles becomes large. The recognition of Telugu characters is two types. That is offline handwritten Telugu Character Recognition and Online handwritten Telugu Character Recognition. Fig.4 shows handwritten Telugu characters.

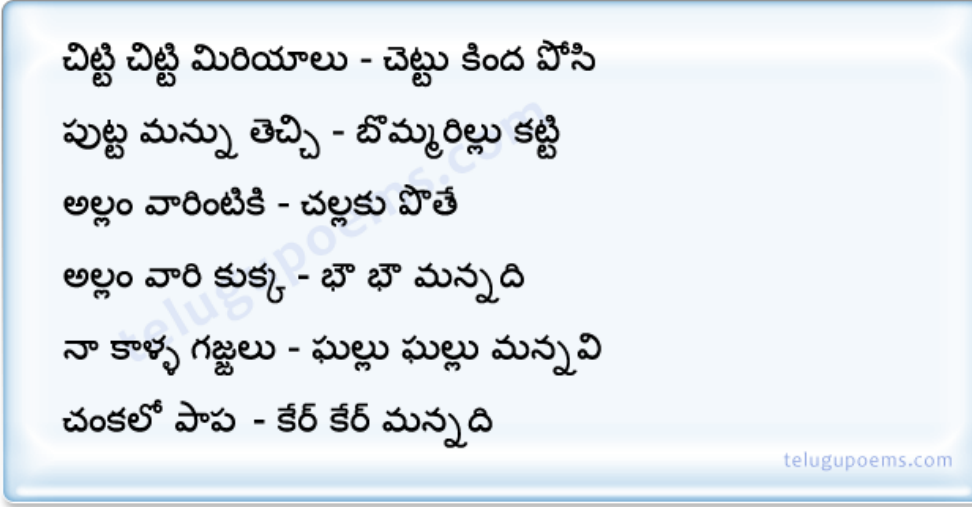


Fig.3: Printed Telugu Characters

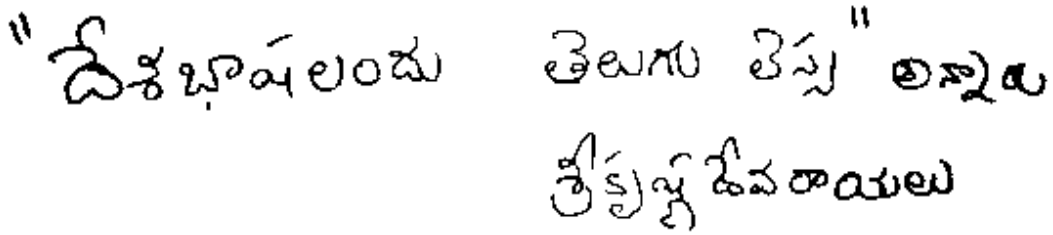


Fig.4: Handwritten Telugu Characters

#### 1) Offline Handwritten Telugu Character Recognition:

Normally printed or handwritten of Telugu documents convert into digital form is called digital document image.

J. Bharathi et.al. [6] segmented touching characters and recognized the characters of Telugu language. P. Pavan Kumar et. al. [7] described how to handle broken characters and segmentation problems in the recognition of Telugu characters. Bilingual recognition method is implemented in [15]. They recognized Hindi and Telugu characters.

Atul Negi et.al [16] recognized Telugu text by identified candidate characters using zoning technique and cavity analysis used to construct structural features. Rinki Singh et.al. [17] extracted height and width of characters, the number of horizontal and vertical lines, number of slope lines and special dots are as features and back-propagation classifier used for classification.

In [18] Multi-Layer Perceptron networks used to recognized Telugu characters. They used different number of hidden nodes and got different accuracy rates.

#### 2) Online Handwritten Telugu Character Recognition:

Online handwritten Telugu characters are recognizing using the movement and strokes of the stylus is also called as digital pen. To increase the utilization of the electronic gadgets online handwritten data also increasing. For this reason online handwritten recognition task becomes very important research challenging task.

Vijay Kumar. K. et.al [19] defines each component in the single character. They identified main stroke, baseline auxiliary, top stroke and bottom stroke are as the four components presented in single Telugu character. By using the strokes of stylus they constructed feature vector and classified with support vector machine. Hidden Markov Model based recognition system was implemented in [20].

Prasanth L. et.al [21] recognized characters based on local features such as Shape Context and Tangent Angle features, Generalized Shape Context feature and the fourth set containing x-y, normalized first and second derivatives and curvature features. For classification they used nearest neighborhood classifier. Rajkumar.J. et. al. [22] implemented

two schemas based on Ternary Search Tree and SVM for recognizing online Telugu Characters.

**V. COMPARATIVE ANALYSIS**

Table.1 and Table.2 provides comparative analysis of Telugu numerals recognition and Telugu characters recognition respectively.

**VI. CONCLUSION**

To recognizing Telugu characters and numerals many recognition systems are implementing. Still handwritten recognition of Telugu language is a challenging task for the researchers. This paper provides the information about recognition of Telugu characters and numerals. This paper also provides comparative analysis on various methods implemented on Telugu characters and numerals. This paper provides information of various recognition methods for Telugu characters and numerals to the researchers.

**TABLE I: COMPARATIVE ANALYSIS OF TELUGU NUMERALS RECOGNITION**

S. No.	Title	Authors	Year	Recognized	Accuracy
1.	Printed Telugu Numeral Recognition based on Structural, Skeleton and Water Reservoir Features in [4].	U. Ravi Babu, Y. V. V. Satyanarayana and S. Marthu Perumal	2013	Printed Telugu numerals	100%
2.	An Approach for Telugu Numeral Recognition by Moment Invariants in Wavelet Transform Domain in [8].	M. Radhika Mani and R. Kavitha Lakshmi	2013	Telugu Numerals	-
S. No.	Title	Authors	Year	Recognized	Accuracy
3.	Kannada, Telugu and Devanagari Handwritten Numeral Recognition with Probabilistic Neural Network: A Script Independent Approach in [9].	B.V.Dhandra, R.G.Benne, and Mallikarjun Hangarge	2011	Handwritten Telugu Numerals	97.20%
4.	Tri-scripts handwritten numeral recognition: a novel approach in [10].	Benne R.G., Dhandra B.V. and Mallikarjun Hangarge	2009	Handwritten Telugu Numerals	98.40%
5.	Efficient Zone Based Feature Extraction algorithm for handwritten numeral recognition of four popular south Indian Scripts in [11].	S.V. Rajashekararadhya, and P. Vanaja Ranjan	2008	Handwritten Telugu Numerals	99%
6.	Handwritten Numeral/Mixed Numerals Recognition of South-Indian Scripts: The Zone-Based Feature Extraction Method in [12].	S.V. Rajashekararadhya, and P. Vanaja Ranjan	2009	Handwritten Telugu Numerals	98.60%
7.	A Script independent approach for handwritten Bilingual Kannada and Telugu digits recognition in [13].	Dhandra B.V., Gururaj Mukarambi, and Mallikarjun Hangarge	2011	Handwritten Telugu Numerals	99.83% 99.80%
8.	Handwritten Multiscript Numeral Recognition using Artificial Neural Networks in [14].	Stuti Asthana, Farha Haneef, and Rakesh K Bhujade	2011	Handwritten Telugu Numerals	96.53%

**TABLE II: COMPARATIVE ANALYSIS OF TELUGU CHARACTERS RECOGNITION**

S. No.	Title	Authors	Year	Recognized	Accuracy
1.	OCR of printed Telugu text with high recognition accuracies in [5].	C. Vasantha Lakshmi, Ritu Jain, and C. Patvardhan	2006	Printed Telugu Characters	98.5%
2.	Improvement of Telugu OCR by	J. Bharathi and P.	2014	Telugu	83%

	segmentation of Touching Characters in [6].	Chandrasekhar Reddy		Characters	
3.	Towards Improving the Accuracy of Telugu OCR Systems in [7].	P. Pavan Kumar, Chakravarthy Bhagvati, Atul Negi, Arun Agarwal, B. L. Deekshatulu	2011	Telugu Characters	-
4.	A bilingual OCR for Hindi-Telugu documents and its applications in [15].	C. V. Jawahar, M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran	2003	Telugu Characters	97.5%
5.	Localization, extraction and recognition of text in Telugu document images in [16].	Atul Negi, K. Nikhil Shanker and Chandra Kanth Chereddi	2003	Telugu Characters	97%-98%
6.	OCR for Telugu Script Using Back-Propagation Based Classifier in [17].	Rinki Singh and Mandeep Kaur	2010	Telugu Characters	-
7.	Improvement in Efficiency of Recognition of Handwritten Telugu Script in [18].	K. Vijay Kumar and R.Rajeshwara Rao	2013	Handwritten Telugu Characters	78.1% 84.9% 81.4%
8.	Online Handwritten Character Recognition for Telugu Language Using Support Vector Machines in [19].	K. Vijay Kumar, R.Rajeshwara Rao	2013	Online handwritten Telugu Characters	96.69%
<b>S. No.</b>	<b>Title</b>	<b>Authors</b>	<b>Year</b>	<b>Recognized</b>	<b>Accuracy</b>
9.	HMM-based Online Handwriting Recognition System for Telugu Symbols in [20].	Jagadeesh Babu. V, Prasanth. L Raghunath Sharma. R, Prabhakara Rao G.V., Bharath. A	2007	Online handwritten Telugu Characters	91.6%, 98.7%
10.	Elastic matching of online handwritten Tamil and Telugu scripts using local features in [21].	Prasanth, L., Jagadeesh Babu, V., Raghunath Sharma, R., Prabhakara Rao, G. V., and Dinesh, M.	2007	Online handwritten Telugu Characters	90.6%
11.	Two schemas for online character recognition of Telugu script based on Support Vector Machines in [22].	Rajkumar.J, Mariraja K., Kanakapriya.K., Nishanthini, S. and Chakravarthy, V.S.	2012	Online handwritten Telugu Characters	90.55% 96.42%

#### REFERENCES

- [1] Vijaya Kumar Koppula and Atul Negi. "Fringe Map Based Text Line Segmentation of Printed Telugu Document Images." In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp. 1294-1298. IEEE, 2011.
- [2] M. Swamy Das, C.R.K. Reddy, A. Govardhan, and G. Saikrishna "Segmentation of Overlapping Text lines, Characters in printed Telugu text document images." *International Journal of Engineering Science and Technology*, 2(11), 6606-6610, 2010.
- [3] 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 and Vision*, Vol. 13 Issue.5, pp. 11-16, 2013.
- [4] 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*, 10(7), 1815-1824, 2013.
- [5] C. Vasantha Lakshmi, Ritu Jain, and C. Patvardhan, "OCR of printed Telugu text with high recognition accuracies." In *Computer Vision, Graphics and Image Processing*, pp. 786-795, Springer Berlin Heidelberg, 2006.
- [6] J. Bharathi and P. Chandrasekhar Reddy, "Improvement of Telugu OCR by segmentation of Touching Characters.", *International Journal of Research in Engineering and Technology*, Vol.3 Issue. 10, 2014
- [7] P. Pavan Kumar, Chakravarthy Bhagvati, Atul Negi, Arun Agarwal, B. L. Deekshatulu, "Towards Improving the Accuracy of Telugu OCR Systems", In *Document Analysis and Recognition (ICDAR), 2011 International Conference*, pp. 910-914. IEEE, 2011.
- [8] 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, 2013.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] C. V. Jawahar, M. N. S. K. Pavan Kumar, S. S. Ravi Kiran, "A bilingual OCR for Hindi-Telugu documents and its applications." In *2013 12th International Conference on Document Analysis and Recognition*, Vol. 1, pp. 408-408, IEEE Computer Society, 2003.
- [16] Atul Negi, K. Nikhil Shanker and Chandra Kanth Chereddi, "Localization, extraction and recognition of text in Telugu document images", In *2013 12th International Conference on Document Analysis and Recognition*, Vol. 2, pp. 1193-1193, IEEE Computer Society, 2003.
- [17] Rinki Singh and Mandeep Kaur, "OCR for Telugu Script Using Back-Propagation Based Classifier", *International Journal of Information Technology and Knowledge Management*, Volume 2, No. 2, pp. 639-643, 2010.
- [18] K. Vijay Kumar and R.Rajeshwara Rao, "Improvement in Efficiency of Recognition of Handwritten Telugu Script", *International Journal of Inventive Engineering and Sciences*, Vol. 2, Issue 1, pp. 1-4, 2013
- [19] K. Vijay Kumar, R.Rajeshwara Rao, "Online Handwritten Character Recognition for Telugu Language Using Support Vector Machines", *International Journal of Engineering and Advanced Technology*, Vol.3, Issue.2, pp. 189-192, 2013.
- [20] Jagadeesh Babu. V, Prasanth. L Raghunath Sharma. R, Prabhakara Rao G.V., Bharath. A, "HMM-based Online Handwriting Recognition System for Telugu Symbols", In *Document Analysis and Recognition, ICDAR, Ninth International Conference*, Vol.1, pp. 63-67, IEEE, 2007.
- [21] Prasanth, L., Jagadeesh Babu, V., Raghunath Sharma, R., Prabhakara Rao, G. V., and Dinesh, M., "Elastic matching of online handwritten Tamil and Telugu scripts using local features." In *Document Analysis and Recognition, ICDAR 2007, Ninth International Conference*, Vol. 2, pp. 1028-1032, IEEE, 2007.
- [22] Rajkumar.J, Mariraja K., Kanakapriya,K., Nishanthini, S. and Chakravarthy, V.S., "Two schemas for online character recognition of Telugu script based on Support Vector Machines." In *Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition*, pp. 565-570. IEEE Computer Society, 2012.
- [23] Telugu poems, (2014), *chitti-chitti-miriyalu.png* [ONLINE]. Available at: [http://telugupoems.com/telugu\\_children\\_rhymes.php](http://telugupoems.com/telugu_children_rhymes.php) [Accessed 24 November 14].