International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 12, Number 2 (2016), pp. 149-157
© Research India Publications
http://www.ripublication.com

Post-Processing for Glyph-Based Offline Handwritten Telugu Characters Recognition System

Ch. N. Manisha[1], E. Sreenivasa Reddy[2] and Y.K. Sundara Krishna[3]

[1]Research Scholar, Krishna University, Machilipatnam, Andhra Pradesh, India.
[2]Professor, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.
[3]Professor, Krishna University, Machilipatnam, Andhra Pradesh, India.
[1]ch.n.manisha@gmail.com
[2]esreddy67@gmail.com
[3]yksk2010@gmail.com

Abstract

Post-processing is the final step of a character recognition system. This method is used to improve the recognition accuracy rate of the character recognition system. In this paper, we proposed a post-processing method for glyph-based recognition of offline handwritten Telugu characters. We identified the characters having confused glyphs. Therefore, we grouped the characters on the basis of their confused glyphs, which were identified at the classification step. For each group, we rectified the incorrectly recognized glyphs of characters to improve the recognition accuracy rate. We got overall success rate is 99.06%

1. INTRODUCTION

Post-processing is the final step of a character recognition system, which is used to improve the recognition accuracy rate. This step occurs after the feature extraction and classification steps. This method deals with the confused glyphs of the offline handwritten Telugu characters. "Telugu is the primary language of Andhra Pradesh and Telangana states" [8]. This method deals with the glyphs of offline handwritten Telugu characters. Glyph is "A graphical representation of either a character, a part of a character, or a sequence of characters." [5]. In this proposed method, glyph is represents as a part of a character.

Post-processing rectifies the recognition errors that occur during the classification step. Glyphs can be classified as top glyph, main glyph, and bottom glyphs. When main glyphs are extracted from offline handwritten Telugu characters, some glyphs may have conflicts because of their structural similarities. Therefore, the classifier recognizes incorrect glyphs and the recognition accuracy rate decreases. In the post-

processing step, the recognition errors are rectified to improve the recognition accuracy rate.

The structure of the paper organized as follows. Section-2 describes the literature review, and Section-3 discusses about the proposed method. Section-4 discusses the experimental results, and conclusions of this method are presents in the final section.

2. LITERATURE REVIEW

While conducting the literature review, we found various post-processing methods for character recognition system. Yuan-Xiang et al. [6] developed contextual post-processing method for offline handwritten Chinese script recognition. They corrected the characters based on confusion matrix. They corrected characters of the word with the help of adjoining characters matching in the sentence. Ankita Karia et al. [3] developed recognition system for Devanagari script. At post-processing stage they group the characters based on sequence of characters in the strings. Reference [1] deals with confused group of English characters. From each group they differentiate the characters with the help of signature plots and their spike threshold values.

John F. Pitrelli and Michael P. Perrone applied various confidence-scoring techniques applied to verify whether the recognized output is valid or not [4]. With this method they improved the recognition accuracy rate. Youssef Bassil and Mohammad Alwani developed an error correction algorithm for OCR post-processing in [9]. They developed this algorithm based on spelling suggestions of the goolge. Hisao Niwa et al. [10] corrected the characters based on keyword information at post-processing stage. They used zipf's law to extract the keywords.

3. PROPOSED METHOD

This section contains the proposed post-processing method for glyph-based offline handwritten Telugu characters recognition system.

We developed a glyph segmentation method for offline handwritten Telugu characters at the pre-processing step in [2]. We developed a glyph-based recognition method in [7] for performing feature extraction and classification of the offline handwritten Telugu glyphs.

The proposed post-processing method deals with the groups of confused glyphs, which were recognized at the classification step. Confused glyphs arise because of their structural similarities. We identified five groups of confused Telugu characters based on their glyphs. Each group contained characters with confused glyphs.

We rectified the incorrectly recognized characters, which had confused glyphs from each group based on six variables. If 'G' is the glyph, then we identified six variables at various stages.

1. Let 'n' represent the number of glyphs at the glyph segmentation step.
2. Let 'G_T' be the recognized top glyph at the classification step.
3. Let 'G_M' be the recognized main glyph at the classification step.
4. Let 'G_B' be the recognized bottom glyph at the classification step.
5. Let 'L_M' be the link variable at the glyph segmentation step.

If top glyph is connected to the main glyph before glyph segmentation, then

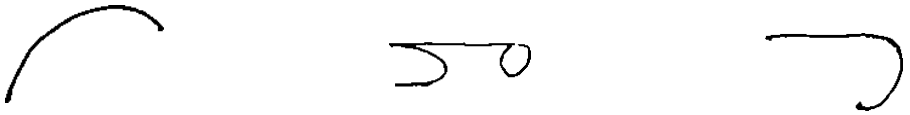
$L_M \leftarrow$ true;

Else

$L_M \leftarrow$false;

End if

6. Let 'S' be the number of small glyphs, which is positioned above the main glyph.



T_1 T_2 T_3

(a) **(b)** **(c)**

Fig. 1. Top glyphs

Let 'M_i' and 'T_j' represent the main glyph and top glyphs with index 'i' and 'j' respectively. We corrected the confused glyphs with the help of top glyphs T_1, T_2 and T_3 are shown in Fig.1.

This section has been divided into five sub-sections; each sub-section represents each group of confused characters.

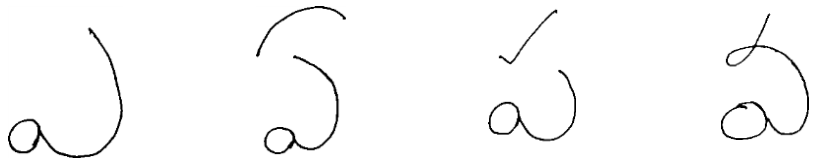
3.1 Group-1

In this sub-section, we deal with four offline handwritten Telugu characters that have confused glyphs identified as Group-1. The confused glyphs of Group-1 characters are shown Fig.2.

The recognized glyph G_M will correct with the help of top glyph G_T, link variable L_M and main glyphs M_i where i=1, 2, 3, 4.

The following steps are used to correct G_M for group-1 confused glyphs.

1. If $G_T = \phi$ then $G_M \leftarrow M_1$.
2. If $G_T = T_1$ then $G_M \leftarrow M_2$.
3. If above two conditions failed then if L_M = false then $G_M \leftarrow M_3$ else $G_M \leftarrow M_4$.



M_1 M_2 M_3 M_4

(a) **(b)** **(c)** **(d)**

Fig. 2. Confused Glyphs of Group-1 Characters

3.2 Group-2

In this sub-section, post-processing was performed for four offline handwritten Telugu characters with confused glyphs identified as Group-2. The confused glyphs of Group-2 characters are shown Fig.3.

These confused glyphs were rectified by correcting G_M with the help of top glyph G_T, bottom glyph G_B, number of small glyphs 'S' and main glyphs M_i, where i= 1, 2, 3, 4. Let $G_M \in \{M_1, M_2, M_3, M_4\}$ then the following steps are used to correct G_M for group-2 confused glyphs.

1. If $G_T = \phi$ and S=0 and $G_B = \phi$ then $G_M \leftarrow M_1$
2. If ($G_T = T_1$ or S≠0) and $G_B = \phi$ then $G_M \leftarrow M_2$
3. If $G_T = T_2$ and $G_B = \phi$ then $G_M \leftarrow M_3$
4. If above three conditions failed then $G_M \leftarrow M_4$

M_1 M_2 M_3 M_4

(a) (b) (c) (d)

Fig. 3. Confused Glyphs of Group-2 Characters

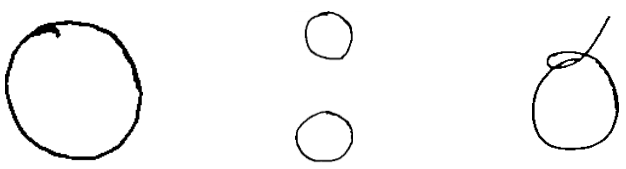
3.3 Group -3

In this sub-section, post-processing was performed for three offline handwritten Telugu characters with confused glyphs as identified as Group-3. The confused glyphs of Group-3 characters are shown Fig.4.

These confused glyphs were rectified by correcting G_M with the help of number of glyphs 'n', link variable 'L_M' and main glyphs 'M_i', where i= 1, 2, 3.

Let $G_M \in \{M_1, M_2\}$ then the following steps are used to correct G_M for group-3 confused glyphs.

1. If n = 1 then $G_M \leftarrow M_1$
2. If n = 2 and L_M= false then $G_M \leftarrow M_2$
3. If above two conditions failed then $G_M \leftarrow M_3$



M_1 M_2 M_3

(a) **(b)** **(c)**

Fig. 4. Confused Glyphs of Group-3 Characters

3.4 Group-4

In this sub-section, post-processing was performed for five sub-groups of offline handwritten Telugu characters with confused glyphs as identified as Group-4. Each sub-group contains the confused glyphs of offline handwritten Telugu characters. The confused glyphs of Group-4 characters are shown Fig.5.

These confused glyphs were rectified by correcting G_M with the help of link variable 'L_M' and main glyphs '$M_{i,j}$', where i= 1, 2, 3, 4, 5 and j=1, 2.

Let $G_M \in \{M_{i,1}, M_{i,2}\}$, where i= 1, 2, 3, 4, 5 then the following conditions are used to correct G_M for group-4 confused glyphs.

1. If L_M = true then $G_M \leftarrow M_{i,1}$
2. If L_M = false then $G_M \leftarrow M_{i,2}$

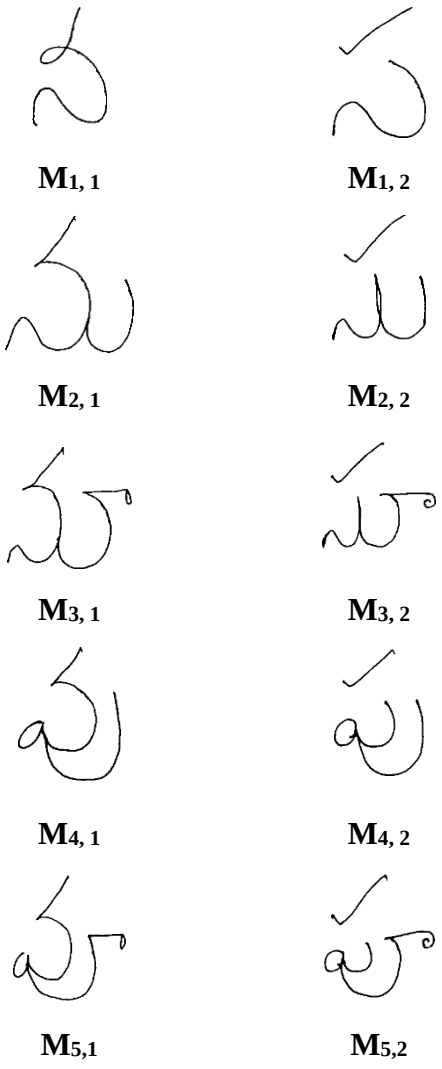


Fig. 5. Confused Glyphs of Group-4 Characters

3.5 Group-5

In this sub-section, post-processing was performed for four sub-groups of offline handwritten Telugu characters with confused glyphs as identified as Group-5. The confused glyphs of Group-5 characters are shown Fig. 6.

These confused glyphs were rectified by correcting G_M with the help of top glyph G_T and main glyphs '$M_{i,j}$' where i= 1, 2, 3, 4 and j=1, 2.

Let $G_M \in \{M_{i,1}, M_{i,2}\}$ then the following conditions are used to correct G_M for group-5 confused glyphs.

1. If $G_T \neq T_3$ then $G_M \leftarrow M_{i,1}$
2. If $G_T = T_3$ then $G_M \leftarrow M_{i,2}$

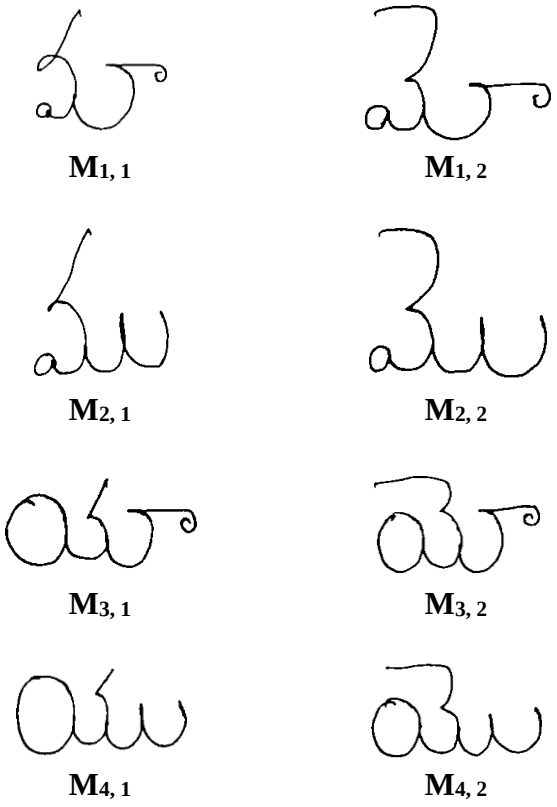


Fig. 6. Confused Glyphs of Group-5 Characters

4. EXPERIMENTAL RESULTS

We collected handwritten Telugu characters from 250 people of different age groups with different professions. Since handwritten Telugu characters does not have any standard datasets [11].

We conducted experiments using this proposed post-processing method for glyph-based recognition of offline handwritten Telugu characters. Confused glyph of the each character result analysis shows in Table-1.

Table 1: Confused Glyphs Result Analysis

Groups	**Glyphs**	**No. of Samples**	**No. of Rectified Glyphs**	**Success Rate**
Group – 1	M_1	250	244	97.60%
	M_2	250	250	100%
	M_3	250	250	100%
	M_4	250	248	99.20%
Group - 2	M_1	250	241	96.40%
	M_2	250	237	94.80%
	M_3	250	240	96%
	M_4	250	239	95.6%
Group - 3	M_1	250	250	100%
	M_2	250	250	100%
	M_3	250	248	99.20%
Group - 4	$M_{1,1}$	250	250	100%
	$M_{1,2}$	250	250	100%
	$M_{2,1}$	250	246	98.40%
	$M_{2,2}$	250	250	100%
	$M_{3,1}$	250	250	100%
	$M_{3,2}$	250	249	99.60%
	$M_{4,1}$	250	250	100%
	$M_{4,2}$	250	250	100%
	$M_{5,1}$	250	250	100%
	$M_{5,2}$	250	248	99.20%
Group - 5	$M_{1,1}$	250	250	100%
	$M_{1,2}$	250	250	100%
	$M_{2,1}$	250	247	98.80%
	$M_{2,2}$	250	250	100%
	$M_{3,1}$	250	250	100%
	$M_{3,2}$	250	246	98.40%
	$M_{4,1}$	250	249	99.60%
	$M_{4,2}$	250	250	100%
Total Glyphs		7250	7182	99.06%

The comparative result analysis of each group is presented in Fig.7. At result comparison of the five groups we got highest success rate for group-3. We got overall success rate is 99.06%.

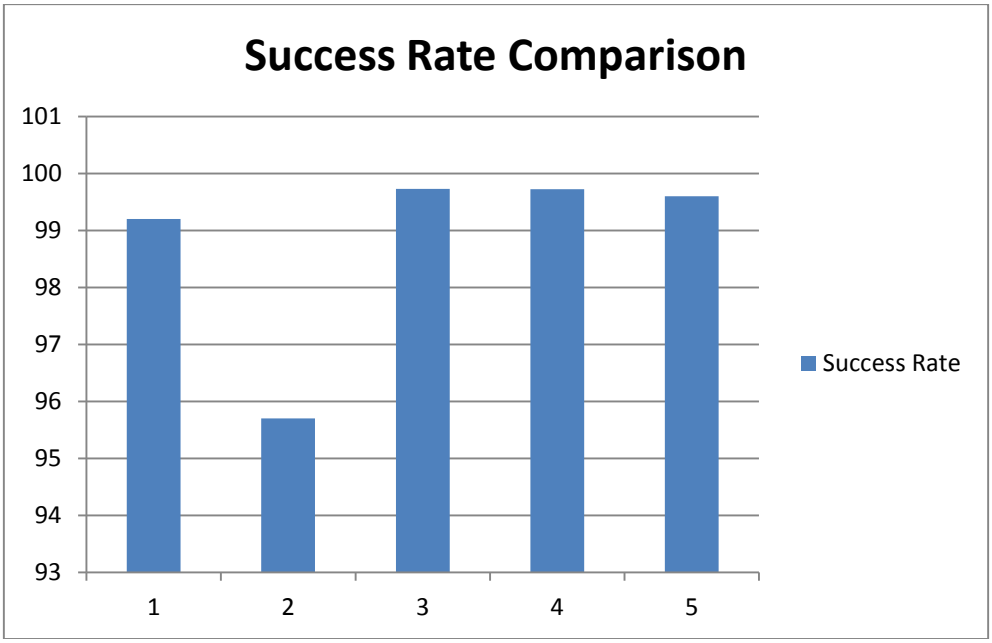


Fig. 7. Result Comparison of Five Groups

5. CONCLUSION

We proposed a post-processing method for glyph-based recognition of offline handwritten Telugu characters, which dealt with the confused glyphs of offline handwritten Telugu characters. We identified five groups of characters, each having confused glyphs. With this post-processing method, we corrected the glyphs that were incorrectly recognized at the classification step. Therefore, we improved the recognition accuracy rate with an overall success rate of 99.06%.

REFERENCES

[1] Binod Kumar Prasad, Goutam Sanyal, "A Model Approach to Off-line English Character Recognition", International Journal of Scientific and Research Publications, Vol. 2, Issue. 6, June 2012, PP. 1- 6.

[2] Ch. N. Manisha, Y.K. Sundara Krishna, E. Sreenivasa Reddy, "Glyph Segmentation for Offline Handwritten Telugu Characters", Proceedings of IC3T 2016, Springer, in press.

[3] Ankita Karia, Sonali Sharma, Reevon Rodrigues, Maitreya Save, "Character Recognition (Devanagari Script)", International Journal of Engineering Research and Applications, Vol. 5, Issue 4, (Part -6), April 2015, pp.109-114.

[4] John F. Pitrelli and Michael P. Perrone (2003, August). Confidence-scoring post-processing for off-line handwritten-character recognition verification. In Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on (pp. 278-282). IEEE.

[5] Glyph, https://glosbe.com/en/te/glyph.

[6] Yuan-Xiang Li, Chew Lim Tan, Xiaoqing Ding, Changsong Liu, (2004). Contextual post-processing based on the confusion matrix in offline handwritten Chinese script recognition. Pattern Recognition, 37(9), 1901-1912.

[7] Ch. N. Manisha, E. Sreenivasa Reddy, Y.K. Sundara Krishna, "Glyph-Based Recognition of Offline Handwritten Telugu Characters", Submitted.

[8] Telugu language. (2016, October 3). In Wikipedia, The Free Encyclopedia. Retrieved 18:39, October 5, 2016, from https://en.wikipedia.org/w /index.php?title=Telugu_language&oldid=742443661.

[9] Youssef Bassil, Mohammad Alwani. "OCR Post-processing Error Correction Algorithm using Google's Online Spelling Suggestion", Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No. 1, January 2012.

[10] Hisao Niwa, Kazuhiro Kayashima, Yasuham Shimeki, "Postprocessing for Character Recognition using Keyword Information, MVA'92 IAPR Workshop on Machine Vision Applications, PP. 519-522.

[11] Sastry, P. N., Lakshmi, T. V., Rao, N. K., Rajinikanth, T. V., Wahab, A., "Telugu Handwritten Character Recognition Using Zoning Features," International Conference on IT Convergence and Security (ICITCS 2014), IEEE, pp. 1-4, October, 2014.