Slant Correction for Offline Handwritten Telugu Isolated Characters and Cursive Words

Ch. N. Manisha*
Research Scholar, Department of Computer Science,
Krishna University, Machilipatnam, Andhra Pradesh, India.

Y.K. Sundara Krishna
Professor, Department of Computer Science, Krishna University,
Machilipatnam, Andhra Pradesh, India.

E. Sreenivasa Reddy
Professor, Department of Computer Science Engineering,
Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

Abstract
Slant correction is one of the preprocessing steps for the character recognition. This proposed method is simple and efficient for correcting the slant of offline handwritten Telugu isolated characters and cursive words. This method is motivated from the chain code of the word contour slant correction method but this proposed method does not identify the chain code of the word or character contour. This method uses the vertical starting and ending co-ordinates for estimating the slant. It estimates the slant direction using the maximum vertical projection profile and corrects the slant of offline handwritten Telugu isolated characters and cursive words using shear transformation. The experimental results demonstrate comparative analysis of the efficiency of the proposed method compared with that of the existing method. The proposed method efficiently reduced the processing time compared with the existing methods.

Keywords: Slant, Telugu, Characters, Offline, Handwritten, Correction.

Introduction
Slant correction is a major step influencing the segmentation step. It is a crucial step of pre-segmentation. In offline handwritten character recognition we identified slant correction is one of the important level at preprocessing stage in [4]. Slant correction uses vertical co-ordinates of the character for correcting the slant. The proposed method deals with offline handwritten Telugu isolated characters and cursive words. The cursive and non-cursive Telugu words that differ with the writing style of people. Some people wrote the characters cursively, segmentation of which is the toughest task because the cursive words comprise slant errors. Whereas, the non cursive Telugu words comprise non uniform slant errors for each character. In this paper, we corrected the non-uniform slanted Isolated characters and cursive words. We proposed a novel approach for estimating and correcting the slant of offline handwritten Telugu isolated characters and cursive words. This method is a very efficient for correcting slant of offline handwritten Telugu isolated characters and cursive words. A proper slant corrected cursive words can easily segment and a proper slant corrected isolated characters increases the recognition accuracy. Therefore, slant correction is required for Telugu isolated characters and cursive words.

The remaining paper is organized as follows. Section-2 presents the related work of slant correction in various languages, Section-3 explains the preprocessing step, and Section-4 presents the proposed method describes the slant estimation and correction methods for offline handwritten Telugu isolated characters and cursive words. Section-5 discusses and presents the results in two subsections; First subsection presents the comparative result analysis of the existing method and proposed method and Second subsection discusses comparative analysis of the processing time of the existing method and proposed method. Section-6 provides the conclusion of the study.

Related Work
Alceu de S. Britto Jr. et. al. [1] implemented a modified KSC method to correct slant of the numeral strings. This method deals with overlapped numerals. Taira et. al. [21] corrected non-uniform slant of handwritten English words. Dynamic program based algorithm developed for that. Mohamed et. al. [14] analyzed the slant of the online signatures with vector-rule based approach. Bušta et.al. [3] proposed five skew estimators to correct the text skew in scene image. They proposed vertical dominant, VD on the convex hull, longest edge, thinnest profile and symmetric glyph for it. Yamaguchi et. al. [23] classified digits for telephone numbers on digitized signboards. In that they corrected the slant of numbers by circumscribing digits with the tilted rectangles method. Kimura et.al. [11, 12] corrected the slant of handwritten words with chain code of the word contour. Madhvanath et.al., [13] used chain codes to extract vertical line elements. They
calculate slant angles with the start and end points of each line and global slant angle calculate by the average of all vertical line slant angles. They corrected by the shear transformation of the angle.

Ali and Jumari [2] detected slant angle by weiner-ville distribution. The authors calculate all histograms by WVD. The highest intensity of histograms taken as non slanted word. Sun and Si [20] were detected slant of characters in text line by gradient operation, corrected by shear operation and connected components can be corrected by fitting parallelograms.

Ding et.al., [8] proposed simple high speed iterative method and eight directional chain code to corrected the slant for English words. Ding et. al., [7] performed four, eight, twelve and sixteen directional chain codes used for correcting slant of words in the English language. The authors shown eight directional chain codes gave better results than other directional chain codes.

Ishaan et. al., [9] reconstructed the images by Zernike moments. They observed radon transform gives better result than Hough transform. They corrected slant of both English and Hindi text. Kavallieratou et. al., [10] were developed a slant removal algorithm for English and Greek text. They estimated slants by vertical projection and WVDs. Papandreou and Gatos [17, 18] detected core region depends on the upper and lower baseline and black run profiles used to removed small fragments and corrected the slant. de Zeeuw [5] Identified slants by using the vertical projection histogram. They calculate different vertical projection histograms between -45° to 45° and which has the highest peak of histogram consider for slant correction.

Preprocessing
Before applying propose method, we preprocessed the image for noise removal, binarization, and skeletonization. Noise removal is a crucial step in preprocessing. Most handwritten documents suffer from salt-and-pepper noise. Removal of noise discussed by Nair et. al., [15]. The conversion of a grayscale image to a binary image is called binarization. For binarization, we used the Otsu’s threshold method. The Otsu’s threshold method described by Otsu in [16]. Skeletonization is a process of extracting the essential structure of characters. Deng et.al., [6] developed a fast parallel algorithm for skeletonization. Figure 1 shows the Telugu character and cursive word.

![Figure 1: Before Slant Correction (a) Offline Handwritten Telugu Character and (b) Offline Handwritten Telugu Cursive Word](image)

Proposed Method
In this section, we present the proposed method for slant estimation and correction of offline handwritten Telugu isolated characters and cursive words.

We proposed a novel approach for estimating the slant of offline handwritten Telugu isolated characters and cursive words. The slant estimation method is motivated from the chain code word contour slant correction method in [12]. The chain code sequence (see Figure 2) and the slant estimation angle from Kimura et.al., [12] in Eq. (1). The proposed method does not use the chain code of the word or character contour.

For the extraction of the isolated characters and cursive words from the document images we used horizontal and vertical projection profiles. For isolated characters based words contains non uniform slant errors. Therefore each isolated character can be corrected separately.

\[
\theta = \tan^{-1} \left( \frac{n_1 + n_2 + n_3}{n_1 - n_3} \right) \tag{1}
\]

Let ‘m’ and ‘n’ be the width and height of the character image ‘I’. Furthermore, ‘x’ and ‘y’ represent the vertical and horizontal co-ordinates of the pixel, respectively. We identified the starting and ending vertical co-ordinates of the character in the image, which is represented as ‘x₁’ and ‘xₘ’ respectively (see Figure 3).

We simplified Eq. (1) for reduce the processing time and increase the efficiency of the method. Instead of ‘m’ and ‘n’, we replaced with ‘x₁’ and ‘xₘ’. Instead of ‘n’, we replaced with Eq. (2).

\[
n_2 = \left( \frac{x_1 + x_m}{2} \right) \tag{2}
\]

For the slant estimation of offline handwritten Telugu characters, ‘x₁’ and ‘xₘ’ were used. We substituted ‘x₁’ and ‘xₘ’ in Eq. (1) to obtain the slant estimation angle ‘θ’ for the offline handwritten Telugu characters (see Eq. (3)).

\[
\theta = \tan^{-1} \left( \frac{1.5(x_m + x_1)}{x_m - x_1} \right) \tag{3}
\]

It is necessary to identify the slant direction. Because it decides positive or negative sign of the slant angle.

For correcting the slant of a character, shear transformation is widely used. It is also called Transvection [19]. Shear transformation is generally of two types, horizontal and vertical shear transformations. We used the vertical shear transformation that shifts each vertical co-ordinate in a particular direction. For applying the shear transformation, we used ‘θ’ and get new pixels (Δx, Δy), as shown in Eq. (4) and Eq. (5).

\[
\Delta x = x + y \times (\theta \times 0.01) \tag{4}
\]

\[
\Delta y = y \tag{5}
\]

With our experiments it is necessary to adjust the angle ‘θ’ for shear transformation for better slant correction. From our experiments we used Trial-and-Error approach to set 0.01 as a value to adjust ‘θ’.
We estimated slant angle ‘\(\theta\)’ for the offline handwritten Telugu characters. It is necessary to identify the direction of the slant. Direction of the slant will decide the sign of the angle that is positive or negative angle of slant correction angle.

De Zeeuw [5] estimated slant angle with the help of vertical projection histograms. The authors shear transformed the image with different angles from \(-45\) to \(+45\) then computed the vertical projection histograms for each angle. Among the vertical projection histograms of the word, the highest peak of vertical projection histogram was considered as the slant corrected word.

In this paper, we considered only two angles for verification. For selecting negative or positive \(\theta\), we used \(-\theta\) and \(+\theta\) in eq. 4 and obtained two images, say \(I_1\) and \(I_2\) (See Eq. (6)), we determined the maximum vertical projection profile value for \(I_1\) and \(I_2\) (See Eq. (7)).

\[
V_j = \sum_{i=1}^{n} I(j,k)
\]

(6)

\[
L = \text{MAX}(V_1, V_2, ..., V_m)
\]

(7)

Let \(L_1\) and \(L_2\) be the maximum vertical projection profile values for \(I_1\) and \(I_2\), respectively. The image corresponding to higher \(L\) value between the \(L_1\) and \(L_2\) is selected as the slant corrected image (see Eq. (8)).

\[
L_3 = \begin{cases} 
I_1 & \text{if } L_1 > L_2 \\
I_2 & \text{else}
\end{cases}
\]

(8)

Figure 4 shows the offline handwritten Telugu character and cursive word after slant correction. Processing time and slant angle analysis are shown in TABLE.1.

![Figure 4: After Slant Correction](image)

### Experimental Results

We collected offline handwritten Telugu cursive words written by teachers and students and characters from the hpl-telugu-iso-char-offline dataset [22]. In this section, we present a comparative analysis of the existing chain code of the word contour slant correction method (CCWC) and proposed method (PM). In First subsection, we compare the results of the chain code of the word contour slant correction method and proposed method. In Second subsection, we compare the processing time of the chain code of the word contour slant correction method and proposed method.

#### Comparative Analysis of Testing Results

The comparative analysis of the chain code of the word contour slant correction method result and proposed method result shown in Figure 5. We compared results of the chain code of the word contour slant correction method and proposed method for offline handwritten Telugu isolated characters and cursive words.

We analyzed five offline handwritten Telugu cursive words. The results of the comparative analysis shown in Figure 6. The comparative analysis of slant angle correction for offline handwritten Telugu isolated characters and cursive words shown in Figure 7 and Figure 8 respectively.
**Figure 5:** Slant Correction Comparison of Offline handwritten Telugu isolated characters (a) Before Slant Correction (b) Slant corrected with Chain code of the Word Contour Method (c) Slant corrected with Proposed Method

**Figure 6:** Slant Correction Comparison of Offline handwritten Telugu cursive words (a) Before Slant Correction (b) Slant corrected with Chain code Word Contour Method (c) Slant corrected with Proposed Method

**Comparative Analysis of Processing Time**
We used a 3.00 GHz processor and 32-bit operating system for this experiment. We compared the processing time of chain code of the word contour and proposed methods for offline handwritten Telugu isolated characters and cursive words, as shown in Figure 9 and Figure 10, respectively. Average processing time comparison between chain code of the word contour method and proposed method for offline handwritten Telugu isolated characters and cursive words as shown in Figure 11.

**Figure 7:** Comparative Analysis of Slant Angle Correction for Isolated Characters.

**Figure 8:** Comparative Analysis of slant Angle Correction for Cursive words

**Figure 9:** Comparative Analysis of Processing Time for Offline Handwritten Telugu Isolated Characters

**Figure 10:** Comparative Analysis of Processing Time for Offline Handwritten Telugu Cursive words

**Figure 11:** Avg. Processing Time Comparison of Chain Code Word Contour and Proposed Method for Offline Handwritten Telugu Isolated Characters and Offline Handwritten Telugu Cursive words
Conclusion
In this study, we proposed a novel and simple method for slant correction of offline handwritten Telugu isolated characters and cursive words. The slant correction for offline handwritten Telugu isolated characters and cursive words was motivated from the chain code of the word contour slant correction method in [12]. However, the proposed method does not identify the chain code of the character contour. This method deals with non-uniform slanted offline handwritten Telugu isolated characters and cursive words.

In this proposed method, the starting and ending vertical co-ordinates of offline handwritten Telugu characters were considered for slant angle estimation. To transform the direction of the vertical co-ordinates, shear transformation was used. To identify the slant direction, the estimated negative and positive slant angles were applied to shear transformation. For the slant angle estimation, maximum vertical projection profiles corresponding angle selected as the slant corrected image. We used handwritten Telugu characters and cursive words from different people and dataset [22] for testing, and the proposed method efficiently corrected the slant of offline handwritten Telugu isolated characters and cursive words.

References