International Journal of Engineering & Technology, 7 (3) (2018) 1846-1853

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i3.13286
Research paper

A novel method for segmenting and straightening of text lines in handwritten Telugu documents based on smearing and regression approach

Mslb. Subrahmanyam [1] *, V. Vijaya Kumar [2] , B. Eswara Reddy [3]

[1] Research Scholar, JNTU Kakinada, Kakinada, 533001, India
[2] Director for CACR, Dean Dept. of CSE &IT, Anurag Group of Institutions (Autonomous), Hyderabad, India
[3] Professor of CSE Dept. & Principal of JNTU-A College of Engineering, Kalikiri, India
**Corresponding author E-mail: [1]subrahmanyam_mtech@rediffmail.com*

Abstract

In handwritten document images, segmenting text lines is a very challenging task due to various reasons like variability in intra baseline skew and inter line distance between text lines. So far, no work is reported in the literature for the straightening of handwritten Telugu languages. Telugu is one of the most popular languages of India that is spoken by more than 80 million people especially in South India. Telugu characters are mostly compound characters and that is way the straightening task of Telugu document is more challenging tasks than European languages. This paper introduces a novel approach for segmenting and straightening text lines of handwritten Telugu documents based on smearing and regression approach (SRA). This method initially performs preprocessing and estimates parameters by dividing into connected components of Telugu script. A horizontal and vertical run length-smearing algorithm is used in this paper to shape text lines. To identify text lines more precisely cubic polynomial regression is used between vertical midpoints of two blocks of compound handwritten Telugu characters. A simple logic is derived on this to achieve final process. We tested the proposed algorithm with three different kind of 1000 handwritten documents. The performance of proposed method is evaluated by using matchScore, detection rate, recognition accuracy and F-measure. The experimental results indicates the efficiency of the proposed method over the existing methods.

Keywords: *Telugu Languages Text Lines; Compound Characters; Run Length Smearing; Cubic Polynomial Regression.*

1. Introduction

Digitization playing major rule now a days for reducing physical storage of huge paper documents. Optical character recognition plays important role in this area. Preprocessing, segmentation, feature extraction and classification are major stages in this process. In Segmentation stage lines, words and characters are extracted from input document image. Text lines identification or extraction from the document image is one of major and challenging task in segmentation stage. It effects the accuracy of remaining all other stages in this process if not identified properly. Text line segmentation is very difficult in case of handwritten documents compared to printed documents because of variability in intra baseline skew and inter-line distance. Touching or overlapping of characters between adjacent lines at single point or multiple point's also frequent problem in handwritten documents.

In the literature [1], [42-48] a lot of Text line segmentation methods were presented which can be broadly divided into seven major categories [1] based on Projection-based methods[12], [37], [38], [40], Smearing methods [20], [42], [42],Grouping methods [7], [15], [16], [29], [39], [41], Methods based on the Hough transform[10], [16], [24], Repulsive–attractive network method[48] , Stochastic method and Processing of overlapping and touching Components. This method well suited for printed documents which contains uniform spacing between text lines but not suitable for unconstrained handwritten documents. In [2] Zahour, A et al. used partial projections to modify this method for unconstrained handwritten Arabic documents. Bar-Yosuf et al [3] used piecewise projection profile for extraction text lines. Shi and Govindaraju [4] identified text lines using fuzzy directional run length. In [5] Darko Brodi proposed method for extracting lines in handwritten document using water flow algorithm. In this paper, they used power function counterpart instead of linear water flow function. Extracting text lines from multi skewed handwritten documents Rajath. A.N proposed method in [6]. G. G. Rajput et al [7] applied histogram and connected component analysis for extracting text lines from Handwritten documents. In [8] Dibyayan Chakraborty proposed method for detecting base line from multi-lingual multi-turn handwritten document images. Problems and review of various text line segmentation methods for handwritten documents are presented in [9]. In [10] Satadal Saha et al. proposed method using Hough transform technique for extraction text line from handwritten documents. In [11] Sunanda Dixit et al. proposed method for Hindi and English documents using Hough Transform. Amreen Singh et al. [12] proposed method for extracting text lines from handwritten Gurumukhi script using horizontal and vertical projection profile. In [13] M.Ravi kumar et al. proposed thresholding approach using clustering method. Jayant Kumar et al. [14] introduced graph-based method and EM algorithm to correct errors for handwritten document images. Using connected component approach Jewoong Ryu [15] proposed language-independent method. G. Louloudis et al. [16] proposed text line segmentation with Hough transform on connected components and post processing using skeletonization technique. In [17] Zaidi Razak et al. presented a paper on survey of of-

fline text line detection from different existing handwritten document images and in [18] Chethana H T et al. proposed a paper on survey of line segmentation from handwritten Kannada documents images. In [19] authors proposed for Persian script using Morphological approach. Samir Malakar et al.[20] proposed spiral run length smearing approach. Energy minimization frame work was proposed by Hyung Il Koo et al. [21] for segmenting text lines in Chinese documents. Piecewise painting algorithm was proposed by authors in [22] for Persian handwritten scripts. Rodolfo P[23] proposed using morphology and histogram. G. Louloudis et al. [24] proposed based on connected component and Hough transform approach. In [25] authors proposed text line segmentation in printed or handwritten document images based on fringe maps. Using interdependency between text lines and inter line gap Bidyut B. Chaudhuri et al proposed method in [26]. In [27] P. Nagabhushan et al. proposed painting technique for Persian or Arabic scripts. In [28] Afaz Uddin Ahmed et al. proposed segmentation of number plate for secure garage system. In [29] Laurence Likforman-Sulem and Claudie Faure proposed an iterative process by perceptual grouping technique. From survey, more research available for English, Arabic, Chinese, Persian handwritten text line segmentation but relatively less work has been reported for Indian scripts. Because complexity of scripts in Indian languages [30].

The straightening of text lines of handwritten document very important for improving OCR accuracy, beautification of the document and also human readability. Still no specific paper was not published for straightening of Telugu documents in the literature. This paper proposes a novel method for identifying and straightening of handwritten Telugu text lines using smearing and regression approach.

This paper is organized as follows details about Telugu script characteristics are given in Section 2, section 3 describes our proposed method, Section 4 contains experimental results and conclusions are given in Section 5.

2. Telugu script characteristics

Telugu is one of the most popular languages of India that is spoken by more than 80 million people especially in South India [45]. Telugu is widely spoken Dravidian language and ranks third by the number of native speakers (more than 74 million people). Telugu contains both simple and compound characters. There are 52 simple characters, out of which 16 represents basic vowel sounds, called achchus and 36 represents simple consonants, called hallus. In addition there are two groups of symbols, semi-vowel or vowel modifiers called as 'maatras' and half-consonants called as 'voththus'. The compound characters of Telugu languages are obtained by combination of simple consonants (hallus) with vowel modifiers (maatras) or with half consonants (voththus) or with both maatras and voththus[47]. The number of computed characters of Telugu language is estimated between 50000 and 100000 [47]. The Telugu alphabet set is shown in Fig. 1.Fig 2 shows the some Telugu simple and compound characters.

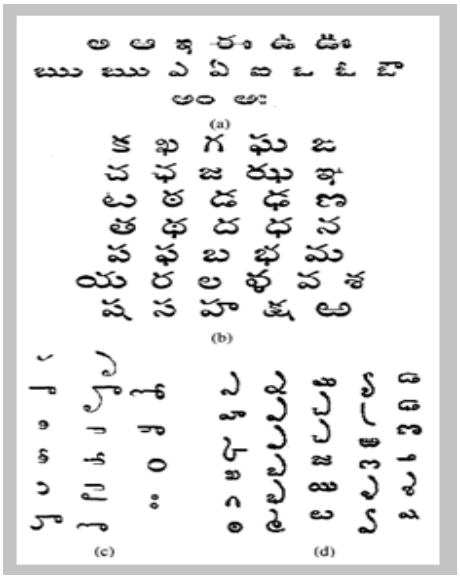


Fig. 1: Telugu Alphabet (A) Achchlu (B) Halulu (C) Maatras (D) Voththulu.

Fig. 2: Samples of Telugu Simple and Compound Characters.

3. Related work

Segmenting and straightening of Text lines in handwritten documents for Indian languages in general more complex than those of European languages due the linguistic complexity of large number of vowel, consonants and various different combinations of vowels and consonants. Relatively less work has been reported until now for Indian languages. In [31] and [32] authors proposed methods for handwritten Kannada script. In [33] Saiprakash Palakollu et al. proposed method for text segmentation of Hindi handwritten documents. In [34] authors proposed for Handwritten Bangla document based on water reservoir principle. Vishwas H. S et al. [35] proposed smearing techniques for Kannada document. In [36] authors proposed midpoint detection approach for Gurmukhi handwritten documents. In [37] Dr.S.Pannirselvam et al. proposed projection method for Tamil handwritten documents. For Devanagari skewed and overlapped script Rahul Garg et al. [38] proposed method for text line detection using projection profile technique. In [44] N. Shobha Rani et al proposed technique for handwritten Telugu script recognition. In [46] authors proposed for Telugu printed documents.

4. Methodology

The segmenting and straightening process of handwritten Telugu document is more challenging task because unlike European languages, majority of Telugu alphabets are compound characters with maatras and voththus. Also with similar looking shapes, sizes and lot of variation between different writing styles and presence of more number of compound characters makes it most challenging task for straightening text lines of a Telugu handwritten text. After an in-depth study on the Telugu handwritten documents this paper proposes a novel method for identifying and straightening text lines using SRA model. Fig 3 shows the block diagram of proposed SRA model.

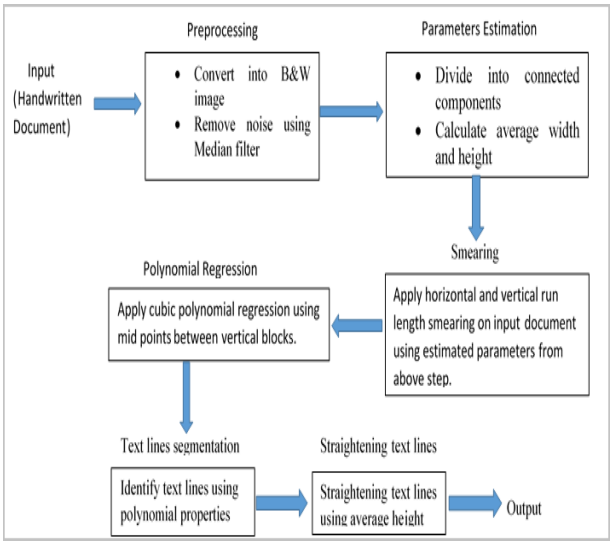


Fig. 3: Block Diagram of Proposed SRA Method.

The proposed detecting and straightening of text lines in handwritten Telugu document is given below.

Step 1: The proposed SRA method initially performs a preprocessing operation based on Otu's [47] algorithm to convert given color or gray scale document into black and white image. Fig 4(A) shows the original Telugu handwritten text document. Fig 4(B) shows image after converting into black and white from original Telugu handwritten document 4(A).

Step 2: To estimate parameters for further steps the present papers initially derives a connected component (CC) on Telugu document using 8-connectivity neighborhood algorithm and calculates the average width(aw) and average height(ah) of all CC's[43] using following formulas.

$$aw = \frac{1}{N}\sum_{k=0}^{N} w_k \quad (1)$$

$$ah = \frac{1}{N}\sum_{k=0}^{N} h_k \quad (2)$$

Where N is total count of CC's with in the document and $\{w_1, w_2, w_N\}$ and $\{h_1,h_2\ldots..h_N\}$ are width and heights of individual connected components.

Step 3: To identify text lines the present paper derived two run length smearing algorithms using following procedure. The first algorithm runs in horizontal direction with threshold t_1 and second algorithm runs in vertical direction with threshold t_2 using following equations 3 and 4 respectively.

$$t_1 = t_w * aw \quad (3)$$

$$t_2 = t_h * ah \quad (4)$$

Where aw and ah are calculated from equations 1 and 2 respectively and t_w and t_h are threshold coefficients for width and height respectively. We experimented with different values of t_w and t_h on large number of Telugu handwritten document images and finally achieved optimal solution with selection of 4 and 0.5 respectively.

Algorithms for run length smearing:

A character is composed of set of pixels. The set of pixels that are part of character (foreground) is represented with intensity equal to 0 and the set of pixels which are not part of character (background) is represented with intensity equal to 1.Let W denote width and H denote height of the original image.

Algorithm 1: horizontal smearing

```
Begin
 For each row y in H
 Begin
   var count = 0;
   var flag = 0;
   var position = -1;
   For each column x in W
    Begin
      if(Intensity [x,y] == 1)//check for bg pixel
       Begin
          if(position == -1)
            position = y;
        flag = 1;
        count++;
        End
      else if (flag == 1 && count <= t1)
     Begin
      For each z in [position count]
        Intensity[z,y] = 0; //convert into fg pixel
        count = 0;
        flag = 0
        position = -1;
       End
     End
End.
```

Algorithm 2: vertical smearing

```
  Begin
   For each column x in W Do
      Begin
        var count = 0;
        var flag = 0;
        var position = -1;
       For each row y in H Do
           Begin
             if(Intensity[x,y] == 1)//check for bg pixel
                Begin
                    if(position == -1)
                      position = y;
                     flag = 1;
                     count++;
                End
              Else if (flag == 1 && count <= t2)
             Begin
              For each z in [position count]
                  Intensity[x,z] = 0; //convert into fg
             flag = 0;
             count = 0;
             position = -1
            End
          End
       End
  End.
```

To make all blocks equal size the present paper extend each one with line to nearest edge neighbor boundaries. Fig 4(c) represents image after smearing.

Step 4: The handwritten text lines of Telugu characters are basically complex in nature and it may contain different flow of writing, style, character stroke etc. Also since Telugu language contains compound characters it is very difficult to separate two adjacent text lines. To identify such complex text lines more precisely the present paper proposed cubic polynomial regression technique to estimate vertical space between two blocks. To estimate them the present paper initially divides the document into connected components to shape text lines and then applies cubic polynomial regression technique between two adjacent vertical blocks using following procedure.

Cubic polynomial regression:

$$\text{Let } f(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \quad (5)$$

Where b_0, b_1, b_2 and b_3 are constant coefficients and f(x) be function of x. Let $\{(x_1,y_1),(x_2,y_2),\ldots\ldots\ldots(x_n,y_n)\}$ be the set of mid points between two adjacent vertical blocks and its normal equation in matrix form can be represented using equation 6.

$$Y = Xb \quad (6)$$

Where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & x_{n-1}^3 \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

And

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

From this, we can calculate matrix 'B' values using following formulas (7) and (8).

$$B = (X^T X)^{-1} X^T y \quad (7)$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \\ x_1^2 & x_2^2 & x_3^2 & & x_n^2 \\ x_1^3 & x_2^3 & x_3^3 & \dots & x_n^3 \end{bmatrix} \quad (8)$$

Where X^T and X^{-1} represents transpose inverse of X respectively.
Fig 4(D) represents image after cubic polynomial regression between adjacent vertical blocks on 4(C).
Algorithm 3: polynomial regression on vertical midpoints between blocks.
Let N be the total number of connected components after step 4. Let us be the set of vertical midpoints between two adjacent connected components. Let R be the set of all cubic polynomial regressions for the entire document.

```
Begin
    For each connected component c in N Do
    Begin
       Consider next connected component c+1.
       For each column h in W do   //for each column.

        Begin
           Calculate vertical midpoint (xmh,ymh) between c and c+1.
           (xmh,ymh) ε S//assign to S.
         EndFor
            Apply cubic polynomial regression on S using equations 7 and 8 and let it be Q
        Q ε R.//add to the list of R.
        S = φ //make S empty for next iteration.
    End for.
End
```

Step 5: After polynomial regression between two vertical blocks the block may fall above or below or on the polynomial curve f(x). The present paper derives the following algorithm to identify the text line from all blocks from step 2 whose center of gravity coordinate (xg , yg) satisfies the equation 9. The center of gravity coordinate (xg , yg) of each block are calculated using equations 10 and 11.

$$Yg >= f(x_g) = b_0 + b_1 x_g + b_2 x_g^2 + b_3 x_g^3 \quad (9)$$

$$x_g = \frac{1}{N}\sum_{k=0}^{N} x_k \quad (10)$$

$$y_g = \frac{1}{N}\sum_{k=0}^{N} y_k \quad (11)$$

Where N is total number of foreground pixels with in the block and $\{(x_1, y_1),(x_2,y_2,),\dots\dots(x_N,y_N)\}$ are coordinates of foreground pixels in the block.
Algorithm 4: Segmentation of Telugu handwritten text lines after polynomial regression.
Input: Let R be total number of regressions after step 4.Let M be the total number of considered blocks for each R from step 2. Let K be the set of all blocks with in the same text line.
Output: Identify text line with all blocks belongs to K.

```
Begin
    For each k in R do
      Begin
        For each p in M do
          Begin
             Calculate center of gravity coordinate (xg,yg) of the
             block 'p' using equations 10 and 11.
             Calculate the value of f(xg) using equation 9.
             If (yg >= f(xg))
               Assign block 'p' to K i.e p ε K.
          End for
      End for.
End.
```

Fig 4(E) represents image after identifying text lines in 4(A).
Step 6: To straightening the identified Telugu handwritten text lines from the above step the present paper derives the following algorithm using center of gravity (x_{lg},y_{lg}) of the each identified text line from the following equations .

$$x_{lg} = \frac{1}{N}\sum_{k=0}^{N} x_{gk} \quad (12)$$

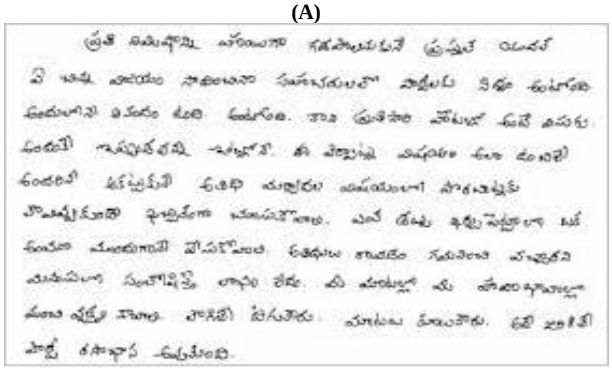
$$y_{lg} = \frac{1}{N}\sum_{k=0}^{N} y_{gk} \quad (13)$$

Where N is the total count of blocks with in the identified text line and {(xg1, yg1), (xg2, yg2),……..,(xgN,ygN)} are coordinates of center of gravity of each block with in the identified text line.
Algorithm 5: straightening of the identified Telugu handwritten text lines.
Let M be the set of all blocks with in the identified text line.

```
Begin
 Calculate center of gravity (xlg,ylg) of all blocks with in the identified text line  using formulas 12 and 13.
For each p in text line blocks M
 Begin
       Calculate center of gravity (xgp,ygp) of the block p using formula 10 and 11.
       if(ylp > ygp)
         move the block 'p' upward with the distance equal to (ylp – ygp).
       If(ylp <= ygp)
         move the block 'p' downward with the distance equal to (ygp – ylp).
     End
End.
```

Fig 6 explains the proposed procedure for straightening of identified text lines from handwritten Telugu document. Fig 4(F) shows image after straightening text-lines in 4(A).

(A)



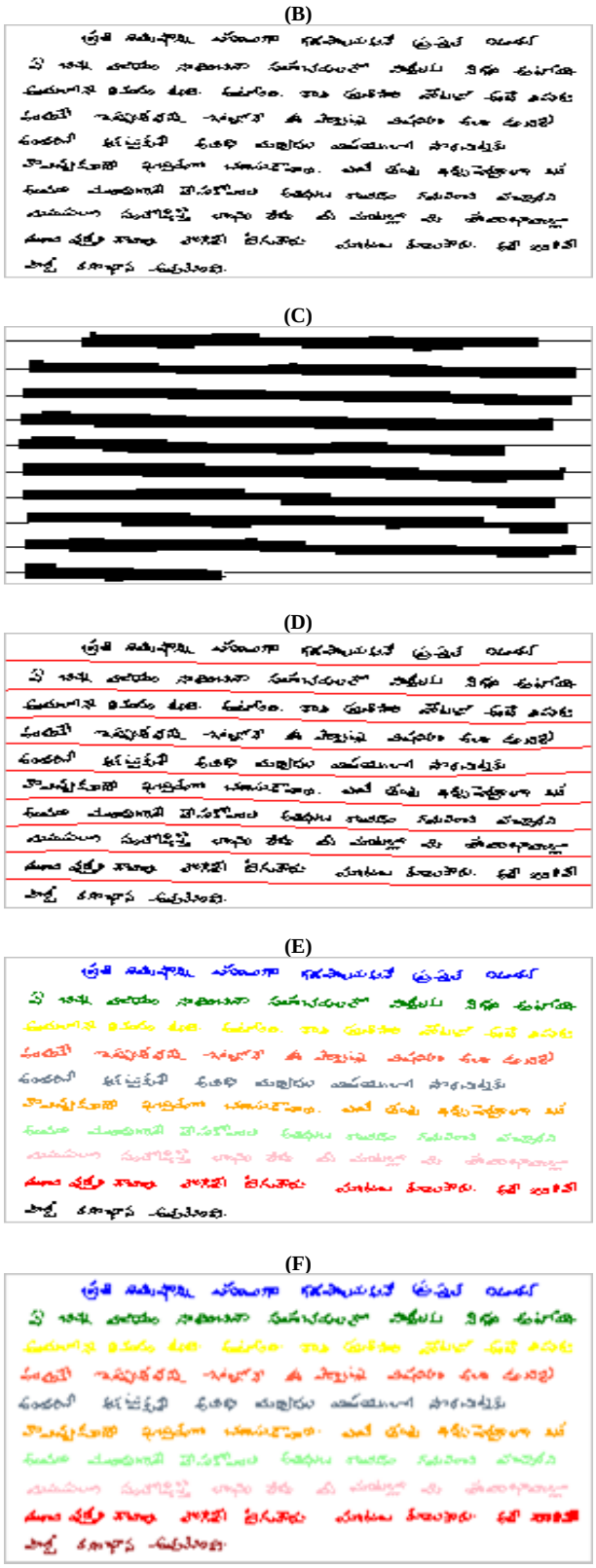


Fig. 4: (A) Represents Original Telugu Handwritten Document (B) Represents after Dividing into Connected Components from (A). (C) Image after Applying Horizontal and Vertical Smearing on (A). (D) Image after Applying Cubic Polynomial Regression between Adjacent Vertical Blocks on (C). (E) Image after Identifying Text-Lines. (F) Image after Straightening the Text Lines.

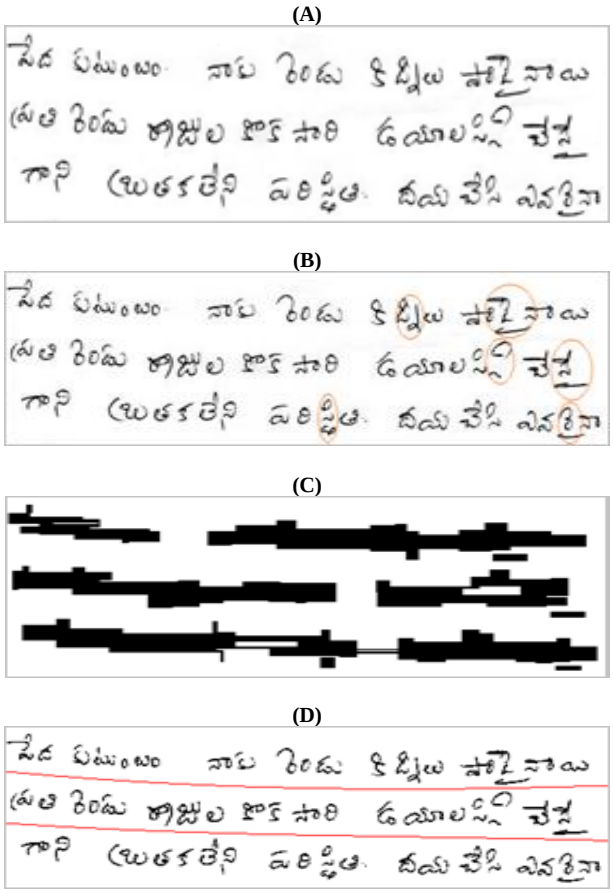


Fig. 5: (A) Telugu Handwritten Image (B) Marked with Red Circles on Compound Characters on Input Image (A). (C) Image after Step 3. (D). Image after Applying Cubic Polynomial Regression between Adjacent Vertical Blocks on (A).

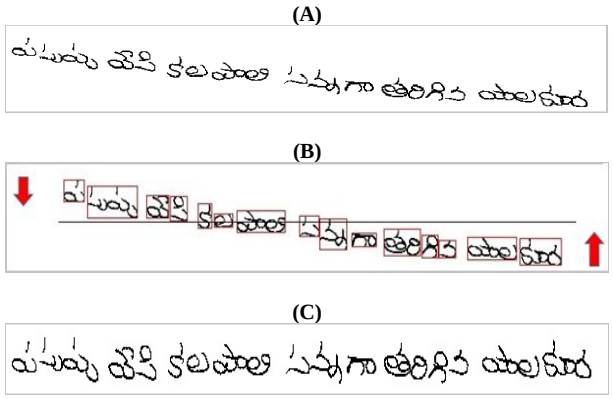


Fig. 6: (A) Identified Text Line (B) Middle Line Represents Center of Gravity of the Text Line and Arrows Represent the Move Upward or Downward the Blocks with in the Text Line Respectively. (C) Straightening Text-Lines of (A).

5. Experimental results

This paper considered a total of 1000 handwritten Telugu documents representing three categories. In category one the present paper considered 550 handwritten documents collected from individuals from school children, house wives, employees etc. In second category the present paper considered 330 Telugu printed documents. Samples of these documents are shown in Fig 7. This paper also considered 120 handwritten text documents from Indian languages (other than Telugu) as category 3 and sample images are shown in Fig 8, 9 and 10.This paper initially experimented with different degrees of polynomials like 3,4 and 5. The proposed method achieved on 99.46%, 94.37% and 89.52% with polynomial degree

of [3-4] and 5 respectively. The results indicates the efficiency of the proposed SRA model.

The performance of proposed SRA model evaluated using Match-Score [15], [49],[50] table, Detection rate (DR), recognition accuracy (RA) and F-Measure(FM) using equations 14,15,16 and 17 respectively.

$$\text{MathchScore (i,j)} = \frac{|G_j \cap R_i|}{|G_j \cup R_i|} \quad (14)$$

Where G_j and R_i represents set of pixels labeled as j^{th} text line from original handwritten document and i^{th} text line after proposed method respectively.

$$\text{Detection Rate (DR)} = \frac{o2o}{N} \quad (15)$$

$$\text{Recognition Accuracy (RA)} = \frac{o2o}{M} \quad (16)$$

Where N and M represents number of text lines in original document and after proposed method respectively. o2o represents the number of one-to-one matches.

$$\text{FM} = \frac{2*DR*RA}{(DR+RA)} \quad (17)$$

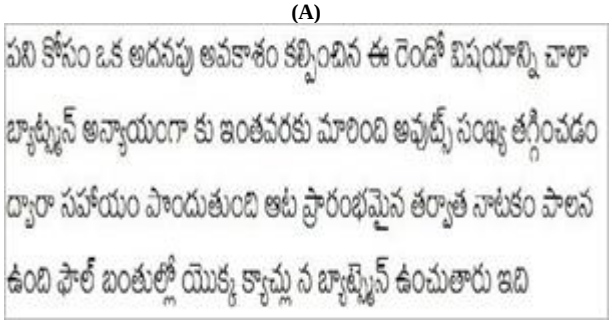
This paper considered the pair to be one-to-one match if Matchscore is greater than 0.95. Table 1 shows the detailed comparison of the results. The last row of Table 1 gives the average of all categories of document i.e. 1000 documents. Fig 9 shows comparison results with all categories.

Table 1: Comparison of the Proposed SRA Model on Three Categories

Dataset category	DA	RA	FM
Category one	99.63%	99.78%	99.70%
Category two	100%	100%	100%
Category three	98.48%	98.91%	98.69%
Average	99.37%	99.56%	99.46%

The important contribution of this paper is the proposed SRA method exhibited high accuracy not only in segmenting and straightening the handwritten Telugu documents. The proposed method is tested on Telugu printed document and on handwritten Indian languages like Tamil, Kannada, Malayalam etc. The straightening results of these handwritten and printed document are shown in Fig 7, 8, 9 and 10. Also the proposed SRA methods works well for different writing styles.

(A)



(B)

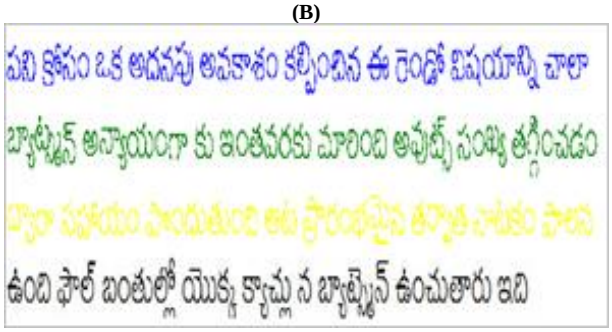
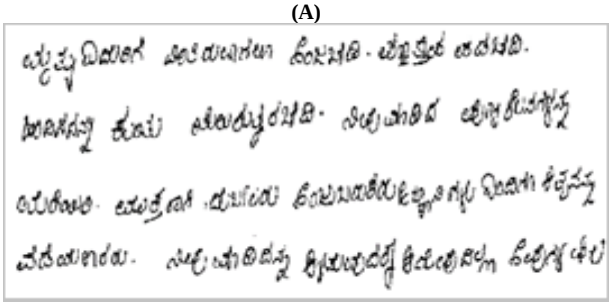


Fig. 7: (A) Represents Printed Telugu Document Image (B) Image after Proposed Method on (A). (Category 2).

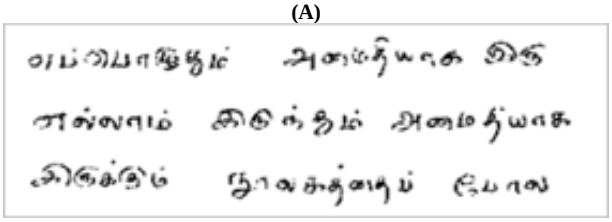
(A)



(B)

Fig. 8: (A) Represents Handwritten Kannada Document Image (B) Image after Proposed SRA Method on (A).

(A)



(B)

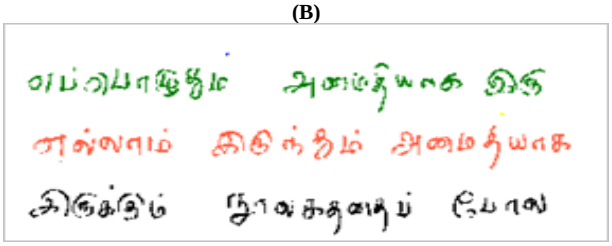
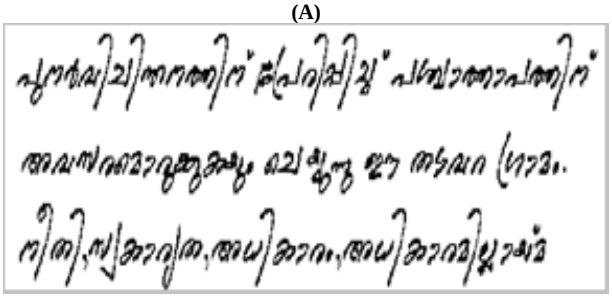


Fig. 9: (A) Represents Handwritten Tamil Document Image (B) Image after Proposed SRA Method on (A).

(A)



(B)

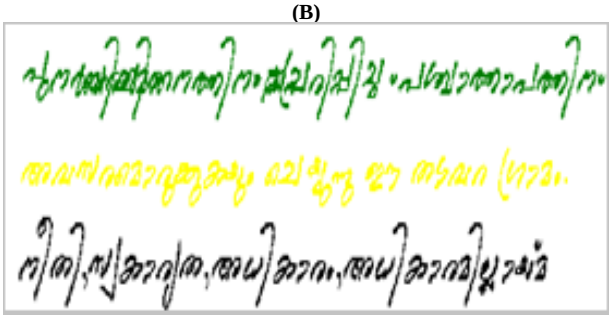


Fig. 10: (A) Represents Handwritten Malayalam Document Image (B) Image after Proposed SRA Method on (A).

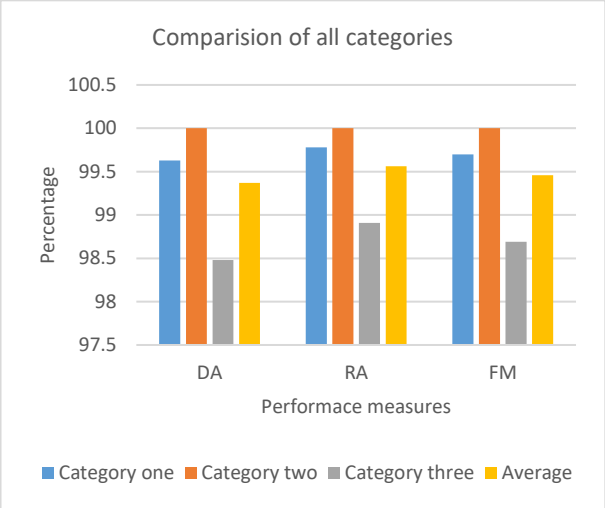


Fig. 11: Comparison of All Categories.

5.1. Discussion

From the experimental result we observed that the proposed method achieved 100% accuracy for document of category 2. The reason for this is uniform line spacing between lines and no concept of overlapping between adjacent text lines. Next we achieved good results in case of category 1 where the Telugu unconstrained handwritten document are collected from students, officers, house wife's etc. These documents does not contain uniform space between adjacent lines and overlapping between adjacent text-lines also possible in this category. The present paper also applied the proposed method on other languages text documents and achieved an accuracy rate of 98.91%. This accuracy rate is litter bit low when compared with type 1and 2 Telugu documents. This indicates that the proposed method also suitable for straightening handwritten text lines from other Indian languages.

5.2. Comparative analysis of different methods

The proposed SRA method is compared with existing methods using H. I. Koo et al [21] and Jewoong Ryu et al. [15] algorithms. The detailed comparison results are shown in the following Table 2 and Fig 12 and shows the efficiency of the proposed SRA model with all other methods. These results reveals that the proposed ZFD method is very suitable for Telugu handwritten documents.

Table 2: Comparison of the Proposed SRA Method with Existing Methods

Method	DA	RA	FM
H. I. Koo et al. [21]	95.01%	95.52%	95.26%
Jewoong Ryu et al. [15]	98.35%	98.41%	98.37%
proposed	99.37%	99.56%	99.46%

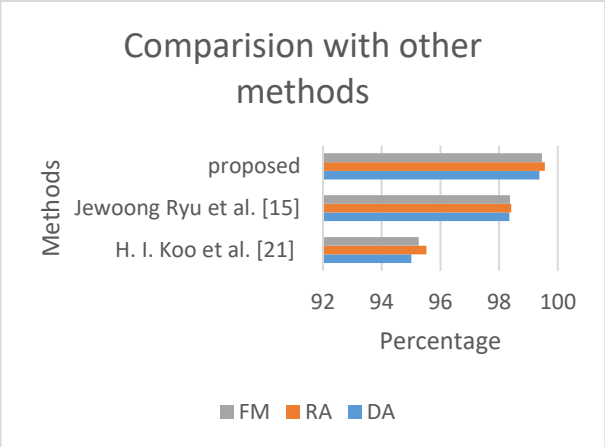


Fig. 12: Comparison with Existing Methods.

6. Conclusion

This Paper introduced a novel method for identifying and straightening of text lines in Handwritten Telugu documents. In this paper smearing and cubic polynomial regression technique is proposed for identifying or extracting text lines from Telugu handwritten documents. The robustness of our method is proved by conducting an experiment on 1000 samples with three categories datasets. The performance of proposed method evaluated using MatchScore table, Detection rate (DR), recognition accuracy (RA) and F-Measure (FM). The experimental analysis revealed the effectiveness of the proposed method. The proposed method can also be extended to other Indian languages like Kannada, Malayalam, Tamil, Hindi etc.

References

[1] L. Likforman-Sulem, Abderrazak Zahour and Bruno Taconet, "Text Line Segmentation of Historical Documents: A Survey", International Journal on Document Analysis and Recognition (2007) 9:123–138. https://doi.org/10.1007/s10032-006-0023-z.

[2] Zahour, A., Taconet, B., Mercy, P., Ramdane, S." Arabic hand-written text-line extraction", ICDAR '01 Proceedings of the Sixth International Conference on Document Analysis and Recognition, pp. 281–285, 2001.

[3] B. Yosef, N. Hagbi, K. Kedem, I. Dinstein, "Line Segmentation for Degraded Handwritten Historical Documents," Proc. 10th ICDAR, pp. 1161-1165, 2009. (4).

[4] Shi, Z., Govindaraju, V."Line separation for complex document images using fuzzy runlength" In: Proceedings of the International Workshop on Document Image Analysis for Libraries, Palo Alto, CA, USA, 23–24 January 2004 (50).

[5] Darko Brodi, "Text line segmentation with water flow algorithm based on power function", Journal of electrical engineering, VOL. 66, NO. 3, 2015, 132–141.(6).

[6] Rajath.A.N," An Adaptive Approach: Text Line Extraction from Multi-Skewed Hand Written Documents" IJCSET (www.ijcset.net). June 2015 Vol 5, Issue 6,158-161.(7).

[7] G. G. Rajput, Suryakant B. Ummapure, Preethi N. Patil ," Text-Line Extraction from Handwritten Document images using Histogram and Connected Component", International Journal of Computer Applications (0975 – 8887) National conference on Digital Image and Signal Processing, DISP 2015.(8)

[8] Dibyayan Chakraborty and Umapada Pala," Baseline Detection of Multi-lingual Unconstrained Handwritten Text Lines". Pattern Recognition Letters (2016), https://doi.org/10.1016/j.patrec.2016.02.003.

[9] Rahul Garg, Naresh Kumar Garg," Problems and Review of Line Segmentation of Handwritten Text Document", International Journal of Advanced Research in Computer Science and Software Engineering 4(4), April - 2014, pp. 1036-1039.(14).

[10] Satadal Saha, Subhadip Basu, Mita Nasipuri and Dipak Kr. Basu, "A Hough Transform based Technique for Text Segmentation", Journal of computing, Volume 2, Issue 2, February 2010, ISSN 2151-9617 (10 -15).

[11] Sunanda Dixit, sneha,Nilotap Utkalit and Suresh H.N ," Text Line Segmentation of Handwritten Documents in Hindi and English", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 4.(16).

[12] Amreen Singh and Er. Sukhpreet Singh," Line Segmentation of Handwritten Documents written in Gurumukhi Script", International Journal of Application or Innovation in Engineering & Management (IJAIEM). Volume 2, Issue 8, August 2013. (17).

[13] M.Ravi Kumar, Nayana N Shetty and B.P.Pragathi, " Text Line Segmentation of Handwritten Documents using Clustering Method based on Thresholding Approach", International Journal of Computer Applications (0975 – 8878) NCACC, April 2012.(18).

[14] Jayant Kumar Le Kang David Doermann Wael Abd-Almageed," Segmentation of Handwritten Textlines in Presence of Touching Components", 2011 International Conference on Document Analysis and Recognition (29).

[15] Jewoong Ryu, Hyung Il Koo and Nam Ik Cho," Language Independent Text-Line Extraction Algorithm for Handwritten Documents", IEEE Signal processing letters, VOL. 21, NO. 9, SEPTEMBER 2014 (15).

[16] G. Louloudisa B. Gatos, I. Pratikakis C. Halatsis,"Text line and word segmentation of handwritten documents", Pattern Recognition 42

(2009) 3169 – 3183. (23) https://doi.org/10.1016/j.patcog.2008.12.016.
[17] Zaidi Razak, Khansa Zulkiflee, Mohd Yamani Idna Idris, Emran Mohd Tamil, Mohd Noorzaily Mohamed Noor, Rosli Salleh," Off-line Handwriting Text Line Segmentation: A Review", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.7, July 2008.(24).
[18] Chethana H T and Mamatha H R," Comparative Study of Text Line Segmentation on Handwritten Kannada Documents", International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016, 26-33.(25).
[19] Abdollah Amirkhani-Shahraki , Amir Ebrahimi Ghahnavieh and Seyyed Abdollah Mirmahdavi," A Morphological Approach to Persian Handwritten Text Line Segmentation", 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.(26).
[20] Samir Malakar, Sougata Halder, Ram Sarkar, Nibaran Das, Subhadip Baus, Mita Nasipuri: "Text line extraction from handwritten document pages using spiral run length smearing algorithm". 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS). (27) https://doi.org/10.1109/CODIS.2012.6422278.
[21] H. I. Koo and N. I. Cho, "Text-line extraction in handwritten Chinese documents based on an energy minimization framework", IEEE Trans. Image Process., vol. 21, no. 3, pp. 1169–75, Mar. 2012.(21).
[22] Alireza Alaei, P. Nagabhushan and Umapada Pal," A New Text-line Alignment Approach Based on Piece-wise Painting Algorithm for Handwritten Documents", 2011 International Conference on Document Analysis and Recognition.(30).
[23] Rodolfo P. dos Santos, Gabriela S. Clemente, Tsang Ing Ren and George D.C. Calvalcanti: "Text Line Segmentation Based on Morphology and Histogram Projection", 2009 10th International Conference on Document Analysis and Recognition.(31) https://doi.org/10.1109/ICDAR.2009.183.
[24] G. Louloudis, B. Gatos, I. Pratikakisb and C. Halatsis," Text line detection in handwritten documents", Pattern Recognition 41 (2008) 3758 – 3772. (32) https://doi.org/10.1016/j.patcog.2008.05.011.
[25] Vijaya Kumar Koppula, and Atul Negi," Using Fringe Maps for Text Line Segmentation in Printed or Handwritten Document Images", 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems.(33).
[26] Bidyut B. Chaudhuri and Sumedha Bera," Handwritten Text Line Identification In Indian Scripts", 2009 10th International Conference on Document Analysis and Recognition. (34) https://doi.org/10.1109/ICDAR.2009.69.
[27] P. Nagabhushan and Alireza Alaei," Tracing and Straightening the Baseline in Handwritten Persian/Arabic Text-line: A New Approach Based on Painting-technique", International Journal on Computer Science and Engineering. Vol. 02, No. 04, 2010, 907-916.(35).
[28] Afaz Uddin Ahmed , Taufiq Mahmud Masum , Mohammad Mahbubur Rahman," Design of an Automated Secure Garage System Using License Plate Recognition Technique", I.J. Intelligent Systems and Applications, 2014, 02, 22-28 https://doi.org/10.5815/ijisa.2014.02.03.
[29] Laurence Likforman-Sulem and Claudie Faure," Extracting text lines in handwritten documents by perceptual grouping", Published in Advances in Handwriting and Drawing : a multidisciplinary approach,C. Faure, P. Keuss, G. Lorette, A. Winter (eds), pp. 21-38, Europia, Paris, 1994.(36).
[30] K.S. Sesh Kumar, A.M. Namboodiri, and C.V. Jawahar," Learning Segmentation of Documents with Complex Scripts", In the proceedings of ICVGIP 2006, LNCS 4338, pp. 749–760, 2006 (37).
[31] Mamatha H R and Srikantamurthy K," Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document", International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868. Volume 4– No.5,October 2012 (19).
[32] Mamatha Hosalli Ramappa and Srikantamurthy Krishnamurthy," Skew Detection, Correction and Segmentation of Handwritten Kannada Document", International Journal of Advanced Science and Technology. Vol. 48, November, 2012.(31).
[33] Saiprakash Palakollu , Renu Dhir and Rajneesh Rani," Handwritten Hindi Text Segmentation Techniques for Lines and characters", Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS 2012, October 24-26, 2012, San Francisco, USA.(28).
[34] Pal, U., Datta, S.," Segmentation of Bangla unconstrained handwritten text", In: Proceedings of Seventh International Conference on Document Analysis and Recognition, pp. 1128–1132, 2003.(3) https://doi.org/10.1109/ICDAR.2003.1227832.
[35] Vishwas H. S. and Bindu A.Thomas," Impact of Smearing Techniques on Text line Localization of Kannada Historical Scripts", International Journal of Computer Applications (0975 – 8887) (NCESCO 2015).(9).
[36] Payal Jindal and Balkrishan Jindal," Line and word segmentation of handwritten text documents written in Gurmukhi script using mid-point detection technique", International journal of advance research in science and engineering Vol. No. 4,Special issue(01), November 2015.(10).
[37] Dr.S.Pannirselvam , S.Ponmani," A Novel Hybrid Model For Tamil Handwritten Character Segmentation", International Journal of Scientific & Engineering Research, Volume 5, Issue 11, November-2014. ISSN 2229-5518.(12).
[38] Rahul Garg1 and Naresh Kumar Garg," An algorithm for Text Line Segmentation in Handwritten Skewed and Overlapped Devanagari Script",International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 5, May 2014.(13).
[39] Hashem Ghaleb, P. Nagabhushan and Umapada Pal," Graph Modeling based Segmentation of Handwritten Arabic Text into Constituent Sub-words", I.J. Image, Graphics and Signal Processing,2016, 12, 8-20, Published Online December 2016 in MECS (http://www.mecs-press.org/), DOI: 10.5815/ijigsp.2016.12.02.(47).
[40] Srinivasa Rao A.V,"Segmentation of Ancient Telugu Text Documents", I.J. Image, Graphics and Signal Processing,2012, 6, 8-14, Published Online July 2012 in MECS (https://doi.org/10.5815/ijigsp.2012.06.02.
[41] Sarbjit Kaur," An Automatic Number Plate Recognition System under Image Processing", I.J. Intelligent Systems and Applications, 2016, 3, 14-25, Published Online March 2016 in MECS (http://www.mecs-press.org/), DOI: 10.5815/ijisa.2016.03.02(49).
[42] Ali Benafia, Smaine Mazouzi and Benafia Sara," Handwritten Character Recognition on Focused on the Segmentation of Character Prototypes in Small Strips", I.J. Intelligent Systems and Applications, 2017, 12, 29-45, Published Online December 2017 in MECS (http://www.mecs-press.org/), https://doi.org/10.5815/ijisa.2017.12.04.
[43] MSLB. Subrahmanyam, V. Vijaya Kumar and B. Eswara Reddy," A new algorithm for skew detection of Telugu language document based on Principle-axis farthest pairs Quadrilateral (PFPQ)", I.J. Image, Graphics and Signal Processing,2018, 3, 47-58, Published Online March 2018 in MECS (https://doi.org/10.5815/ijigsp.2018.03.06.
[44] N. Shobha Rani and Vasudev T and Pradeep C.H," A Performance Efficient Technique for Recognition of Telugu Script Using Template Matching", I.J. Image, Graphics and Signal Processing,2016, 8, 15-23. https://doi.org/10.5815/ijigsp.2016.08.03.
[45] C. Vasantha Lakshmi, Ritu Jain, C. Patvardhan, "OCR of Printed Telugu Text with High Recognition Accuracies", ICVGIP 2006, pp. 786 – 795. (38).
[46] Vijaya Kumar Koppula , Atul Negi," Fringe Map Based Text Line Segmentation of Printed Telugu Document Images", Document Analysis and Recognition (ICDAR) ,2011 conference.(40).
[47] *Nobuyuki Otsu (1979), "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66.* doi:10.1109/TSMC.1979.4310076(41).
[48] Oztop, E., Mulayim, A.Y., Atalay, V., Yarman-Vural, F."Repulsive attractive network for baseline extraction on document images", Signal Process. 75, 1–10 (1999) (45). https://doi.org/10.1016/S0165-1684(98)00220-5.
[49] B. Gatos, N. Stamatopoulos, and G. Louloudis, "ICDAR 2009 handwriting segmentation contest," in Int. Conf. Document Analysis and Recognition (ICDAR), 2009, pp. 1393–1397. https://doi.org/10.1109/ICDAR.2009.245.
[50] N. Stamatopoulos, B. Gatos, G. Louloudis, U. Pal, and A. Alaei, "ICDAR 2013 handwriting segmentation contest," in Int. Conf. Document Analysis and Recognition (ICDAR), 2013, pp. 1402–1406. https://doi.org/10.1109/ICDAR.2013.283.