

Text line and word segmentation of Indian Script Handwritten Document

Varsha Hole

Computer Engineering Department
Smt. Indira Gandhi College of
Engineering
Koparkhairane, Navi Mumbai, India

Leena Ragha

Computer Engineering Department
Ramrao Adik Institute of Technology
Dr. D.Y. Patil Vidyanagar, Nerul,
Navi Mumbai, India

Pravin Hole

Computer Technology Department
Terna Polytechnic
Koparkhairane, Navi Mumbai, India

ABSTRACT

Based on the analysis of Indian script character shapes and literature survey, it presents a new sequence of line and word segmentation method to handle some of the deformations usually present in the handwritten document like touching components, overlapping components, skewed lines, words with individual skews etc. and build a proper text image with all these deformations removed. Line segmentation procedure is applied using Hough transform. The word segmentation is done with the computation of the distances of adjacent components in the text line image and classification of the previously computed distances as either inter-word gaps or inter-character gaps in a Gaussian mixture modeling framework. The proposed method of line segmentation is a sufficiently accurate to extract the text lines from unconstrained handwritten text documents. Word segmentation procedure also works well on different language scripts. Average result of word segmentation for complex Document on different language script is 76% and average result of word segmentation for good Document of different language script is 90%.

Categories and Subject Descriptors

I.4 [Image Processing and Computer Vision]: Enhancement-Gray scale manipulation, Segmentation-partitioning, Image Representation-morphological; I.7 [Document and text processing]: Document Capture-Optical character recognition OCR.

General Terms

Hough transform, connected components, touching components, overlapping components.

Keywords

Optical character recognition, Pre-processing, Global skew detection and correction, Line segmentation, Word segmentation

1. INTRODUCTION

With years of research in handwriting recognition, writer independent recognition of general handwritten text remains a

challenging task and the most predominant complexities are: skewed document and lines, curvilinear lines, fluctuating lines, touching and overlapping components, irregularity in geometrical properties of the line, such as line width, height, leftmost position, distance in between words and lines. Along with these complexities every script has its own complexities and so the same method may not be applicable to different scripts. In this paper, we focus on line and word segmentation of the document to build a properly aligned document and to test it for different Indian scripts. There is very little work reported for Indian script handwritten documents and most of the work done is found only for devnagari script. Based on the literature survey and the Indian script study, a generalized

technique of line and word segmentation which is applicable to most of the Indian scripts is proposed.

The paper is organized as follows: Section 2 is dedicated to the literature survey on text line segmentation and word segmentation of handwritten documents. In Section 3 a short introduction of Indian and Devnagari script is given. Section 4 presents the proposed line segmentation method. Section 5 describes the proposed word segmentation method. Section 6 describes results followed by conclusion.

2. LITERATURE SURVEY

Text line segmentation is a labeling process which consists of assigning the same label to spatially aligned units (such as pixels, connected components or characteristic points). Handwritten text line segmentation is still considered to be a major challenge in document image analysis. In a simple document analysis processing pipeline, it would follow image binarization and line segmentation, and precede word and character segmentation, character recognition etc. Since it is in the beginning of a pipeline of processing, it is very important to minimize errors so that next stages of pipeline get accurate input. Different authors have addressed the complexities in line segmentation using different methods. In [1], a method proposed by shredding text into lines and by shredding their surface with local minima tracers. An approach is based on the topological assumption that for each text line, there exists a path from one side of the image to the other that traverses only one text line. In [2], a new dual method of line identification that is based on inter dependency between text line and inter-line gap. In [4], methods that make use of the projection profiles include partitioning into vertical strips and horizontal run calculation as well as calculation of the projection profiles of every vertical strip (chunk) and traverse around any obstructing handwritten connected component (CC) by associating it to the text line above or below. In [7] and [9] authors have proposed methods that make use of the Hough transform. Hough transform is applied on the binarized edge map to generate the Hough image of it. For line segmentation, the parameters of the Hough transform, like delta Ro, delta Theta, start Theta, end Theta, connect Distance and pixels Count are initialized or tuned in such a way that the lines are extracted as a set of connected words.

Algorithms dealing with word segmentation in the literature are based primarily on analysis of geometric relationship of adjacent components. Components are either Connected component CCs or overlapped components (OCs). An OC is defined as a set of CCs whose projection profiles overlap in the vertical direction. Related work for the problem of word segmentation differs in two aspects. The first aspect is the way the distance of adjacent components is calculated. The second aspect concerns the approach used to classify the previously calculated distances as either between word gaps or within word gaps.

In [7], methods that make use of the Hough transform and the parameters of the Hough transform are initialized or tuned in such a way that the words are extracted as a set of connected characters. Each and every connected component is labeled using 4-connected neighbors' approach. In [9], word segmentation is addressed as a two class problem. The distances between adjacent overlapped components in a text line are calculated using the combination of two distance metrics and each of them is categorized either as inter or an intra-word distance in a Gaussian mixture modeling framework. In [6], a method based on contours of the word is proposed; contour of each word is examined along with threshold for inter-word gaps to extract words with high confidence. In [5], the approach for word segmentation claimed two methods: (i) The gap metric was computed by combining three different distance measures, which avoided the weakness of each of the individual one and thus provided a more reliable distance measure (ii) Besides the local features, such as the current gap, a new set of global features were also extracted to help the classifier make a better decision. In [3], a segmentation-based approach to handwritten Devnagari word recognition is proposed. On the basis of the head line, a word image is segmented into pseudo characters.

3. INTRODUCTION TO INDIAN SCRIPT

India has 18 official languages which include Assamese, Bangla, English, Gujarati, Hindi, Konkani, Kannada, Kashmiri, Malayalam, Marathi, Nepali, Oriya, Punjabi, Rajasthani, Sanakrit, Tamil, Telugu and Urdu. All the Indian languages do not have the unique scripts. Some of them use the same script. For example,

Some Indian scripts, like Devanagari, Bangla, Gurumukhi and Assamese have some common properties. Most of the characters have a horizontal lines at the upper part called *headline* and primarily the characters of words in these scripts are connected by a these headlines (shown in Fig. 2). Due to these properties they can be differentiable from the Roman (English), Telegu, Oriya, Urdu and other scripts. Furthermore, some characters have a part extended above the headline in these scripts. In these scripts headline connects all the characters of word together, the word segmentation can be analyzed easily. As a result the head-line features divide the total scripts into two sub-groups containing Bangla, Devanagari, Gurumukhi, scripts in one group and English, Tamil, Telugu, Oriya, Malayalam, Gujrathi, Urdu, Kannada in other group. The scripts in which concept of headline is not present, text is written by isolated characters placing side by side with or without some minimum distance between characters of words, if this distance is not maintained properly between characters and between words then it may result in some characters getting separated as words. So these characteristics of different scripts need to be analyzed with respect to methodologies chosen for word segmentation. Most of the Indian scripts have been originated from Brahmi script; hence there are lots of similarities between the different language scripts. Most of the Indian script follows the similar character formation and can be partitioned into three sub zones similar to Devnagari script. The upper zone usually has *matra* information. The middle zone is the main zone in with consonant information and bottom zone



languages such as Hindi, Marathi, Rajasthani, Sanskrit and Nepali are written using the Devnagari script. Bangla languages are written using the Bangla script, Urdu and Kashmiri are written using the same script and Telugu and Kannada use the same script. In all, ten different scripts are used to write these 18 languages. These scripts are named as Bangla, Devanagari, Roman (English), Gurumukhi, Gujarati, Malayalam, Oriya, Tamil, Kannada and Urdu. The images blocks of these images are shown in Fig.1. Indian scripts are different from Roman script in several ways. Indian scripts are two dimensional compositions of symbols: core characters in the middle strip, optional modifiers above and/or below core characters. Two characters may be in shadow of each other. While line segments (strokes) are the predominant features for English, most of the Indian language scripts are formed by curves, holes, and also strokes. In Indian language scripts, the concept, of upper case and lower-case characters is absent; however the alphabet itself contains more number of symbols than that of English.

Fig.1 Indian Script Languages

may have matra information or may be another consonant in case of complex characters. As the middle zone has maximum information, it therefore consider only middle zone for line detection. Fig.2. shows an example of a Devnagari word with different modifiers & 3 strips of word. The upper and lower zones of two consecutive text lines normally do not overlap or touch in case of printed script, but for handwriting, people have the tendency to write them bigger and with skews, that may lead to overlap and touch. Overall, these characteristics make handwritten Indian text recognition more challenging.

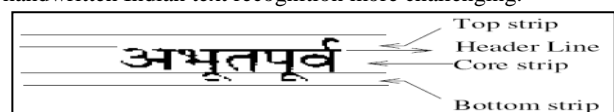


Fig. 2 Three strips of a Devnagari word.

4. TEXT LINE SEGMENTATION

The proposed methodology for text line segmentation in handwritten document images deals with the following challenges:

- (i) Each text line that appears in the document may have an arbitrary skew angle and converse skew angle along the text line,
- (ii) Text lines may have different skew directions,
- (iii) Accents may be cited either above or below the text line
- (iv) Parts of neighboring text lines may be connected. The text line segmentation methodology includes the following stages.

4.1 Connected component extraction

The connected components in a document image are identified by connected component analysis algorithm. Then the average character height AH for the whole document image is calculated based on the average height of all CCs. It assumes that the average character height equals to the average character width AW . The final step includes the partitioning of the CCs domain into three sub-domains which are denoted as "Subset1", "Subset2" and "Subset3". These sub-domains are treated in a different manner by the methodology [8]. "Subset 1" is expected to contain all components which correspond to the majority of the characters with size which satisfies the following constraints:

$$(0.5*AH \leq H < 3*AH) \text{ and } (0.5*AW \leq W) \quad (1)$$

Where H , W denotes the component's height and width, respectively, and AH , AW denotes the average character height and the average character width, respectively. "Subset 2" is expected to contain all large CCs. Large components are either capital letters or characters from adjacent text lines which touch each other. The height of these components is defined by the following equation:

$$H \geq 3 * AH \quad (2)$$

Finally, "Subset3" should contain characters like accents, punctuation marks and small characters. The equation that defines this set is:

$$((H < 3 * AH) \text{ and } (0.5*AW > W)) \text{ or } ((H < 0.5*AH) \text{ and } (0.5*AW < W)) \quad (3)$$

Fig.3. shows an example of partitioning CCs into 3 subsets denoted as "Subset 1", "Subset 2" and "Subset 3".

4.2 Hough transform mapping

In this stage, the Hough transform takes into consideration a subset (denoted as "subset 1" in Fig.4.) of the connected components of the image. This subset is chosen for the following reasons: (i) it is required to ensure that components which appear in more than one line will not vote in the Hough domain; (ii) Components, such as accents, which have a small size, must be rejected from this stage because they can cause false text line detection by connecting all the accents above the core text line. In our approach, instead of having only one representative point for every CC, a partitioning is applied for each CC lying in "Subset1", to equally sized blocks, so as to have more representative points voting in the Hough domain. An exception might be applied on the right most blocks. The width of each block is defined by the average character width AW . An example is shown in Fig.4 correspond to partitioning of CC's. After the creation of blocks, it calculates the gravity centre of the CC contained in each block. The set of all these

points contributes to the Hough transform. Every gravity centre in the subset corresponds to a set of cells in the accumulator array of the (ρ, θ) domain.

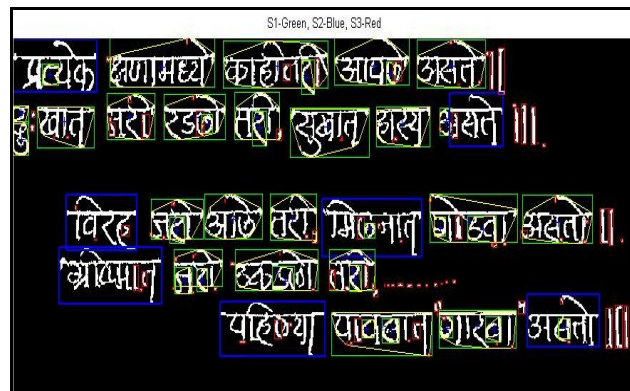


Fig.3 Example showing the connected components partitioned to 3 subsets denoted as "Subset 1" in green, "Subset 2" in blue and "Subset 3" in red.

To construct the Hough domain the resolution along θ direction was set to 1 degree letting θ take values in the range 85 to 95 degrees and the resolution along ρ direction was set to $0.2*AH$ [9]. It detects the cell (p_i, θ_i) having the maximum contribution and assign to the text line (p_i, θ_i) all points that vote in the area $(p_i - 5, \theta_i) \dots (p_i + 5, \theta_i)$. To decide whether a CC belongs to a text line, at least half of the points representing the corresponding blocks must be assigned to this area. After the assignment of a CC to a text line, all votes that correspond to this particular CC are removed from the Hough transform accumulator array. This procedure is repeated until cell (p_i, θ_i) having the maximum contribution contains less than $n1$ votes in order to avoid false alarms. During the evaluation of the procedure, the dominant skew angle of currently detected lines is calculated. In the case that the cell (p_i, θ_i) has a maximum contribution less than $n2$ ($n2 > n1$), an additional constraint is applied upon which, a text line is valid only if the corresponding skew angle of the line deviates from the dominant skew angle less than 2° [7]. Parameters $n1$ and $n2$ in the proposed text line segmentation methodology is experimentally defined in [9] as ($n1 = 5$, $n2 = 9$). Fig.5 shows the result of line segmentation after applying Hough transform.

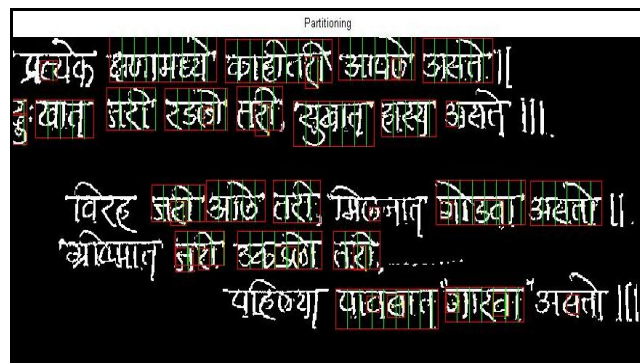


Fig.4 Example showing the partitioning of CC's to block of width AW and corresponding gravity centre in yellow. CC's without block are from "Subset 2" and "Subset 3" which are not considered for this stage

4.3 Post-processing

The post-processing stage consists of two steps. At the first step,

(i) A merging technique over the result of the Hough transform is applied to correct some false alarms and

(ii) CCs of “Subset1” that were not clustered to any line are examined to determine whether a new line is detected (see [8]). After the detection of the final set of lines, all components lying in “Subset3” as well as those unclassified components of “Subset1” become grouped to the closest line. The second step deals with large components lying in the sub domain “Subset2”. All components of this subset mainly belong to n detected text lines ($n > 1$).

As a next step the connected components of “Subset 1” that were not clustered to a line must be checked whether they create a new line that the Hough transform did not reveal. To this end, a grouping technique of the remaining connected components is applied that utilizes the gravity centers of the corresponding blocks (x_i, y_i) , we calculate the distance d_i . For every block with gravity centre (x_i, y_i) we calculate the distance d_i and the closest already detected text line. If d_i between ranges around the average distance of adjacent lines then the corresponding block is considered as a candidate to belong to a new text line. To decide whether a connected component is assigned to a new text line, at least half of the corresponding blocks must be candidates to belong to the new text line.

“Subset 2” includes the components whose height exceeds 3 times the average height (see Fig.5 (a)). These ‘large’ components may belong to more than one text line. This situation may appear when an ascender of one line meets a descender of an adjacent line. To include a connected component to a text line label, the number of lines that cross the bounding box of the connected component must be calculated. If more than one line crosses it, then this component is assigned to more than one text line otherwise it is grouped to the text line that crosses it.

“Subset 3” includes all the components that do not fall into the previous two categories. Components of “Subset 3” are usually punctuation marks or accents. As a final step all components belonging to this subset as well as the unclassified components of “Subset 1” are grouped to the closest line. In more detail, for any of these connected components, the distance from every line detected in the previous stages is calculated. This distance is the length of the vertical line that starts from the gravity centre of the connected component and finishes to the point that reaches the text line. Fig. 5(a) shows the CC’s in Subset 2 and 3 which need to be processed. Fig. 5(b) shows Final line segmentation after post processing stage.

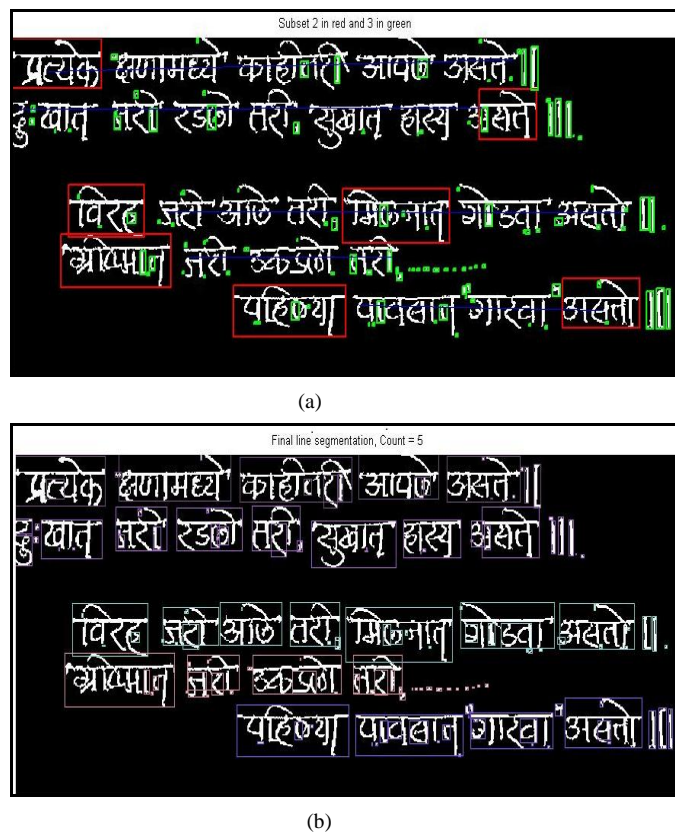


Fig.5 (a) “Subset 2” in Red and “Subset 3” in Green which were not considered in hough transform stage need to be processed in post processing stage (b) Final result of line segmentation after post processing stage with lines indicating in different colors.

5. WORD SEGMENTATION

The word segmentation procedure is divided into two steps. The first step deals with the computation of the distances of adjacent components in the text line image and the second step concerns the classification of the previously computed distances as either inter- word gaps or inter-character gaps. For the first step, it proposes the average of two different metrics: the Euclidean distance metric and the convex hull-based metric. The classification of the computed distances is performed using a well-known methodology from the area of unsupervised clustering techniques, the Gaussian mixtures [9].

5.1 Distance computation

In order to calculate the distance of adjacent components in the text line image, a pre-processing procedure is applied. The computation of the gap metric is considered not on the CCs but on the OCs, where an OC is defined as a set of CCs whose projection profiles overlap in the vertical direction. We define as distance of two adjacent OCs the average value of the Euclidean distance and the convex hull-based distance. The Euclidean distance between two adjacent OCs is defined as the minimum Euclidean distance among the Euclidean distances of all pairs of points of the two adjacent OCs. For the calculation of the Euclidean distance we apply a fast scheme that takes into consideration only a subset of the pixels of the left and right OCs instead of the whole number of black pixels. In order to define the subset of pixels of the left OC, we include in this subset the right most black pixel of every scan line. The subset of pixels for the right OC is defined by including the left most

black pixel of every scan line. Finally, the Euclidean distance of the two OCs is defined as the minimum of the Euclidean distances of all pairs of pixels. We calculate the convex hull-based metric as follows: Given a pair of adjacent OCs C_i and C_{i+1} , let H_i and H_{i+1} be their convex hulls, respectively. Let L be the line joining the centers of gravity (or centroid) of H_i and H_{i+1} . Let P_i and P_{i+1} be the points of intersection of L with the hulls H_i and H_{i+1} , respectively. The gap between the two convex hulls is defined as the Euclidean distance between the points P_i and P_{i+1} (see Fig. 6).

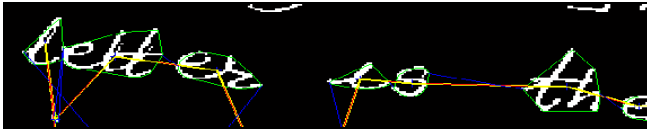
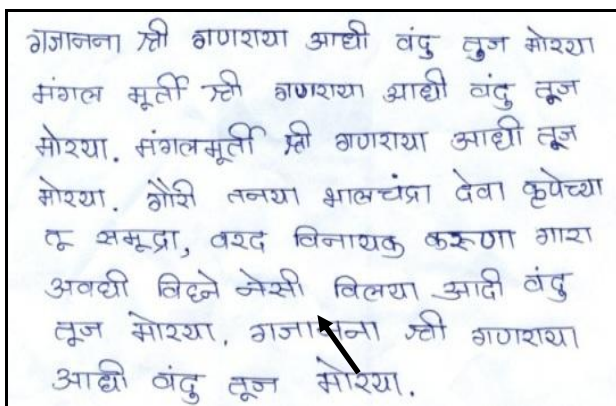


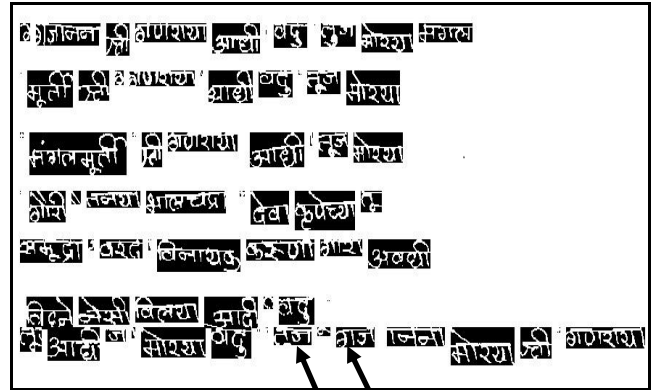
Fig.6 zoomed version of handwritten text line image. The convex hulls are defined with the green color. The yellow line determines the line segment that connects two gravity centers. Finally, the red line segment is the convex hull-based distance.

5.2 Gap classification

For the gap classification problem a novel approach is used. This approach is based on the unsupervised classification of the already computed distances into two distinct classes representing the word inter-class and the word intra-class, respectively. To this end, we adopt the use of Gaussian mixtures, a methodology which, to the best of our knowledge, was never used in previous works on word segmentation. A mixture model based clustering is based on the idea that each cluster is mathematically presented by a parametric distribution. We have a two clusters problem so every cluster is modeled with a Gaussian distribution. The algorithm that is used to calculate the parameters for the Gaussian s is the EM algorithm.



(a)



(b)

Fig.7 (a) Input document in Marathi Language (b) Result of word segmentation with saperate words. Arrow indicating oversegmentation (one word is split into two separate words) of word segmentation.

We use this methodology since the Gaussian mixture is a well known unsupervised clustering technique with many advantages which comprise:

- (i) the mixture model covers the data well,
- (ii) a density estimation for each cluster can be obtained
- (iii) a “soft” classification is available [9]

Fig 7. Shows the result of word segmentation with saperate words.

6. RESULTS AND ANALYSIS

We have collected many handwritten document pages from different people. These persons belong to both sexes and different age groups. The text documents contain different scripts such as Marathi, English, Gujarati, Tamil, Kannada, Bangla and Malayalam. The text contains samples of all basic characters and modifiers for the respective script alphabet. The approach has been tested to work well on writing with multiple sizes, variable skew and spacing as well as varied line length and line position, like the verses in poems. Hence it is quite versatile and comparable with other methods. The experimental results for line segmentation of the proposed method are encouraging and give very accurate line segmentation for different language scripts as input to further stages.

The word segmentation module takes as input the result of the proposed text line segmentation technique. Table I shows the result of word segmentation module applied on the documents with less complexities as shown in Fig.9. Average result shown of word segmentation for good documents for different Indian language scripts is 90%.

Table II shows the result of word segmentation module applied on complex documents shown in Fig.10. Average result shown of word segmentation for complex document for different Indian language scripts is 76%.

Fig.8 summarizes some problems that are encountered at the word segmentation procedure. The scripts like English, Tamil, Kannada, Malayalam, in which concept of headline is not present, text is written by isolated characters placing side by side with or without some minimum distance between characters of words, if this distance is not maintained properly between characters and between words then it may result in some characters getting separated as words. For such scripts inter and intra word gap calculations have to be performed for

word segmentation. Due to non uniform spacing between adjacent words there are cases that parts of adjacent words are merged (under segmentation) and cases where parts of the same word are split into two or more words (over segmentation).

The word segmentation of Devnagari Scripts like Marathi, Hindi, Sanskrit, in which concept of Shirorekha or headline is present, will not consider inter-word gap and intra-word gap calculations.

TABLE I

Result of word segmentation for good Documents of different languages

Script	Total No. Of words in each script	No. Of actually detected words	% of word detection
English	131	127	96.94 %
Marathi	47	45	95.74 %
Bangla	21	19	90.47 %
Kannada	27	24	88 %
Tamil	44	37	84.09 %
Malayalam	25	21	84 %

TABLE II

Result of word segmentation for complex Document of different languages

Script	Total No. Of words in each script	No. Of actually detected words	% of word detection
Tamil	25	22	88
Gujarati	48	38	79.16
Marathi	47	36	77
English	133	97	72.93
Marathi	29	18	62.6

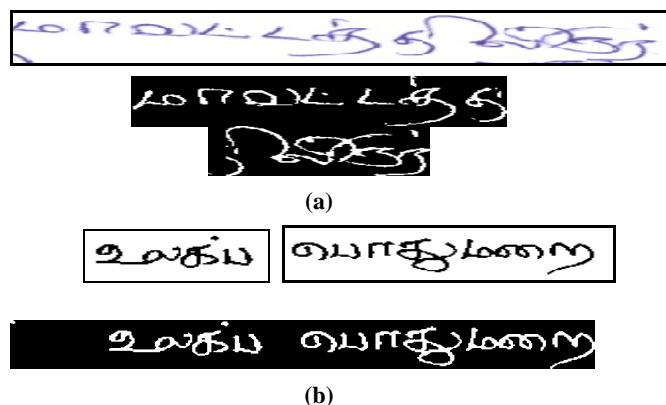


Fig.8 Typical examples of indicative errors of the proposed word segmentation methodology (a) A Tamil Word over segmented as two separate words (b) Malayalam Word merged as one single word of two separate words .

In these scripts headline connects all the characters of word together. So for the word segmentation it will simply consider the gap between two successive headlines or Shirorekha to separate the word. For such scripts it is a constrain that every character of word should be connected to headline, even all the

matra's of the word should be connected to shirorekha to have proper word segmentation results. If such constrained is not present then, matra's of the upper zone and lower zone can also be considered as separate CC's. Due to lack of the context of word in the image while trying to merge this small CC's to the corresponding word it is difficult to decide whether merging of the CC's is to be applied to the above or below word.

7. CONCLUSION

In this paper we present techniques for line and word segmentation of unconstrained handwritten document which are applicable to many Indian languages. A new text line detection method for unconstrained handwritten documents is presented. The main novelties of the proposed approach consist of (i) the partitioning of the connected component space into three subsets each treated in a different manner and (ii) the splitting of connected components into equally spaced blocks each of them voting in the Hough domain. The proposed method is a sufficiently accurate method to extract the text lines from unconstrained handwritten text documents.

A new word segmentation technique based on an efficient distinction of inters and intra-word gaps using the combination of two different distance metrics. The distance metrics that we use comprise the Euclidean distance metric and the convex hull-based metric. The distinction of the two classes is considered as an unsupervised clustering problem for which we make use of the Gaussian mixture theory in order to model the two classes. The word segmentation of Devnagari Scripts like Marathi, Hindi, Sanskrit, in which concept of Shirorekha or headline is present, will not consider inter-word gap and intra-word gap calculations. It will simply consider the gap between two successive headlines or Shirorekha to separate the word. Average result of word segmentation for complex Document on different language script is 76% and average result of word segmentation for good Document of different language script is 90%.

8. REFERENCES

- [1]A. Nicolaou, and B. Gatos "Handwritten Text Line Segmentation by Shredding Text into its Lines", 10th International Conference on Document Analysis and Recognition, IEEE Computer society, 2009, 626-630.
- [2] Bidyut B. Chaudhuri, Sumedha Bera, "Handwritten Text Line Identification In Indian Scripts", 10th International Conference on Document Analysis and Recognition, 2009,636-640.
DOI=
<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05277570>
- [3] Bikash Shaw,Swapan Kumar Parui, Malayappan Shridha, "A Segmentation Based Approach to Offline Handwritten Devanagari Word Recognition", International Conference on InformationTechnology,2008,256-257.DOI=
<http://doi.ieeeecomputersociety.org/10.1109/ICIT.2008.32>
- [4] Bruzzone, E., Coffetti, M.C. (1999), An algorithm for extracting cursive text lines., 1999. Proceedings of ICDAR '99, 20-22Sept.,749-752.DOI=
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=791896
- [5] C. Huang, S. Srihari, "Word segmentation of off-line handwritten documents", in: Proceedings of the Document Recognition and Retrieval (DRR) XV, IST/SPIE Annual Symposium, San Jose, CA, USA, January 2008.
- [6] Fajri Kurniawan , Amjad Rehman Khan, Dzulkifli Mohamad, " Contour vs Non-Contour based Word Segmentation from Handwritten Text Lines: an Experimental

Analysis" International Journal of Digital Content Technology and its Applications Volume3,Number2,June2009,127-131.DOI=

http://www.aicit.org/jdcta/ppl/jdcta_version10_Part17.pdf

[7] 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 2010,134-14.

[8] G. Louloudisa, B.Gatosb,I.Pratikakisb, C.Halatsisa, " Text line and word segmentation of handwritten documents", Pattern Recognition42,2009,3169-3183.DOI=

http://users.iit.demokritos.gr/~bgat/Louloud_1_2009.pdf

[9] J.M. Marin, K. Mengersen, C.P. Robert, Bayesian Modelling and Inference on Mixtures of Distributions, Handbook of Statistics, vol. 25, Elsevier-Sciences, Amsterdam, 2005.

(c)A document

ಅಜ್ಞಾನದ ಸಾಕ್ಷಿಗಳೆಂದರೆ
 ಸೇ ಸೇ ಸರ ಸರ - ಕಾಪನೆ ।
 ಜೋಡು ಕೂಡುಗಳು ಜೋಡು,
 ಜೋಡು ಜೋಡು ಜೋಡು ಜೋಡು,
 ಜೋಡು ಜೋಡು ಜೋಡು ಜೋಡು ಜೋಡು
 ಜೋಡು ಜೋಡು ಜೋಡು ಜೋಡು ಜೋಡು

in Bangla Script

गजानना स्त्री गणशया आही वंदु तुज मोशया
 संगल मूर्ति स्त्री गणशया आही वंदु तुज
 मोशया. संगलमूर्ति स्त्री गणशया आही तुज
 मोशया. गौरी तनया भालचंद्रा देवा कृपेच्या
 तू ममूद्रा, वरद विनायक करुणा गारा
 अवली विठ्ठे जेसी विलया आदी वंदु
 तुज मोशया, गजानना स्त्री गणशया
 आही वंदु तुज मोशया.

(a) A document in Marathi Script

ನನ್ನ ಹೆಸರು ಬಸವರಾಜ ವಾಚನ ಗುರುಪ್ರದ.
 ನಮ್ಮ ಉರು ಬಾಗಲಕೋಟೆ ಗುಡ್ಡು ಕೆಡು ಕನಕಬೆರೆ
 ರಾಜ್ಯವಲ್ಲು ಇಲ್ಲವು. ಬಾಗಲಕೋಟೆ ಸಾಂಪರವಾದ ನಗರವಾಚು
 ವಿಲ್ಲು ಕೆಡುಕನ ನಗರವ ನೆಲವುಗನ್ನೆ ಒಳಗೊಂಡಿರುತ್ತದೆ.
 ನಾನು ನನ್ನ ಬರಬರೆ ಬರಬರೆ ಬರಬರೆಯುವೆನು ಜೊಡೆಯರೊಡನೆ

(d)A document in Kannada Script

Software development process is an iterative process where different artifacts undergo cycles of draft → review → rework → baseline and further updation as and when essential. Requirement specification design, code files, project plan and test plan undergo continuous updation to suite the changes in related artifacts. It is necessary to track the changes in the work products and reference material so that the updated documents are available to people concerned. This also helps in labeling and building the work product to provide it to different users/customers as per their needs. Configuration management, also known as change control management, helps an organization in achieving the goal of controlling changes during life cycle. It is not necessary to use any automation tool, though many tools are available for achieving software configuration management.

(b) A document in English Script

ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು, ಮಹಾಶಯವು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು
 ಶ್ರೀಕೃಷ್ಣನು ಎಂಬುದುಂಟು ಶ್ರೀಕೃಷ್ಣನು ವಾಚನವು

(e)A document in Tamil Script

(f)A document in Malayalam Script

(c)A document in Marathi Script

(a)A document in Tamil Script

(d)A document in English Script

(b)A document in Gujarati Script