Markov Random Field Region Based Text Detection and Segmentation by Stroke Width Transform

Renuka
PG Student
Department of Computer Science and Engineering
PDA College of Engineering, Kalburgi

Dr. Sujata Terdal
Professor
Department of Computer Science and Engineering
PDA College of Engineering, Kalburgi

Abstract

Text detection in handwritten image has gained widespread interests. Detection of the texts from handwritten images is a challenging problem due to the multiple fonts, different sizes, various orientations and alignment, reflections, shadows, the complexity of image background. Text detection and segmentation from handwritten images are useful in many applications. We present a method called Markov Random Method for image operator that seeks to find the value of each image pixel, and demonstrate their use on the task of text detection in natural, which makes it fast and robust enough to eliminate the need for multi scale computation or scanning windows. A notable work, which is Markov Random Field method (MRF), has been attracting much interest due to its simplicity and efficiency. However, the Stroke Width Transform (SWT), and OCR has difficulty in situations like blur, low contrast, and illumination change, since it is highly relies on the outcome from the edge detector. Here region based approach MRF (Markov Random Field) with stroke width transform (SWT) method is proposed for automatic detection and extraction of text from handwritten images and explains the methodology to extract and recognize text. The applications of region based image segmentation by MRF for text detection from image has given the scope to us to include the important technologies like Text Information Extraction, Stroke Width Transformation etc. which will helps to improve the efficiency of work.

Keywords: Bounding box, discrete wavelet transform, Markov random field, Text localization and Stroke width transform

I. INTRODUCTION

Text detection on handwritten images has gained much interested in real world applications like assisting the visually impaired people, the tourist's navigation, and enhancing safe vehicle driving etc. the text based information has great interests and it contains lots of useful information which can be easily understood both by human and the computer, but analyzing of text information is difficult due to the variations of size, font, color and alignment. Detection of text in both indoor and outdoor environments it provides contextual clues for a wide variety of vision applications. And it has been shown that the performance of image detection algorithms depends on the performance of their text detection modules. Text localization and extraction of the background in different images is the main purpose of automatic text detection approaches. The text based search has been successfully applied in many applications and the robust and computational cost of feature matching algorithm is depends on other high-level features that are not efficient enough to be applied to large databases. For the complex background and high variations of font, size, and color, the text have to be robustly detected and one of notable works on the scene text detection is the Markov Random Field (MRF).The MRF is attracting and is based on its simplicity and efficiency. The simplicity can be seen from which the edge is used for each edge pixel, it traverses based on its gradient orientation until another pixel is encountered. Then, the path is saved by its length value of path in an image.

The main objective of the work is to develop a powerful and reliable tool for detecting text regions in an image, by using the Markov Random Field (MRF). The approach of MRF is grouping pixels together in a correct way, instead of looking for each separate pixel. By using the MRF we are able to relax the assumptions that are mentioned above, and maintain a high quality of results. Our goal is to implement and improve the algorithm which is defined and most of the text in the natural will be discovered with the little noise.the Stroke Width Transform (SWT), since it transforms the image data from containing color values per pixel to containing the most likely stroke width. The resulting system is able to detect text regardless of its scale, direction, font and language. When applied to images of natural scenes, the success rates of OCR drop drastically. There are several reasons for this. First, the majority of OCR engines are designed for scanned text and so depend on segmentation which correctly separates text from background pixels. While this is usually simple for scanned text, it is much harder in natural images. Second, handwritten images exhibit a wide range of imaging conditions, such as color noise, blur etc. Finally, while the page layout for traditional OCR is simple and structured, in handwritten images it is much harder, because there is far less text, and there exists less overall structure with high variability both in geometry and appearance.

II. RELATED WORK

Markov random field (MRF) based post processing has been applied to exploit the context in the already obtained results. This leads to significant improvement of the results. Our segmentation scheme makes use of purely image based features. Thus, it has the advantage of locating text independent of scripts, font, font-size, geometric transformation, geometric distortion, and background texture. Text line extraction or segmentation is an important problem that does not have an universal accepted solution in the context of automatic handwritten document recognition systems [1]. Text characteristics can vary in font, size, shape, style, orientation, alignment, texture, color, and contrast and background information. These variations turn the process of word detection complex and difficult [2]. In the case of handwritten manuscripts, differently from machine printed, the complexity of the problem even increases. Since handwritten text can vary greatly depending on the user skill, disposition and even cultural background. Here, we present a method to segment text lines based on morphology and histogram projection. Morphological operations[3] are used to produce a binary image. In the process of text line extraction from video images containing text information. In their application, precise box containing the region of the text is used as output of the system to identify machine printed text in different video contexts. We have adapted and improved this idea for handwritten text line segmentation problem. An important fact in relation to image analysis based on contrast is that this characteristic is robust in relation to changes in illumination and it is invariant to different image transformations such as scaling, translation and skewing. Once the page document has been preprocessed, a technique based on projection profiles is applied. Projection profiles are commonly used for printed document segmentation and can also be adapted to handwritten documents [4]. the projection curves[5] are used to segment music sheets in order to extract the basic symbols and their positions. The segmentation approach proposed is divided in 3 levels and utilizes projection profiles along the Y and X axes alternately. Projection profiles in the horizontal direction to segment words of historical handwritten documents [6] during the line segmentation stage. In this work, a projection profile in the horizontal direction is initially applied to obtain the text lines positions. Some improvements were necessary in this procedure to correctly identify the line segments, so a recovery process is also developed. A similar process is used to obtain the word borders of a line using projection profiles in the vertical direction. We refer to projection profile as histogram projection. Experiments are performed on handwritten documents randomly selected from the IAM-database [7], showing that the proposed technique produces encouraging results.

Instead of extracting all the contours of the text, we only use the outline of the text that is also called outer contour. It is because the inner contour sometimes might become a distraction in the later steps for summarizing the characteristics of samples. After filling the holes inside letters with the Sandwich method proposed in[8], a Canny edge detector is applied to extract edges.[9] first assigned a bounding box to the boundary of each candidate character in the edge image and then detected text characters based on the boundary model (i.e., no more than two inner holes in each bounding box of alphabets and letters in English).A group of filters to analyze texture features in each block and joint texture distributions between adjacent blocks by using conditional random field. One limitation of these algorithms is that they used noncontent-based image partition[10] to divide the image spatially into blocks of equal size before grouping is performed. Noncontent- based image partition is very likely to break up text characters or text strings into fragments which fail to satisfy the texture constraints. The input image is decomposed into multiple foreground images. Individual foreground images go through the same processing steps, so the connected component analysis and text identification modules can be implemented in parallel on a multiprocessor system to increase the processing speed. Finally, the outputs of all the channels are fused to locate the text in the input image. Text location is represented as the coordinates of the bounding box surrounding the text. Details of our algorithm are provided in [11]. Detecting, segmenting, and recognizing text in images which are part of web pages is also a very important issue, since more and more web pages present text in images. Existing text-segmentation and text-recognition algorithms cannot extract the text. Thus, all existing search engines cannot index the content of image-rich web pages properly [12]. Automatic text-segmentation and text-recognition also helps in automatic conversion of web pages designed for large monitors to small liquid crystal displays (LCDs) of appliances, since the textual content in images can be retrieved. Two simple methods to locate text in complex images [13]. The first approach is mainly based on finding connected monochrome color regions of a certain size, while the second locates text based on its specific spatial variance. Both approaches are combined into a single hybrid approach. Since their methods were designed primarily to locate text in scanned color CD cover images, they are not directly applicable to video frames. Usually, the signal-to-noise ratio (SNR) is much higher in scanned images, while its low value in videos is one of the biggest challenges for text segmentation. A text region detector is designed to estimate the text existing confidence and scale information in image pyramid, which help segment candidate text components by local binarization to efficiently filter out the non-text components, a conditional random field (CRF) model considering unary component properties and binary contextual component relationships with supervised parameter learning is proposed. Finally, text components are grouped into text lines/words with a learning-based energy minimization method[14]. A clustering-based technique[15] has been devised for estimating globally matched wavelet filters using a collection of ground truth images and extended text extraction scheme for the segmentation of document images into text, background, and picture components. Multiple, two-class Fisher classifiers have been used. An approach to robustly detect and localize texts in natural scene images.

III. PROBLEM STATEMENT

Detection of text in a natural scene image is an important part in number of Computer vision applications. Such as the performance of optical character recognition (OCR) algorithms can be improved by first identifying the regions of text in the image. Text detection in natural is highly researched and developed field. There are various approaches for solving this type of problem. The

most text detection schemes are restricted for the particular languages, scale and direction of the text. A tradeoff between the numbers of restrictions we can apply and get the quality of the result. And we limit our search the less noise is encounter, hence with aid of the Markov Random Field (MRF) method for image text recognition. It is possible to overcome the problems mentioned and get result accurately.

IV. PROPOSED METHOD

In the existing method drawbacks like non-uniform texture of characters, edge distorted due to shape un-uniform etc. allow us to proposed new way of text extraction by using the image segmentation by MRF (Markov Random Field) for region of interest to locate extract the textual part form the images by using Stroke Width Transform method.

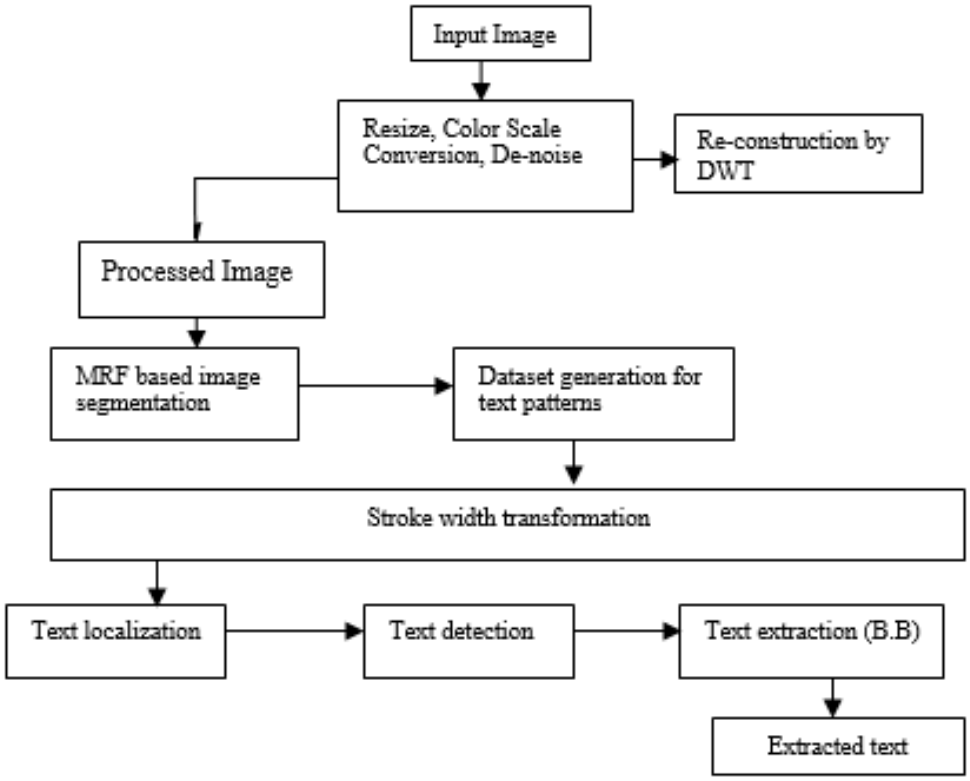

Fig. 1: Proposed System Block Diagram

In the above proposed system block diagram (Fig.1) the proposed system includes the three major phases they are Reconstruction of the image, MRF based image segmentation and SWT based text detection. These phases are further processed in the work to get the desired user result.

V. METHODOLOGY

There are two main methods they are:

- Text Detection
- Text Localization and Segmentation

A. Text Detection

In the text detection, there is no prior information on the input image that contains any text that is existence or nonexistence of text in image that must be determined in this step. The several approaches are used for the certain types of video frame or contain text. In the Detection of text there are number of processes are used that are Image Acquisition for text detection, Preprocessing, Discrete wavelet Transformation (DWT) and the Image Enhancement.

1) Image Acquisition:

In the Image acquisition process, the recognition system acquires a scanned image as an input image. The input image should have a particular format such as BMP or JPEG etc. This image is acquired through a digital camera or any other suitable digital input devices.

2) Pre Processing:

The pre- processing is a series of operations that are performed on the scanned input image. Which is essentially enhances the image rendering and it is suitable for segmentation. The main role of pre-processing is to segment the interesting pattern from the

background. Generally, for the filtering the noise, smoothing and normalization should be done in this step. The pre-processing step also defines a compact representation of the pattern.

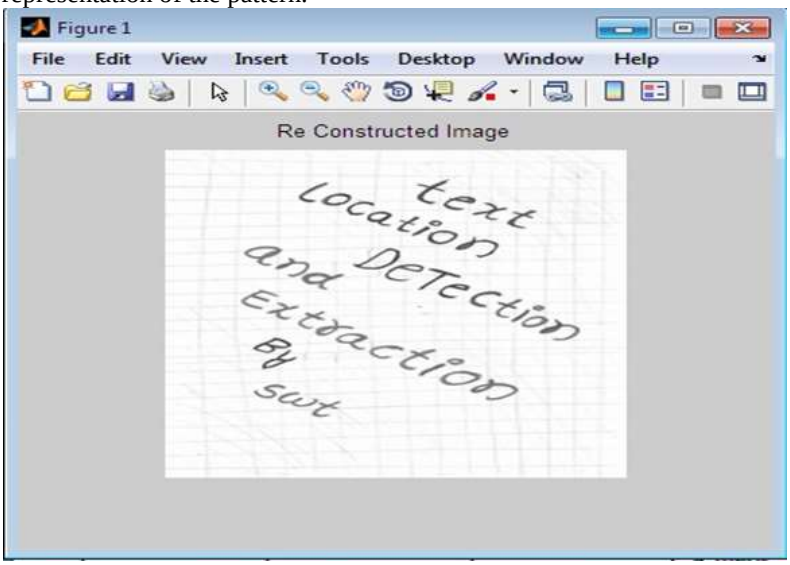

Fig. 2: The pre-processed image

3) Image Enhancement:

Image Enhancement is one of the most important and difficult techniques in image research. The image enhancement is used to improve the visual appearance of an image, and provide a better transform representation for image processing. It is very necessary to enhance the contrast and remove the noise to increase quality of the image enhancement technique is different from one field to another field according to its objective.

Image enhancement includes the image resizing, color space transformation and de-noising of the image. The de-noising of the image is done by using the DWT (Discrete Wavelet transformation) method with the use of weiner filter.

B. Text Localization and Segmentation

The text localization method can be divided into two types that are region based and text based. It deals with text localization in the compressed domain. The method contains two approaches is difficult to categorize. The performance measures are presented for each approach that is based on experimental results when it is available. Using the Localization and Segmentation of text the region based method and the Markov Random Field method is used with the thresholding, text Information Extraction and Tracking and Extraction of text.

1) Region based methods:

Region-based methods are used for the properties of color or gray scale in a text region or their differences with the corresponding properties of the background. These methods can be divided into two sub parts that is connected component based and edge based

2) Markov Random Field (MRF) Method:

Markov Random Field method is used for many computer vision applications. The main probability of the data being observed is inconsequential. The probability distributions on labels y and an image x have modeled equally, the probability of image being ignored at the classification time. The Markov random fields are probability distributions are parameterized by a graph G = (V, E). The typical generated random fields are treating the interaction between local data and its labels and its neighboring labels.

3) Text Information Extraction:

The text tracking, extraction, and enhancement methods are used for importance of verification, enhancement, speedup, etc. To enhance the system performance it is necessary to make changes in sequence of frames. The text tracking process can be serve to verify the result of text localization, if text tracking process performed in a short time than the text detection and localization this would speed up the system In some cases where text is located in different frames, text tracking can help to recover the original image.

VI. RESULT AND DISCUSSION

A. Recognition of the Text in the Image

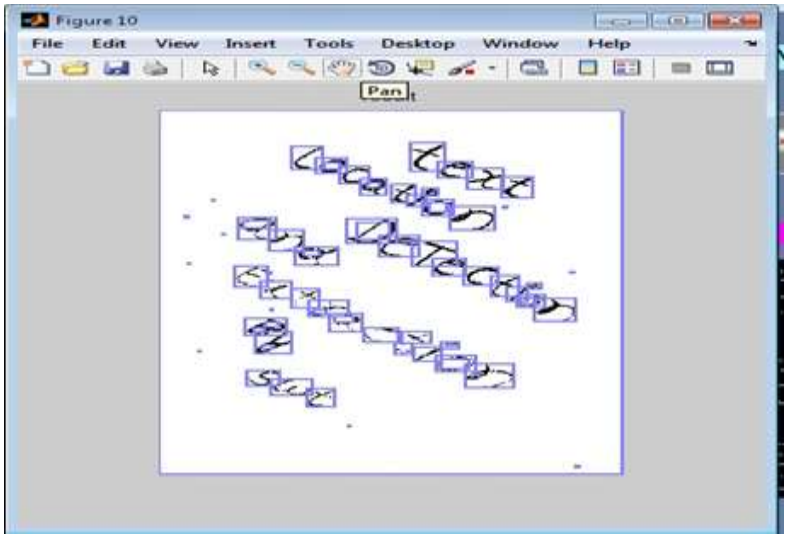

Fig. 3: The Recognized text

Result is accepted to be achieved if the applied strategy gives the expected output at each stage. For simplicity of understanding the output of the acquired are obtained at each major process. The project has been programmed to properly extract the text present in the frame that is acquired and recognize it. The proposed solution, based on the standard method gives the recognized text as its final solution.

The experimental result has shown the user the recognized text from the input image given. The image is transformed during the system processing where it goes through the umber of the phases as image reconstruction, image segmentation by MRF, text pattern extraction etc as mentioned in the above segments.

VII. ACCURACY CALCULATION

The accuracy is calculated using the formula

$$\text{Percentage accuracy (\%)} = \frac{\text{Correctly Recognized Image Samples}}{\text{Total Number of Test Image Samples}} * 100.$$

=43/50*100
=86%

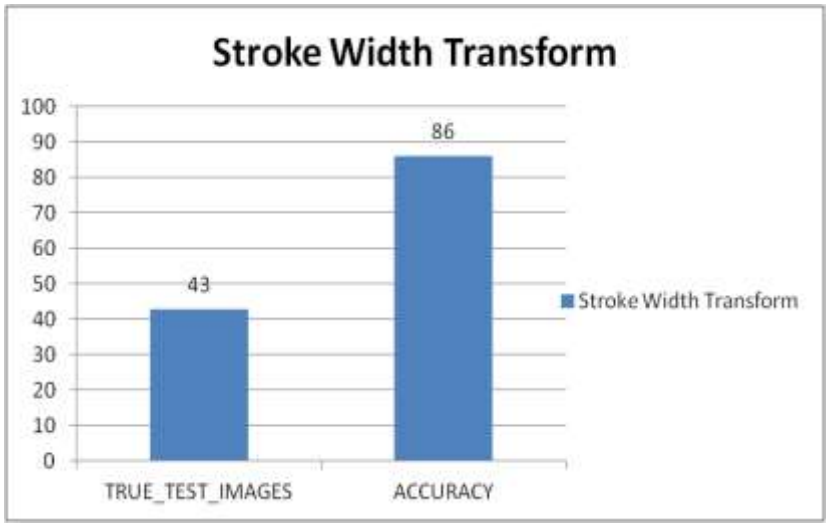

Fig. 4: Accuracy result of Stroke Width Transform

The accuracy of the system is measured by using the number of input images versus the accurate results. The experiment has been conducted on 50 image data set which has given the efficient result of 86% in respect to the dataset.

VIII. CONCLUSION AND FUTURE ENHANCEMENT

We firstly present a novel work after attempting number of iterations, we have come across the problems associated with text extraction and we have used effective features and methods to overcome the majority of problems .the proposed work has been implemented by using Matlab-2013 development tool which has given accurate result.

The applications of region based image segmentation by MRF for text detection from image has given the scope to us to include the important technologies like Text Information Extraction, Stroke Width Transformation etc. which will helps to improve the efficiency of work.

The experiment has been conducted in the collected dataset and given the accurate results. The future work should be acquiring the image by using the live digital image collector board which has the machine interfacing with the system.

REFERENCES

[1] L. Likforman-Sulem, A. Zahour, B. Taconet, "Text line segmentation of historical documents: a survey", International Journal on Document Analysis andRecognition, 2007, pp. 123-138.
[2] K. Junga, K.I. Kimb, A.K. Jain, "Text information extraction in images and video: a survey", Pattern Recognition, 2004, pp. 977-997.
[3] J.C. Wu, J.W. Hsieh, Y.S. Chen, "Morphology-based textline extraction", Machine Vision and Applications, 2008, pp. 195-207.
[4] S. Marinai, P. Nesi, "Projection Based Segmentation of Musical Sheets", Document Analysis and Recognition, ICDAR 1999, pp. 515-518.
[5] R. Manmatha, J.L., Rothfeder, "A scale space approach for automatically segmenting words from historical handwritten documents", IEEE Trans. Pattern Anal. Mach. Intell., 2005, pp. 1212-1225.
[6] U.V. Marti, H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition", International Journal on Document Analysis and Recognition, 2002, pp. 39-46.
[7] B. Gatos, A. Antonacopoulos, N. Stamatopoulos, "Handwriting Segmentation Contest", Document Analysis and Recognition, ICDAR 2007, pp. 1284-1288.
[8] C. L. Tan, S. Lu, and L. Li, "Document image retrieval through word shape coding," Pattern Analysis and Machine Intelligence,vol. 30, pp. 1913–1918, 200.
[9] T. Kasar, J. Kumar, and A. G. Ramakrishnan, "Font and background color independent text binarization," in Proc. 2nd Int. WorkshopCamera-Based Document Anal. Recognit., 2007, pp. 3–9.
[10] J. Weinman, A. Hanson, and A. McCallum, "Sign detection in natural images with conditional random fields," in Proc. IEEE Int. WorkshopMach. Learning Signal Process., 2004, pp. 549–558.
[11] A. K. Jain and B. Yu. Automatic text location in images and video frames. Technical Report MSUCPS: TR97-33, Dept.of Computer Science, Michigan State University, 1997.
[12] D. Lopresti and J. Zhou, "Locating and recognizing text in WWW images,"Info. Retrieval, vol. 2, pp. 177–206, May 2000.
[13] Y. Zhong, K. Karu, and A. K. Jain, "Locating text in complex color images," Pattern Recognit., vol. 28, pp. 1523–1535, Oct. 1995.
[14] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," Pattern Recogn., vol. 37, no. 5, pp. 977 997, 2004.
[15] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and mrf model," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2117–2128, Aug. 2007.