Scene Text Detection & Language Translation Using SWT and EBMT

[1]S. Khandait, [2] P. Khandait and [3] P. Jambhulkar

[1,3] Head of Information Technology Dept., K.D.K. College of Engineering Nagpur, Maharashtra, India
[2]Head of Electronics Engineering Dept., K.D.K. College of Engineering Nagpur, Maharashtra, India
{prapti_khandait@yahoo.co.in prabhakark117@gmail.com jambhulkar.pravin29@gmail.com}

Abstract: In recent years, the popularity of portable devices for capturing images, video, text extraction, etc. become key problems to detect and recognize the text in images. Extraction of text information from images or scene involves binary, text detection, text localization, word segmentation, and enhancement and character recognition. But there are variations involved in text such as font style, font size, orientation, alignment, reflections and illumination effect, hybrid background and low contrast image make text extraction process too difficult and more challenging. A large number of approaches and methods have been proposed to address this problem but still none of them are perfect. This paper presents an effective approach where we come up with hybrid approach that combines SWT with OCR for text detection and recognition in an image and EBMT to translate detected text into Hindi language. General challenges for performing scene text detection are also discussed.

Keywords: ***Word segmentation, text extraction, text detection, character recognition, localization.***

1 Introduction

Detecting and recognizing texts in images provides appropriate clues for a wide variety of vision related tasks and translation of detected text to target language helps for culture conversion. A text in images specially road side images or place contains more information related to the place and helps and direct us to understand and the objective more easily. The rapid growth in gadgets equipped with megapixel cameras and invention of touch screen digital devices like PDA, mobile, etc. increases the need and demand of information retrieval. Text Detection segmentation of text from natural scene images are useful in many applications. Text detection is quite easier but recognizing text out of detecting text area is a challenging problem due to the variety of colors, existence of complex backgrounds, fonts, and the length of the text strings. Text data in images contain useful information for automatic annotation of images, indexing, and structuring. The Extraction process of such type of information involves detection of text region in a giver input image, next is localization used to determine the location of text in the input image & generating bounding boxes around the text, tracking is used to reduce the processing time for localization, enhancement is to magnify small text at a higher resolution, and lastly recognition in which extracted text image will be transformed into text of the editable text from a given input image. However there are certain condition that makes the difficulty in automatic text extraction really challenging [9] like variations of text due to differences in alignment, size and orientation, style, as well as complex background and low image contrast.

The contents of an image could be more clearly understandable if the text is quoted in it. This extracted text in images is widely used by many applications today. Such applications like cheque clearing house and banking system in which physical transfer of cheque is replaced by cheque truncation system in which only an image of cheque is send to system and the text extracted from that cheque is used to clear the cheque. The success of all such type of application totally depends on a Textual Information Extraction system in which how proficiently the system detects, localize and recognize the text information present in an image.

Textual Information Extraction system mainly performed in two phases: the first in which text detection and localization is performed and in the second phase the regions of detected text are specified to the OCR which in turn recognizes the character present in the image [10].

B. Iyer, S. Nalbalwar and R. Pawade (Eds.)
ICCASP/ICMMD-2016. Advances in Intelligent Systems Research.
Vol. 137, Pp. 547-555.
© 2017- The authors. Published by Atlantis Press
This is an open access article under the CC BY-NC license (http://creativecommons.org/licens)es/by-nc/4.)

There are numerous kind of images comes in our daily life that are very flexible to handle and the problem arises in text detection and recognition is that images can have different aliasing and shadowing, blur and uneven lighting in images, different font and size of character in text.
The proposed approach should handle such challenges and must be independent of all problem faced in text detection and recognition previously [11].

2 Related Work

Scene text detection algorithm broadly classified into two types: Region-based and Connected Component (CC)-based approaches. In Region-based methods a sliding window scheme is adopted, which is basically a brute force approach which takes a lots of trial and local decisions. Therefore, the region-based methods primarily concerned about an efficient binary classification (text and non text) in an images. Text Localization is of fundamental importance in image understanding and content based retrieval. Before applying Optical Character Recognition (OCR) the localization must always be achieved. The second approach is about localizing the individual characters using the local parameters of an image like intensity, stroke-width, color, gradient *etc*. Feature extraction plays an important role in image localization process. The aim of feature extraction is to achieve maximize accuracy rate with minimum number of elements used in it.
A scene text detection algorithm proposed by Hyung Il Koo and Duck Hoon [3] follows two machine learning classifiers: candidate word regions is created by first classifier and other classifier is used to sort out non-text regions present in image. Maximally stable external region algorithm is used to extract Connected components (CCs) in an images and then clusters are formed using these extracted CCs so that candidate regions are generated. Normalization are performed on these candidate word regions and determined that whether every region contains text or not. A text or non-text classifier for normalized images is created because the skew, font, orientation, scale and color of every candidate can be expected from Connected Components. The accuracy of this classifier is depend upon multilayer perceptions and using a single free parameter recall and precision values can be controlled.
A novel image operator proposed by Boris Epshtein , Eyal Ofek, Yonatan Wexler [4], is used to obtain the value of stroke width for each pixel of image, and exhibit its use in text detection for natural images. The proposed image operator is considered to be local and data dependency makes it efficient, fast and robust to reduce the need for multi-scale computation. The letters are grouped together and its accuracy can be enhanced by assuming the directions for the improved strokes and the algorithm can detect curved text lines as well.
X. Chen, J. Yang, J. Zhang, A. Waibel [5], used a hybrid approach in which multiresolution and multiscale edge detection, color analysis, adaptive searching and affine rectification in a hierarchical manner which is used for sign detection having dissimilar priority at every stage to handle different orientations of text in images, font sizes, color distributions and complex backgrounds. Proposed Methodology used affine rectification that will improve text regions deformation that may caused by an improper camera angle. It extracts features from an image directly instead of using binary information for OCR. The proposed methodology have local intensity normalization method to handle the problem of variations in lighting followed by a Gabor transform that will find local features and then a linear discriminate analysis (LDA) method is applied for feature selection. The approach is utilized for developing a Chinese sign system, which can detect and recognize Chinese signs and then translate the recognized text into English.
K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon [6], developed a technique that will handle the text-localization problem using a CC-based approach which first detect character strokes followed by a threshold and stroke-width that are subsequently used for character segmentation. The accuracy and sensitivity of the detection algorithm is evaluated against four parameter: character recall, word recall, stroke precision, and computing time. The detection algorithm not performed well on italic fonts or when characters of a word have slight different orientation. The possible ways to improve the performance of the system is by directly working on color space in image to detect character strokes.
Y. Pan, X. Hou, and C. Liu [7], come up with a hybrid approach that integrates region information with a robust CC-based method. The parameters of a conditional random field (CRF) model are optimized by supervised Learning and the binary contextual component relationships with the unary component properties are incorporated in the CRF model. The proposed hybrid approach can be further improved because this approach fails on some texts that are difficult to segment. The speed of the proposed hybrid approach needs to be accelerated further. Text recognition should be included with text localization to complete the need of text information extraction as well.
Connected Component (CC) based methodology for text detection in natural scene images is proposed by Yao Li and Huchuan Lu [8]. A potential text region is first captured by utilizing a MSERs and skeleton is then

applied on it to extract stroke width. The robust CC approach groups the characters into separated words and also eliminates false positives values at the same time.

Ravina Mithe, et al. [9], also proposed the hybrid approach of the functionality of Optical Character Recognition (OCR) with speech synthesizer. In this paper, the proposed technique used character recognition method presented by OCR technology and use android phone having higher quality camera. The Optical Character Recognition is used for the recognition of characters present in images. Reliably extracting text from real-world captured photos is a challenging task due to variations in environmental, camera view angle, font size.

3 Proposed Methodology

In this section, we have presented a method for text detection and recognition in an image and translation of detected text into other language. Our project aim is to detect text region in an image by using Stroke Width Transform (SWT) and to recognize text in an image by using Optical Character Recognition (OCR).

A. *STROKE WIDTH TRANSFORM (SWT)*

In the image, a stroke is form by a continuous band of a nearly related and constant width. A RGB image is taken as input by SWT and gives an output image of the same shape and size in which the portions that have text are marked and underlined to distinguish it with background. This SWT algorithm has 3 main steps: the stroke width transform, based on the stroke width in image it groups the pixels having same width into letter candidates, and then group letter candidates into a complete regions of text. Basically he SWT Text Detection algorithm is used to locate and mark the regions of text present in an image. SWT provides the features of the Stroke Width Transform, that will subsequently capable to detect text present in an image regardless of its orientation, font size, background, scale, direction and language.

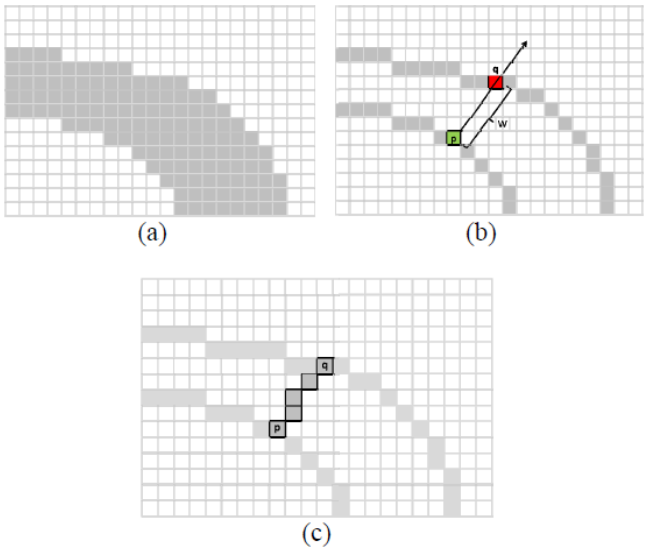

Figure 1: Stroke Width Transform [22]

B. Optical Character Recognition (OCR):

Optical Character Recognition technique is used to recognize the text present in an image. TO achieve the greater accuracy in OCR orientation of text, words, the text lines, symbols in an image should be segmented properly. The accuracy and sensitivity of text recognition is directly affected by how properly and correctly the text are detected and segmented. Text detection composes of line segmentation and word segmentation which are most important part of OCR systems. Segmentation of text line in an image is achieved by SWT which is major and critical component of an Optical Character Recognition (OCR) system. In general, line segmentation

and word segmentation is collectively called as text segmentation. Text segmentation is the process in which a text and character is isolated from the background of the image.
Segmentation is first performed on image that will first line the text present in image further it will segment individual word of text and lastly segment the character present in each word; and this all is achieved in our work by SWT. Texts in scene images contain important information and thus isolation of text strings is an important issue. Most of the business card images contains complex graphic backgrounds. SO in such cases to identify text separation of text region from background plays an important role to achieve higher accuracy. OCR [10] technique enables us to extract the textual data from an image and convert it into an editable form as it further taken as input by EBMT. It has a wide application as bank cheque processing by truncation system, map interpretation, form processing, engineering drawing interpretation and postal address sorting, extraction of text is important.

C. Example Based Machine Translation

EBMT used a translation technique by analogy. That means it work very fine by a previously successful translation of sentences and their respective translations in the target language, EBMT use such example to translate new sentences with same grammar and structure of test and train sentence. It simply mean if a previously successful translated sentence occurs again, the same translation need not to check again it simply match and gives output.
The basic principal of EMBT is that, it has a database that contains previously translated sentences and it simply matches the sentences; if not found it matches phrases again not found then it go for word by word. So bigger the database greater is the accuracy.

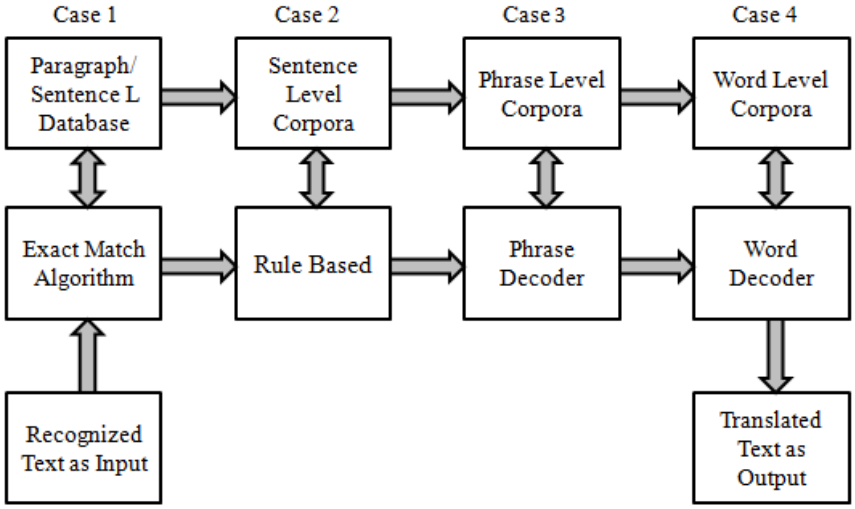

Figure 2: Block Diagram of EBMT

The EBMT system divided into four different modules.

Case 1: Exact Match Algorithm

In this process, the test sentence (i.e. English language) is first matched with each and every sentence in the available bilingual corpora for a perfect match. If it is found, the respective target sentence (i.e. Hindi sentence) is generated and displayed as output.

In this case when the input text is a paragraph, then firstly the input is broken into sentences, and each sentence is taken one after another and translated into targeted language.

Case 2: Rule Based Translation

Every sentence is organized in a grammatical way as all languages follow grammar and have dictionary as well. Here we are considering translation of English to Hindi language. English language follows the Subject-Verb-

Object topology while Hindi language follows Subject-Object-Verb. To illustrate this example, compare the following two sentences:

English: Rahul *plays* cricket.

Hindi: Rahul cricket *khelta hai.*

Case 3: Phrase Decoder

When the previous approaches fail to translate the text, we broke the sentences into phrases and on each phrases we apply algorithms based on the statistical machine translation technique to search and get the most probable and suitable translated output of the input sentence.

Case 4: Word Decoder

When all the previous cases fails to translate the text into target language then, EBMT divide the sentence into number of words. For each word, it tries to seek the thesaurus translation and simply assemble the outputs into a combine translated sentence.

The Proposed methodology is divided into different modules as shown in figure 3 below

(a) Preprocessing
(b) Text localization
(c) Text and non-text classification
(d) Text Detection
(e) Character recognition/extraction
(f) Text Translation using EBMT

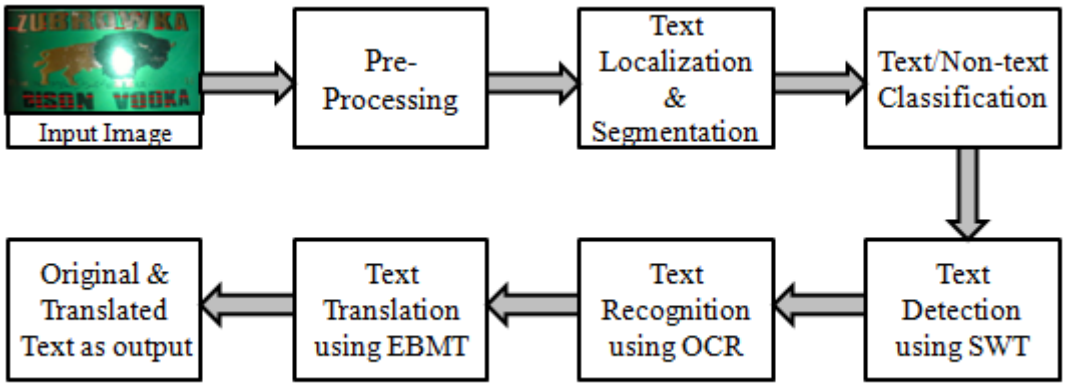

Figure 3: Proposed Methodology for text extraction from image

a)Preprocessing

In preprocessing, an RGB image is taken as input, as mostly we captured color images using camera. This RGB input image is first converted into gray-scale image because conversion into grey scale image reduces the processing overload. The filtering process is applied to grey scale images to remove any of the noises if present in the image. Lastly we apply an edge detection algorithm to detect and extract the edges from the final image.

b)Text localization

Edge detection algorithm gives a gray scale image that has detected edges of text, object, etc is first converted into binary image which help to distinguish text from other objects. Then thresholding is used to locate text candidate regions in an image. At last all the related edges are connected using closing operation which performed after the entire connected component is extracted using structuring element.

Figure 4: Text Localization

c) Text and non-text classification

The localized image contain both the text and non-text component hence we have to separate these components by making the bounding box for all the objects, text and connected components, for this a statistical approach are used to remove the non-text components. A very small component as well as a height & width of bounding box to remove the very big components a thresholding are used.

d) Text Detection

Text detection is performed after text and non-text classification using SWT as it make the text detection process more accurate. SWT detect the same intensity and width of region and consider it as a portion of text and make it available for OCR for recognition.

e) Character Recognition / Extraction

After SWT gives a detected portion of text; OCR extract that text and recognized it and display it as output**.** An OCR system takes a processed image of SWT as input and generates a character set (i.e. text) in editable form as an output because the same is provided as input to EBMT.

f) Text Translation using EBMT

EMBT takes a detected text as input and translate it into a target language using any of the case that translates it accurately.

4 Experimental Setup

The experiment is performed on manually selected and captured images using camera and banners of advertisement and create own dataset and we made a comparison with very recent approach based on natural language processing and pattern recognition. The dataset have 30 different images that have captured by different devices and in different environment. Complete dataset have 76 sentences and 430 words and this data is used to calculate the efficiency of our work.

Following images shows the localization and detection of text in images using stroke width transform after preprocessing of image i.e. binary image.

An image taken as input to a system is shown in figure 6 (a) below. The input image is taken by a camera which has oriented text as “ZUBROWKA BISON VODKA” is first processed and converts into binary image. The binary image is easy to localized and differentiate textual and non textual region. Localization and word segmentation is performed to detect and recognized text present in image. A step by step process is shown in figure 6 below.

(a) (b) (c)

Figure 5: (a) (b) (c) shows a text detection of input images using stroke width transform.

(a) (b) (c)

(d) (e) (f)

Figure 6: (a) Input image (b) Binary image of input image; (c) Line Segmentation of binary image; (d) Word Segmentation of image; (e) get detected text as output; (f) measure performance by precision and recall values

In this section, we evaluate the results of our experiment and compare with other approaches on a parameter of precision and recall value.

$$\text{Precision} = T_p / (T_p + F_{p)}$$

$$\text{Recall} = T_p / (T_p + F_{n)}$$

Where T_p is true positive i.e. the predicted condition is positive with positive condition.

F_p is false positive i.e. the predicted condition is positive with negative condition.

F_n is false negative i.e. the predicted condition is negative with positive condition.

Table below shows the precision and recall values of the tested images having different content, orientation, background, font, quality, etc.

Table 1: Result analysis of selected images using precision and recall values

Image No.	Precision	Recall
Image1.jpg	0.74	0.69
Image2.jpg	0.77	0.67
Image3.jpg	0.73	0.677
Image4.jpg	0.77	0.671
Image5.jpg	0.77	0.67

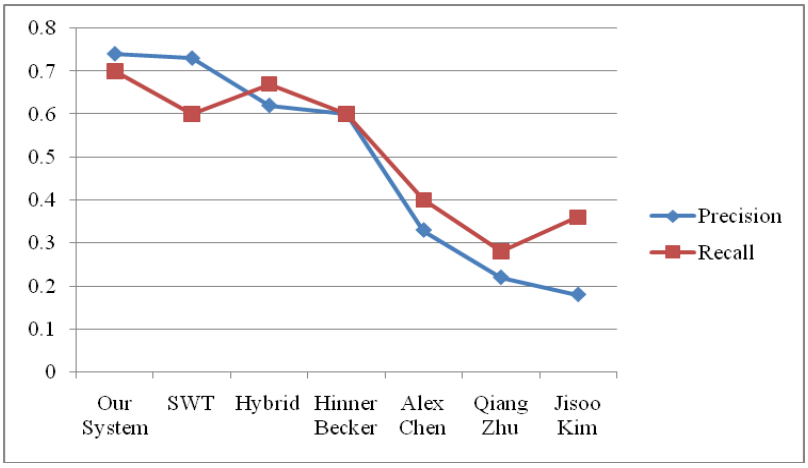

Figure 7: Comparative analysis of proposed system with existing systems

5 Conclusions and Future Scope

In this paper, we have proposed enhanced text detection using SWT and recognition of text in images using OCR and language translation technique using EBMT from images and the proposed method is tested with different types of images i.e. indoor and outdoor images, both images with scene text and caption text. All previous methods specified in references are discussed and studied and the drawbacks are reduced and thus getting an enhanced version of the previous works such that in this work, we get reduced noise levels have more clear detection of text and translation of detected text added advantage for culture conversion.

The accuracy of oriented text detection needs to be improved and one can add more languages to translate detected English text into target language. As a future scope one may also chose more parameter to compare the result.

References

[1] Gen Li ; Jie Liu ; Shuwu Zhang ; Yang Zheng,"Scene text detection with extremal region based cascaded filtering" IEEE International Conference on Image Processing (ICIP),19 August 2016.

[2] Xu-Cheng Yin ; Wei-Yi Pei ; Jun Zhang ; Hong-Wei Hao , "Multi-Orientation Scene Text Detection with Adaptive Clustering" IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 37, Issue: 9, Sept. 1 2015)

[3] Hyung Il Koo and Duck Hoon Kim, "Scene Text Detection via Connected Component Clustering and Nontext Filtering", IEEE Transactions On Image Processing, Vol. 22, No. 6, June 2013.

[4] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2963-2970.

[5] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes" IEEE Transactions on Image Processing VOL.13, NO. 1, January 2004.

[6] K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon, "Character-Stroke Detection for Text-Localization and Extraction", International Conference on Document Analysis and Recognition (ICDAR), 2005.

[7] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. on Image Processing, vol. 20, no. 3, pp. 800, Mar. 2011.

[8] Yao Li and Huchuan Lu, "Scene Text Detection via Stroke Width" International Conference on Pattern Recognition (ICPR 2012), November 2012.

[9] Ravina Mithe, Supriya Indalkar, Nilam Divekar, "Optical Character Recognition", International Journal of Recent Technology and Engineering (IJRTE), Volume-2, Issue-1, March 2013.

[10] K. Jung, "Text information extraction in images and video: A survey," Pattern Recognit., vol. 37, no. 5, pp. 977–997, May 2004.

[11] M. Swamy Das, B. Hima Bindhu , A. Govardhan, "Evaluation of Text Detection and Localization Methods in Natural Images," International Journal of Emerging Technology and Advanced Engineering (IJETAE), Volume 2, Issue 6, June 2012.

[12] Ms. Saumya sucharita Sahoo, "Review of Methods of Scene Text Detection and its Challenges," International Journal of Electronics and Communication Engineering (IJECET) ,Volume 5, Issue 1, January (2014), pp. 74-81

[13] Mohammad Shrif Uddin,Madeena Sultana, Tanzila Rahman and Umme Sayma Busra, "Extraction of texts from a scene image using morphology based approach",IEEE Transactions on image processing, 978-1-4673-1154/12(2012)

[14] Rishpa Sachdeva, Pooja Nagpal, " Text Localization and Extraction in Images Using Mathematical Morphology and OCR Techniques", International Journal of Scientific Engineering and Research (IJSER), Volume 1 Issue 1, September 2013

[15] Saurav Kumar, Andrew Perrault, "Text Detection on Nokia N900 Using Stroke Width Transform" ,December 14,2010.

[16] Gili Werner, "Text Detection in Natural Scenes with Stroke Width Transform", February 2013.

[17] Harpreet Singh, Deepinder Singh, "Text Confining and Extraction in Image Using Mathematical Morphology" International Journal of Science and Research (IJSR), Volume 3 Issue 6, June 2014.

[18] 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.

[19] G. Louloudis, B. Gatos, I. Pratikakis, C. Halatsis, "Line And Word Segmentation of Handwritten Documents"

[20] Vikas J Dongre , Vijay H Mankar, "Devnagari Document Segmentation Using Histogram Approach", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.1, No.3, August 2011.

[21] P. Nagabhushan, S. Nirmala, "Text Extraction in Complex Color Document Images for Enhanced Readability" Intelligent Information Management, 2010, 2, 120-133

[22] Boris Epshtein Eyal Ofek Yonatan Wexler,"Detecting Text in Natural Scenes with Stroke Width Transform".