International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 1, January-February 2015 ISSN 2278-6856

Multilingual Text Localization in Natural Scene Images using Wavelet based Edge Features and Fuzzy Classification

Shivananda V. Seeri[1], J. D. Pujari[2] and P. S. Hiremath[1]

[1]Department of Computer Applications (MCA), B. V. B. College of Engg. & Tech., Vidyanagar, Hubli-580031, Karnataka, India

[2]Department of Information Science and Engg., S.D.M College of Engg. & Tech, Dharwad-580001, Karnataka, India

Abstract

The aim of this study is to propose a new methodology for text region localization and non text region removal in natural scene images with complex background. In this paper, a new hybrid approach is proposed which locates text with different backgrounds. The proposed approach uses Haar wavelet, edge features, K-Means clustering, fuzzy classification and threshold concepts. The text localization algorithm is designed to locate text in different kinds of images. The performance of the approach is demonstrated by presenting promising experimental results for a set of natural scene images taken from ICDAR dataset and own data set and is analyzed in terms of precision and recall rates. The proposed method is developed as a preprocessing technique for text recognition in natural scene images.

Keywords: Natural scene image, Text localization, Haar wavelet, Sobel edge detector, K-means clustering, Mathematical Morphology

1. INTRODUCTION

Large amounts of textual information are embedded in natural scene images which are often required to be automatically processed for object recognition. This requires automatic detection, segmentation and recognition of visual text entities in natural scene images. Scene text may be any textual part of the scene images such as street signs, name plates or text appearing on t-shirts. Text in images contains useful information which can be used to fully understand images. Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos. This content can be in the form of objects, color, texture, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purpose. Since the text data can be embedded in an image with different font styles, sizes, orientations, colors against a complex background, the problem of extracting the candidate text region present in a natural scene image becomes a challenging one. Rapid development of digital multimedia technology has resulted in digitization of all categories of information resources, which are thus available in electronic form. Text extraction from images have many useful applications in document analysis, detection of vehicle license plate, analysis of article with tables, maps, charts, diagrams, keyword based image search, identification of parts in industrial automation, content based retrieval, object identification, street signs, text based video indexing, page segmentation, document retrieving, address block location and visually impaired people's assistance. Text localization can be done in three kinds of images, namely:

a) Document image
b) Scene text image
c) Caption text image

Document images may be in the form of scanned book covers, CD covers or video images. Text in images or videos of a natural scene is classified as scene text and caption text. Scene text is also called as graphics text. Natural images that contain text are called scene text. The name of caption text is meant for artificial text which is inserted or superimposed in an image The Figure 1 shows the three types of texts in images.

(a) (b) (c)

Figure 1: Digital images containing (a) Caption text (b) Scene text and (c) Document text.

A scene text (Figure 1(b)) appears within a natural scene which is often captured by a recording device. Scene texts occur naturally as a part of the scene and contain important semantic information helpful for image understanding for example advertisements that include artistic fonts, names of streets, institutes, shops, road signs, traffic information, board signs, nameplates, food containers, cloth, street signs, bill boards, banners, and text on vehicle and so on.

Scene text extraction can be used in detecting text-based landmarks, vehicle license detection/recognition, and object identification rather than general indexing and retrieval. It is difficult to detect, locate and extract scene text, since it may appear usually in unlimited number of poses, sizes, shapes and colors, resolutions, complex backgrounds, non-uniform lighting or blurring effects of varying lighting, complex movement and transformation, unknown layout, shadowing and variation in font style, size, orientation and alignment. Texts can be in different scripts. The localized text is input to an appropriate OCR for character recognition.

Several techniques have been reported in the literature for extracting text from an image. The existing methods are based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram techniques and so on. All these techniques have their own merits and demerits.

Humans have the peculiar ability of detecting presence of texts in varying scripts of different unknown languages, although they are trained using scripts of only one or two languages. Hence, automation of text detection and recognition involves building suitable machine intelligence that imitates similar to human ability. Text localization is the primary step in the process of text extraction from natural scene images. Hence, the main objective of the present paper is to propose an efficient algorithm to detect and localize multilingual text in natural scene images.

Text extraction in an image comprises five stages, namely, (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and (v) recognition (OCR). Text detection and text localization are closely related and more challenging stages which had attracted the attention of most researchers. The goal of the two stages is to generate accurate bounding boxes of all text objects in a natural scene images and provide a unique identity to each text. In the following, the recent literature focused on text detection and localization is reviewed.

1.1 Region based methods

Region-based methods use the properties of the color or gray-scale in a text region or their differences with the corresponding properties of the background. These methods use a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is needed to merge the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions.

A region based method is basically classified in to two sub categories, namely, edge based and connected component (CC) based methods. An edge based method is mainly focused on the high contrast between text and background region. In this method, firstly text edges are identified in an image and are merged. Finally, some heuristic rules are applied to discard non-text regions. A connected component based method considers text as a set of separate connected components, each having distinct intensity and color distributions. The edge based methods are robust to low contrast and different text size where as CC based methods are somewhat simpler to implement, but fail to localize text in images with complex backgrounds.

Leon [1] presented a method for caption text detection. It is included in a generic indexing system dealing with other semantic concepts which are to be automatically detected. To have a coherent detection system, the various object detection algorithms use a common image description. The author proposed the hierarchical region-based image model for image description and introduced the algorithm for text detection.

Debapratim [2] described the bottom-up approach of line segmentation of handwritten text. In this method, first the picture is sub divided into small 10 x 10 regions. If 50% of the square region is filled up with black pixels then the entire region is filled with black pixels. In this way graphically smooth image is found. Depending on the height and the position information these smoothed blocks are joined to get the individual lines.

Karin [3] designed a method for automatic text location and identification on colored book and journal cover. To reduce the amount of small variations in color, a clustering algorithm is applied in a preprocessing step. Two methods have been developed for extracting text hypotheses. One is based on a top-down analysis using successive splitting of image regions. The other is a bottom-up region growing algorithm. The results of both methods are combined to robustly distinguish between text and non-text elements. Text elements are binarized using automatically extracted information about text color. The proposed method is applicable to the extraction of text from other types of color images also.

Ankita, et al. [16], proposed a method based on histogram thresholding and entropy filtering, and connected components to extract Bengali characters and Bengali text from a certain level of complex multimedia images with a desired accuracy level.

1.2 Edge based methods

Edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge strength, density and the orientation variance are three distinguishing characteristics of text embedded in images, which can be used as main features for detecting text. The edges of the text boundary are identified and merged, and then several techniques are used to filter out the non-text regions. Edge-based text extraction algorithm is a general-purpose method, which can quickly and effectively localize and extract the text from both document and natural images.

Xiaoqing Liu[4] has developed method consisting of three stages: candidate text region detection, text region localization and character extraction. In the first stage, the magnitude of the second derivative of intensity as a measurement of edge strength is used, as this allows better detection of intensity peaks that normally characterize text in images. The edge density is calculated based on the average edge strength within a window. Considering effectiveness and efficiency, four orientations (0^o, 45^o, 90^o, 135^o) are used to evaluate the variance of

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 1, January-February 2015 ISSN 2278-6856

orientations, where 0^o denotes horizontal direction, 90^o denotes vertical direction, and 45^o and 135^o are the two diagonal directions. Edge detector is carried out by using a multiscale strategy, where the multiscale images are produced by Gaussian pyramids after successively applying low-pass filter and down-sample the original image reducing the image in both vertical and horizontal directions. In the second stage, characteristics of clustering can be used to localize text regions. In the third stage, existing OCR engine was used.

Xin Zhang [5] has used the color and edge features to extract the text from a video frame. In this work, two methods are combined, called color-edge combined algorithm, to remove text background. One method is based on the exponential changes of text color, called transition map model, which the other one uses the text edges in gray level image. After removing complex background, text location is determined using the vertical and horizontal projection method. This algorithm is robust to the images with multilingual text. To improve the efficiency of this method, the edge feature is added to remove background and then edge detection is performed on each color image using Canny operator and some morphology operation. Finally the background of text is removed with the help of transition map model.

Partha Sarathi Giri [4] had compared two basic approaches for extracting text region in images, which are edge-based and connected-component based. The algorithms are implemented and evaluated using a set of images that vary in terms of dimensions of lighting, scale and orientation. Accuracy, precision and recall rates for each approach are analyzed to determine the success and limitations of these approaches.

1.3 Morphological based methods

Mathematical morphology is a topological and geometrical based approach for image analysis. It provides powerful tools for extracting geometrical structures and representing shapes in many applications. Morphological feature extraction techniques have been efficiently applied for character recognition and document image analysis. It is used to extract dominant text contrast features from the processed images. The feature is invariant under geometrical image transformation, namely, translation, rotation, and scaling. Even after the lighting condition or text color is changed, the feature still can be maintained.

Jui- Chen Wu [6] presented a morphology-based text line extraction algorithm for extracting text regions from cluttered images. The method defines a novel set of morphological operations for extracting important contrast regions as possible text line candidates. In order to detect skewed text lines, a moment-based method is then used for estimating their orientation. According to the orientation, an x-projection technique can be applied to extract various text geometries from the text-analogue segments for text verification. However, due to noise, a text line region is often fragmented into different pieces of segments. Therefore, after the projection, a novel recovery algorithm is then proposed for recovering a complete text line from its pieces of segments. After that, a verification scheme is then proposed for verifying all extracted potential text lines according to their text geometries.

Rama Mohan [7] performed the edge detection operation using the basic operators of mathematical morphology. Using the edge the algorithm found out text candidate connected components. These components have been labeled to identify different components of the image. Once the components have been identified, the variance is found for each connected component considering the gray levels of those components. Then the text is extracted by selecting those connected components whose variance is less than some threshold value.

1.4 Texture- based methods

Texture-based methods use the observation that text in images has distinct textural properties that distinguish them from the background or other non text regions. The transform based methods, namely, Gabor filters, Wavelet, FFT, spatial variance can be used to detect the textural properties of a text region in an image. The main drawback of such methods is their complexity, while these are more robust than the CC based methods in dealing with complex background.

Chu Duc [8] presented a novel texture descriptor based on line-segment features for text detection in images and video sequences, which is applied to build a robust car license plate localization system. Unlike most of the existing approaches which use low level features (color, edge) for text / non-text discrimination, the aim is to exploit more accurate perceptual information. Scale and rotation invariant - texture descriptor which describes the directionality, regularity, similarity, alignment and connectivity of group of segments are proposed. An improved algorithm for feature extraction based on local connective Hough transform has also been investigated.

Kwang [9] used a novel texture-based method for detecting texts in images. A support vector machine (SVM) is used to analyze the textural properties of texts. The intensities of the raw pixels that make up the textural pattern are fed directly to the SVM, which works well even in high-dimensional spaces. Next, text regions are identified by applying a continuously adaptive mean shift algorithm (CAMSHIFT) to the results of the texture analysis. The combination of CAMSHIFT and SVMs produces both robust and efficient text detection, as time-consuming texture analyses for less relevant pixels are restricted, leaving only a small part of the input image to be texture-analyzed. The performance criterion was the classification accuracy of the SVMs for text and non-text patterns rather than the overall text detection results. For this purpose, 100 training images were divided into two different classes of 70 training images and 30 validation images from which training patterns and validation patterns were collected, respectively. The SVMs were then trained using the training patterns and tested using the validation patterns.

Neha and Banga [10] proposed a new hybrid approach to locate text in different backgrounds. The text localization algorithm system is designed to locate text in different

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 1, January-February 2015 ISSN 2278-6856

kinds of images and eliminates the need to devise separate method for various kinds of images. The approach uses various concepts, namely, Haar discrete wavelet transform (DWT), Sobel edge detector, morphological dilation and connected components. Finally, using some specific condition, the text is extracted in a bounding box. This algorithm is not sensitive to image color or intensity, uneven illumination and reflection effects. This algorithm can be used in large variety of application fields, e.g. vehicle license plate detection to detect number plate of vehicle, mobile robot navigation to detect text based land marks, object identification, identification of various parts in industrial automation, analysis of technical papers with the help of charts, maps, and electric circuits. The algorithm only analyses text box not a single character. Therefore, it requires less processing time which is essential for real time applications.

Goel, and Sharma [11] proposed the algorithm for text extraction in color images using 2D Haar wavelet transform (2D-DWT) along with mathematical morphological operators. The work is carried out in MATLAB 12.0 image processing tool and simulated results are tested on vehicle plate text extraction and the document images.

Sumathi and Gayathri Devi [12] proposed a new methodology for text region extraction and non text region removal from complex background colored images. This method is based on Gamma correction by determining a gamma value for enhancing the foreground details in an image. The approach also uses gray level co-occurrence matrices, texture measures, threshold concepts. The proposed method is a useful preprocessing technique to remove non text region and to show the text region in an image. Experimental results show good performance of the proposed method.

Adesh Kumar, et al. [13] proposed an approach based on Haar DWT, Sobel edge detector, and morphological operator. These mathematical tools are integrated to detect the text regions from the complex images. The proposed method is robust against language scripts and font size of the texts. The proposed method is also used to decompose the blocks including multi-line texts into single line text.

Deepa and.Victor [15] proposed text extraction algorithm for Tamil text extraction from images. It is demonstrated that the proposed method achieved expected accuracy of the text extraction for all the examples. The method uses Dual Tree complex wavelet transform and morphological dilation. It is robust against various conditions such as shadows, degradations, non-uniform illuminations, highlights, different font style and size and low contract images. The experiment results show that the proposed method reasonably extracts text regions with elimination of most non -text regions. This is can be further binarized and used by visually challenged persons through text to audio conversion software.

Satish Kumar, et al. [17] proposed method to localize potential text regions using connected component based heuristics from color layers(RGB). The system takes colored images as input; it detects text on the basis of certain text features such as frequency, orientation and spatial cohesion.

Keshava Prasanna, et al. [18] proposed an efficient algorithm to detect, localize and extract Kannada text from images with complex backgrounds. The proposed approach is based on the application of a color reduction technique, a standard deviation base method for edge detection, and the localization of text regions using new connected component properties.

Murthy and Y S Kumaraswamy [19] proposed a novel approach of extracting a text from image using two dimensional Haar Discrete Wavelet Transformation and K-Means Clustering. Proposed system has been experimented with images with single / multiple text, multiple text of different sizes / style / languages, images with uniform and non-uniform background.

Angadi and Kodabagi [20], proposed a methodology to detect and extract text regions from low resolution natural scene images. The proposed work is texture based and uses DCT based high pass filter to remove constant background. The texture features are then obtained on every 50x50 block of the processed image and potential text blocks are identified using newly defined discriminant functions. Further, the detected text blocks are merged and refined to extract text regions.

Teofilo, et al. [21] proposed object categorization framework based on a bag-of-visual-words representation for recognizing characters in situations that would traditionally not be handled well by OCR techniques. Proposed method tested on an annotated database of images containing English and Kannada characters.

2. METHODOLOGY

The block diagram of the proposed method for text localization in a natural scene image is shown in the Figure 2. The input image may be a color or gray scale image. If the image is color image, then it is converted into grayscale image and preprocessing operation is applied on the resultant image. In the algorithm, a color image is input data to the system and the segmented text with black background is the output. The proposed method consists of three stages, namely, preprocessing, feature extraction and classification which are described below.

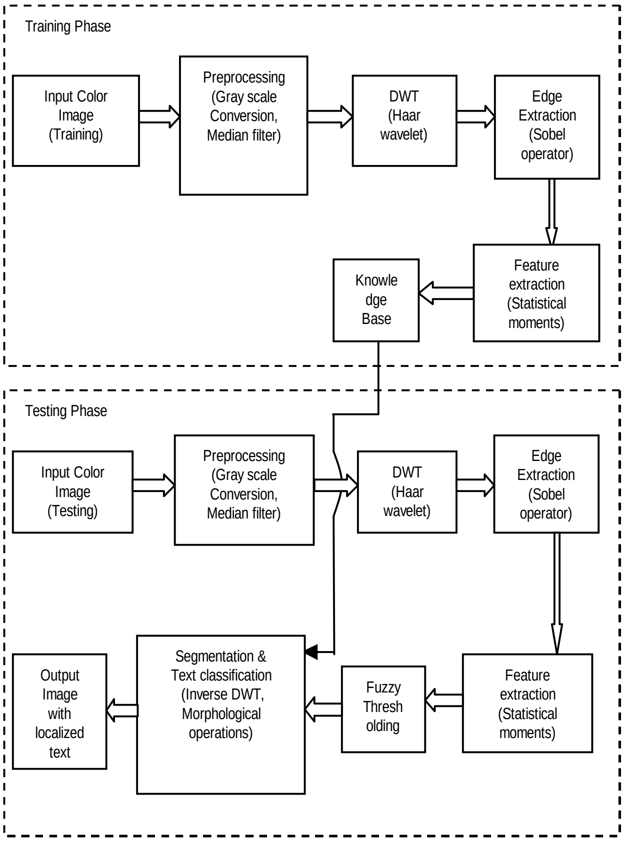


Figure 2: Block diagram of the proposed method

Preprocessing

The input image is pre-processed to facilitate easier detection of text regions. The input image is RGB colour image. The input image is converted to grayscale image. It will reduce the computational complexity as well as memory requirements significantly. The grayscale image is then filtered using a median filter. Median filter preserves sharp transitions or edges in the image.

Feature extraction

In this phase, Haar discrete wavelet transform is used, which provides a powerful tool of multi-resolution for modeling the characteristics of textured images. Most textured images are well characterized by their contained edges. It can decompose signal into different components in the frequency domain. The 2-d Haar DWT decomposes input image into four components or sub-bands, one average component (A) and three detail components (H, V, D) as shown in the Figure 3. The detail component sub-bands are used to detect candidate text edges in the original image. By finding the edges in the three sub bands, namely, horizontal (H), vertical (V) and diagonal (D) sub bands, fusion of the edge information contained in the three sub bands is done. In this way, candidate text regions can be found. In this algorithm, Sobel edge detector is used because it is efficient to locate strong edges pertaining to text. The next step is to form a combined edge map using logical "OR" and "AND" operators which removes some non text regions.

The statistical moments, namely, mean and standard deviation are computed for every 3 x 3 block of each of edge maps of the five images, namely, horizontal component, vertical component, diagonal component, combined image component, and original image, thus yielding ten statistical moments which form the feature vector. The average of all feature vectors corresponding to all the blocks in training image is obtained. This process is repeated for all the training images. The average of feature vectors of all the training images is computed and stored as the knowledge base, which is used for localization of the text regions in a test image. Training images are natural scene images, in which text regions that are present are manually localized to obtain text features required to build knowledge base used for localization of the text regions in test images.

A	H
V	D

Figure 3: Results of 2-d DWT decomposition

Classification

In the testing phase, the feature extraction and the classification of 3 x 3 blocks of a test image into text and non-text classes is done. The classification step uses fuzzy thresholding. The segmented image is subjected to morphological operators to obtain connected components corresponding to text regions. Then bounding boxes are obtained and drawn around the localized text regions in the input image. The method of fuzzy thresholding for text region is described below.

The Eculidean distances ED, of feature vectors FV of 3 x 3 blocks of an image from the centroid of feature vector cluster C, form a fuzzy set with the membership function μ_{ED} as defined below.

$$\mu_{ED}(x) = \left[\frac{\alpha-1}{ED_{median}-ED_{min}}\right](x-ED_{min})+1, \quad (1)$$

for $ED_{min} <= x <= ED_{median}$;

$$\mu_{ED}(x) = \alpha\left[\frac{1-e^{(x-ED_{max})}}{1-e^{(ED_{median}-ED_{max})}}\right], \quad (2)$$

for $ED_{median} <= x <= ED_{max}$

where x is a FV in C, and ED_{median}, ED_{max}, ED_{min} are the median, maximum, minimum of the set of ED values. The parameter α (with empirical value = 0.75) controls the thresholding of ED value for text and non-text regions. The graph of $\mu_{ED}(x)$ is shown in the Figure 4. The defuzzification is performed using the centroid formula to obtain the threshold value ED_{th}.

The graph of $\mu_{ED}(x)$ is shown in the Figure 4.

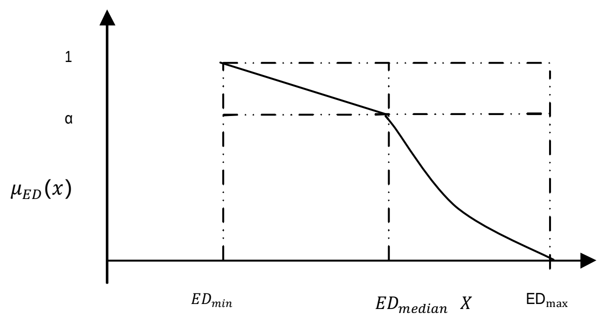


Figure 4: Graph of membership function μ_{ED}

The training and testing algorithms of the proposed method are given below.

Training Algorithm:

Input: A digital natural scene image containing text.

Output: Feature vectors of the text regions in the input image.

Step 1: Read input RGB Image I1.

Step 2: Convert input RGB image I1 to grayscale image I2 and apply median filter to obtain filtered image I3.

Step 3: Apply DWT to I3 using Haar wavelet to yield A, H, V and D component images. Apply Sobel edge operator on H, V, D and I3 to obtain edge maps, namely, EH, EV, ED and EI3 respectively. Obtain combined edge image EI4=EH + EV + ED.

Step 4: Consider 3 X 3 blocks of pixels from EH, EV, ED, EI4 and EI3 images (from top-left to bottom-right) and then compute mean and SD for each block. Compute the average values of mean and SD of all the blocks in EH, EV, ED, EI4 and EI3 and store these values as feature vector (with 10 components) of image I2 in the feature library.

Step 5: Repeat Step 1 to Step 4 for all the natural scene images in the training set.

Step 6: Apply fuzzy K-means (K=2) clustering algorithm to the set of feature vectors stored in the feature library and obtain two clusters C1 and C2 with centers FV1 and FV2 for the clusters C1 and C2 respectively.

Step7: Store feature vectors FV1 and FV2 as the knowledge base of text regions in the natural scene images containing two different classes, namely, text and non text classes.

Step 9: Stop

Testing Algorithm:

Input: A digital natural scene image I1.

Output: The natural scene image I1with localized text contained in bounding boxes.

Step 1: Read input Image I1

Step 2: Convert input RGB image I1 to grayscale image I2 and apply median filter to obtain filtered image I3.

Step 3: Apply DWT to I3 using Haar wavelet to yield A, H, V and D component images. Apply Sobel edge operator on H, V, D and I3 to obtain edge maps of H, V and D namely, EH, EV, ED and EI3 respectively. Obtain combined edge image EI4 = (EH + EV) * ED.

Step 4: Consider 3 X 3 blocks of pixels from EH, EV, ED, EI4 and EI3 images (from top-left to bottom-right) and then compute mean and SD for each block. Form the feature vector FV with these values of mean and SD as its components for all the blocks.

Step 5: Compute fuzzy thresholds ED1TH and ED2TH for the Eculidean distance between the feature vector FV and the feature vectors FV1 and FV2 of clusters C1 and C2, respectively.

Step 6: For each block, compute Eculidean distances ED1 and ED2 between FV of the block and FV1 and FV2 of clusters C1 and C2, respectively.

Step 7: If ED1 <= ED1TH OR ED2 <= ED2TH, the block belongs to the text region; otherwise, the block is non-text and set the corresponding non-text block to zeros in A, H, V and D components of image I2.

Step 8: Repeat Step 6 and Step 7 for all blocks. Finally, obtain the component images A, H, V, and D with the non-text regions set to zeros.

Step 9: Apply inverse DWT with Haar wavelet to obtain the image I5.

Step 10: Binarize I5 and apply morphological operations (with structuring element disk of radius 3, empirically fixed) on image I5 which yields segmented image I6. Label the connected components in I6. For each labeled component of I6 with area >= 300 (Empirical value), compute the bounding box and its aspect ratio R. If the aspect ratio R < 1 (Empirical value), the corresponding component is text region and its bounding box is drawn in I1, thus yielding localization of text in the input image I1.

Step 12: Stop

3. Results

The proposed algorithm is implemented using Intel Core i5 processor @ 2.5 GHz 4GB RAM and MATLAB R2009a. Experiments were carried out using 238 natural scene images (Including ICDAR data set), out of which 178 are used as training images and 60 as testing images. Each image is a RGB color image in JPEG format. These images contain multilingual texts which vary with respect to font style, font size, scale, lighting and orientation, shadowing of text in the image. The scripts of the text in the training images are English, Kannada and Hindi, where as text in testing images are in different languages, namely, English, Kannada, Hindi, Tamil, Bengali, Telugu, Urdu, Punjabi, Oriya, Malayalam, Marathi, Japanese, Russian and Chinese. The sample results of the proposed algorithm are shown in the Figures 5a and 5b. The performance of the proposed algorithm has been evaluated in terms of precision and recall rates defined as given below.

Figure 5a: Sample results of the proposed method showing input images and corresponding text localized output images

Figure 5b: Sample results of the proposed method showing input images and corresponding text localized output images

$$\text{Precision} = \frac{\text{Correctly localized Texts}}{\text{Correctly localized Texts} + \text{False positives}} \text{ X } 100$$

$$\text{Recall} = \frac{\text{Correctly localized Texts}}{\text{Correctly localized Texts} + \text{False negatives}} \text{ X } 100$$

Precision rate takes into account the false positives, which are the non-text regions in the image and have been localized by the algorithm as text regions. Recall rate takes into account the false negatives, which are text in the image, and have not been localized by the algorithm. Thus, precision and recall rates are the two measures which together determine the accuracy of the algorithm in locating text regions. The proposed method has yielded the precision and recall rates of 79.54% and 89.21%, respectively.

6. DISCUSSIONS

The experimental results demonstrate the effectiveness of the proposed method. Further, it is observed that the method is able to localize text regions of different scripts in natural scene images even though the algorithm is trained with only three scripts, namely, English, Kannada and Hindi, which demonstrates the robustness of the proposed method. It is also noticed that the localized text regions may contain multiple lines of text of different scripts, which need to be segmented for character extraction. However the proposed method fails to detect and localize text with large and thick characters, vertically written text, text with shadows, text with faint colors and text with light reflection. This has led to the false positives and false negatives in the experimental results. The proposed method requires modification to address such issues by proper selection of features and segmentation methods, leading to enhancement of precision and recall rates.

7. CONCLUSION

Text localization in a natural scene image with complex background is a difficult, challenging and important problem. In this paper, a novel method is proposed for text detection and localization using wavelet based features of edge images corresponding to input natural scene images. The proposed method uses Haar discrete wavelet Transform, Sobel edge detector, fuzzy thresholding and the morphological operators for segmenting and classifying the text regions. The proposed method has yielded the precision and recall rates of 79.54% and 89.21%, respectively. The experimental results demonstrate the effectiveness and robustness of the proposed method in detection and localization of multilingual text regions with variations in font style, font size, scale, lighting and orientation of text in an image. The localized text regions may contain multiple lines of text of different scripts that need to be segmented for character extraction, which will be considered in future work.

References

[1] Leon, M., Vilaplana, V., Gasull, A. and Marques, F., "Caption text extraction for indexing purposes using a hierarchical region-based image model," IEEE ICIP 2009, El Cairo, Egypt, 2009.

[2] Debapratim Sarkar, Raghunath Ghosh ,"A Bottom-Up Approach of Line Segmentation from Handwritten Text" ,2009.

[3] Karin Sobottka, Horst Bunke and Heino Kronenberg,"Identification of Text on Colored Book and Journal Covers", Document Analysis and Recognition, 20-22 ,1999.

[4] Xiaoqing Liu and Jagath Samarabandu, " Multiscale Edge-Based Text Extraction From Complex Images" , Multimedia and Expo, 2006 IEEE International Conference,2006.

[5] Xin Zhang, Fuchun Sun, Lei Gu, "*A Combined Algorithm for Video Text Extraction", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*,2010

[6] Jui-Chen Wu · Jun-Wei Hsieh · Yung-Sheng Chen,"Morphology-based text line extraction" Machine Vision and Applications 19:195–207 DOI 10.1007/s00138-007-0092-0,2008.

[7] Rama Mohan Babu, G., Srimaiyee, P. Srikrishna, A., "Text Extraction From Hetrogenous Images Using Mathematical Morphology" Journal Of Theoretical And Applied Information Technology © 2005 - 2010 Jatit.

[8] Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen ,"Robust Car License Plate Localization using a Novel Texture Descriptor", Advanced Video and Signal Based Surveillance, 2009.

[9] Kwang In Kim, Keechul Jung, And Jin Hyung Kim ,"Texture-Based Approach For Text Detection In Images Using Support Vector Machines And Continuously Adaptive Mean Shift Algorithm" , IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 25, No. 12, 2003.

[10] Neha Gupta, V.K Banga, "Localization of Text in Complex Images Using Haar Wavelet Transform", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-6, November 2012.

[11] Akash Goel, Prof. Yogesh Kumar Sharma, Text Extraction of Vehicle Number Plate and Document Images Using Discrete Wavelet Transform in MATLAB, IOSR Journal of Computer Engineering

(IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. V (Mar-Apr. 2014), PP 117-126.

[12] Sumathi, C.P. and G. Gayathri Devi, Automatic Text Extraction From Complex Colored Images Using Gamma Correction Method, Journal of Computer Science 10 (4): 705-715, 2014, doi:10.3844/jcssp.2014.705.715 Published Online 10 (4) 2014, ISSN: 1549-3636.

[13] Adesh Kumar, Pankil Ahuja, Rohit Seth, Text Extraction and Recognition from an Image Using Image Processing In Matlab, Conference on Advances in Communication and Control Systems 2013 (CAC2S 2013).

[14] Partha Sarathi Giri, Text Information Extraction And Analysis From Images Using DigitalImage Processing Techniques, ISSN (Print) : 2319 – 2526, Volume-2, Issue-1, 2013.

[15] S.T.Deepa, Dr.S.P.Victor, Tamil Text Extraction, International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462, Vol. 4 No.05 May 2012, 2176-2179.

[16] Ankita Sikdar, Payal Roy, Somdeep Mukherjee, Moumita Das and Sreeparna Banerjee, A Two Stage Method For Bengali Text Extraction From Still Images Containing Text, Natarajan Meghanathan, et al. (Eds): SIPM, FCST, ITCA, WSE, ACSIT, CS & IT 06, pp. 47–55, 2012. © CS & IT-CSCP 2012, DOI : 10.5121/csit.2012.2306.

[17] Satish Kumar, Sunil Kumar, Dr. S. Gopinath, Text Extraction From Images, International Journal of Advanced Research in Computer Engineering & Technology-2012, ISSN: 2278 – 1323, ISSN: 2278 – 1323.

[18] Keshava Prasanna, Ramakhanth Kumar P, Thungamani.M, Manohar Koli, Kannad Text Extraction From Images And Videos For Vision Impaired Persons, International Journal of Advances in Engineering & Technology, Nov 2011. ISSN: 2231-1963.

[19] Narasimha Murthy K N, Dr. Y S Kumaraswamy, A Novel Method for Efficient Text Extraction from Real Time Images with Diversified Background using Haar Discrete Wavelet Transform and K-Means Clustering, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue , No 3, September 2011, ISSN (Online): 1694-0814.

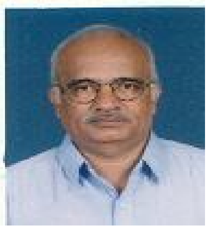
[20] S. A. Angadi, M. M. Kodabagi, A Texture Based Methodology for Text Region Extraction from Low Resolution Natural Scene Images, International Journal of Image Processing (IJIP) Volume(3), Issue(5).

[21] Teofilo´ E. de Campos, Manik Varma, Bodla Rakesh Babu, CHARACTER RECOGNITION IN NATURAL IMAGES, 1Available at http://research.microsoft.com/˜manik/

AUTHORS

Shivananda V. Seeri received the B.E. and M.Tech, degrees in Computer Science and Engineering from Karnataka University Dharwad in 1992 and Visvesvaraya Technological University *(VTU)* Belgaum in 2001, respectively. He is pursuing Ph.D. (Computer Science & Engineering) in VTU, Belgaum. He is presently working as Associate Professor, Dept of Computer Applications (MCA), B.V.B. College of Engg. & Tech., Hubli, Karnataka. He has published two papers on text extraction from natural scene images.

Dr. Jagadeesh D. Pujari has obtained Ph.D. (Computer Science) in digital image processing (CBIR). He has B.E and M.S degrees in Computer Science and Engineering. He has Published 5 papers in International Journals and Presented 19 papers in International conferences (Several of them are IEEE International conferences), 1 in national conference. He has been presently working as Professor and Head, Department of Information Science and Engineering, S.D.M College of Engg. & Tech, Dharwad, Karnataka.



Dr. P. S. Hiremath was born in May 1952, and has obtained Ph.D. (1978) in Applied Mathematics and M.Sc. (1973) in Applied Mathematics, from Karnataka University, Dharwad, Karnataka, India. He had been in the Faculty of Mathematics and Computer Science of various institutions in India, namely, National Institute of Technology, Surathkal (1977-79), Coimbatore Institute of Technology, Coimbatore (1979-80), National Institute of Technology, Tiruchirapalli (1980-86), Karnataka University, Dharwad (1986-1993), Gulbarga University, Gulbarga (1993-2014) and has been presently working as Professor, Dept of Computer Applications (MCA), B.V.B. College of Engg & Tech, Hubli, Karnataka. His research areas of interest are Computational Fluid Dynamics, Image Processing and Pattern Recognition and Computer Networks. He has published 168 research papers in peer reviewed International Journals and Proceedings of International Conferences.