International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 3, Issue 1, ISSN No. 2455-2143, Pages 53-60
Published Online May 2018 in IJEAST (http://www.ijeast.com)

A HYBRID APPROACH TO LOCALIZE TEXT IN NATURAL SCENE IMAGES

Rashmi V
Department of CSE
Jawaharlal Nehru National College Of Engineering
Shivamogga, India

Sankhya N. Nayak
Department of CSE
Jawaharlal Nehru National College Of Engineering
Shivamogga, India

ABSTRACT- Huge number of implementation has been carried out to earnestly and quickly extract the text regions from scene text images. Nevertheless, owing to exceeding challenges and issues, text localization is a flourishing area of analysis. Prewitt edge detector and average filter is applied to the image in order to have an enhanced image and potential text regions are extracted based on edge features in wavelet transformed image. Then, morphological operations are carried out, after which the text regions are tested for text or non-text content with the use of GLCM features and SVM classifier. Around these tested regions, bounding boxes are drawn, yielding localization of the text.

Keywords- **GLCM features, SVM classifier, Haar wavelet transforms**

I. INTRODUCTION

Huge amount of information will be embedded in natural scene images. Recent studies have shown immense interest in extracting data from images. Content can be in many different forms such as, objects, colour, texture, shape and relationship between them. Image gives us the semantic information that can be used for content based image retrieval and will be also helpful in indexing and classification. The problem of localizing the region of interest from an image becomes a challenging task, since the text could be embedded in an image in many sizes, orientations, colours, font styles and with a complex background.

Text will be having few distinct characteristics in terms of frequency, orientation and spatial cohesion information. Images will be classified depending on whether it contains text data or not. Extraction of text has quite a few applications like, analysing of document, vehicle license plate detection, article analysis that has tables, maps, charts, diagrams, keyword based image search, part recognition in automation industry, content based retrieval, street signs, object recognition, text based video indexing, page segmentation and also helps in assistance of visually impaired people.

Natural scene image with text are known as the scene text. These are the one that can be noticed as the content in genuine conditions without any suppositions or any requirements. Regular scene pictures samples' with text are depicted in Fig. 1.

Fig. 1: Samples of Natural scene images

Here in this paper, we have used Prewitt filter for pre-processing. In order to obtain texture features from the image, Gray level co-occurrence matrix (GLCM) technique is adopted. These features are used to train SVM classifier. For testing, Haar wavelet transform is applied and dilation is applied on the wavelet transformed image. Then, connected component analysis is carried out and then bounding box is drawn over the labelled components thus giving us localization of text.

The rest of the paper is organized as follows. We discuss related work in Section II. Section III explains proposed method. Section IV describes experiments and results. We finally conclude by using F-score to measure the effectiveness of our work.

II. LITERATURE SURVEY

Many algorithms have been proposed for recognizing text data in an image. Each method gives robust results for specified set of images. Text information extraction system is designed to extract textural information from the images. Fig. 2 shows the architecture of text information extraction (TIE) system.

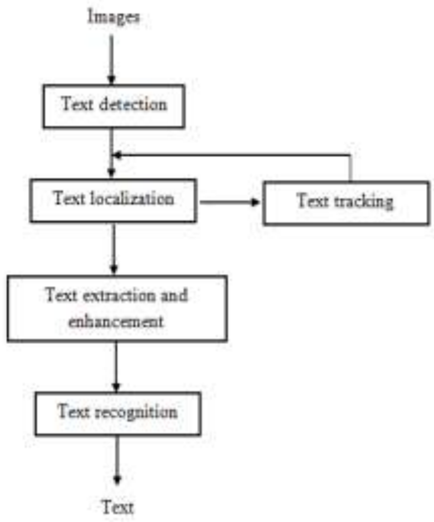

Fig. 2: Architecture of TIE system

The text detection is the identification of the presence of text in the image. Next step is text localization where area of content in the image is detected and bounding boxes are drawn around it. Next step is the process of text tracking which is carried out to deal with the stability of text position over neighbouring frames. Text extraction is the phase where texts in image are extracted from the background. Enhancement is necessary in order to segment text accurately for low resolution and noisy text image. OCR technology [23] is used with extracted text image to transform them into plain text.

Shivananda V. Seeri, J. D Pujari and P. S. Hiremath [1][6][12], have proposed a new methodology to localize text and to remove non-text part from natural scene images that has complicated background. In this reverence, a new hybrid procedure is proposed that identifies multiple language texts [18]. The promising text areas will be derived from the image that is depending on edge features in wavelet transformed image, which will be tested to check whether it is a text data or a non-text data with the help of GLCM features and SVM classifier. To locate text in natural scene images, text localization algorithm has been designed. (2015, 2015, 2016, 2015).

Shraddha and Sankhya N. Nayak [2] has proposed a system that extracts and recognises text using OCR. Here, geometric based noise removal method is applied for pre-processing. Binarization is carried out and textural features will be collected, which has been used by SVM classifier in order to detect text regions with the help of training given to it. (2015).

U. Elakkiya and M. Safa [3] has proposed a hybrid approach for finding and localizing text in natural scene images. Scene texts in noisy images can be detected using text area detector for the estimation of position of text and scale probabilities. This is carried out to separate components of candidate text with the help of local binarization algorithm. Conditional random fields (CRF) are used for conjoining unary and binary components. Next stage is word partition will be done by using word features like word count, centroid distances, bounding box distances, distances between words and ratio between centroid distances within separate words. Eventually, text that corresponds to sub-trees can be derived and the one with tiny components will be eliminated considering them as noises. (2015).

Poonam B Kadam, Latika R Desai [5] has proposed a framework to detect and recognise scheme efficient in urban scene text. Initially, pre-processing and segmentation is done. Next step is to extract feature set which then will be trained. Neural network will be used as a learning mechanism. Here, input character pattern will be matched with the stored character training set. With the help of labels given to the characters, network will itself learn many possible variation of a single pattern and becomes adaptive in nature. This step is called character matching. Based on the calculated scores, system will be called efficient. (2014).

Kumuda and Basavaraj [8] has put forth a different approach for deriving text from image. This is done by joining texture and connected components. Identification of text is done with the help of first and second order statistical features. Then, connected components technique is utilized for isolating text from surroundings. Then heuristic filters are made used in order to remove non text components. (2016).

Poonam and Latika [9] have developed a framework which is capable of identifying and extracting text from CAPTCHA images on its own. This method involves top-down technique and makes use of neural networks. Here, image will be partitioned and will be assigned to many different classes every time. All these classes will be processed in order to obtain the result. (2013).

Thaheera and Abdul [13] have developed a model to extract text from images on its own. Here concepts of text mining is involved, where initially, parsing of input will be carried out with the assistance of language factors. Unrelated components will be eliminated with the help of structured patterns, unary and binary components. To determine the existence of text, text area detector has been used. Finally, testing will be done by feeding the output to OCR engine. (2014).

Tahani Khatib, Huda Karajeh et.al [14] has proposed a multiple level algorithm for deriving text from images. Pre-processing of the image will be performed and edge image will be obtained. This image will be given to multiple levels of system. In each level, a promising candidate for text will be chosen. Later, from these chosen candidate image, background will be removed which gives the text from the image. (2015).

Priyanka Muchhadiya [16] has developed a method that can obtain data from multimedia. An algorithm has been developed in this regard that will extract the text on its own by separating from the background. This method uses colour reduction approach in order to obtain edge image by making use of geometrical properties of text. Text derived from the image will be given to OCR engine so that the text will be identified accurately. This system works fine with different font colours, styles and also for background complexities. (2015).

Pooja Singh [19] has developed a robust system in order to extract text from images. A detector has been designed to determine the presence of text in the image. A model has been developed that eradicates the non text components which make use of unary and binary components relationships. Lastly, with the help of learning based method, text components will be grouped together. (2016).

Amandeep, Manju Bala et.al [20] has been through the quality and nature of the text for efficient text extraction. Many different techniques and approaches have been analyzed and have found out that existing methods can be further made more efficient by developing hybrid algorithms that makes use of combination of many methods. (2016).

Rashedul Islam, Md. Rafiqul Islam and Kamrul Hasan Talukder [7][22], have developed a novel and advanced approach of text derivation by conjoining few factors from edge-based and connected-components. System is tested for its efficiency and has displayed the median accuracy. The system is robust in extracting text from many different kinds of scene images. (2011, 2016).

III. PROPOSED SYSTEM

The proposed method is developed to localize text from complex background. This method has two steps: training and testing phase. In training phase, pre-processing is done to the input image. Prewitt edge detection algorithm is used to locate strong edges. Next step is to derive gray-level co-occurrence matrix features from the image. Along with these four textural features, one of the statistical feature i.e., standard deviation is also computed. These features will be stored as knowledge base which will be used to train the support vector machine (SVM) classifier. In testing phase, 2D Haar wavelet transform is carried out on the edge map of the image which is supposed to be tested that gives us the components of 2D Haar wavelet transform. Morphological operations are applied on those components. Connected components analysis is achieved based on the geometrical properties. Then potential text regions are detected with the help of trained SVM classifier. The architecture of the system is depicted in Fig. 3.

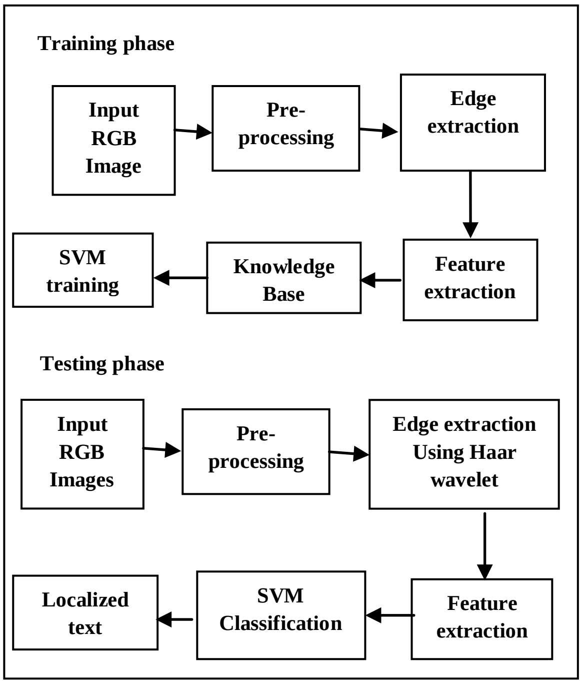

Fig. 3: System architecture

3.1 PRE-PROCESSING

In pre-processing stage, input colour image will be the input to the system. Then, this input image will be transformed into gray scale image to decrease memory requirements. Onto the gray scale image, average filter is applied for noise removal and for preserving sharp edges. These pre-processing steps will help in easier text detection.

3.2 EDGE DETECTION

Next stage is to detect edges. This technique efficiently identifies the points in the image that causes sharp changes in it. Different types of edge detection techniques are available. Here in this work, Prewitt operator is used even though Canny edge detector is efficient because, Canny edge method will detect every edge but here, not all edges are required. So, Prewitt edge method is the best method with respect to this work.

3.3 FEATURE EXTRACTION

In order to train SVM, textural features are extracted from the natural scene images. Here, gray level co-occurrence matrix (GLCM) features are made used in order to train the SVM classifier. From this matrix, fourteen different textural features can be extracted. Here in this work, four significant features are extracted namely, energy, contrast, homogeneity and entropy.

Energy describes smoothness of the image. It is a measure of textural uniformity or the pixel pair repetitions. Energy is calculated as,

$$\text{Energy} = \sum\sum C(i,j) \qquad \text{Eqn. 1}$$

Contrast calculates variation in local gray-level which gives the image spatial frequency. Contrast is determined as,

$$\text{Contrast} = ij(i-j)^2 \sum\sum C(i,j) \qquad \text{Eqn. 2}$$

Homogeneity (H) is also called as an inverse difference moment. It is an estimation of local uniformity in the image. Homogeneity is given by,

$$H = \sum\sum \frac{C(i,j)}{[1+(i-j)*2]} \qquad \text{Eqn. 3}$$

Entropy (E) is the measure of image complexity. It is given by,

$$E = \sum\sum ijC(i,j)\log C(i,j) \qquad \text{Eqn. 4}$$

3.4 POTENTIAL TEXT REGION DETECTION AND SVM CLASSIFICATION

In testing phase, image to be tested will be pre-processed initially. Input RGB test image will be converted to gray scale image. Onto this gray scale image, average filter is applied to remove noise from the image. Then, Haar wavelet transform is performed in order to obtain average (A), horizontal (H), vertical (V) and diagonal (D) components. Haar wavelet transform is the simplest among all the wavelet transform. The components of 2D Haar wavelet transform are as depicted in Fig. 4. After obtaining these components, morphological operations are performed on them.

A	H
V	D

Fig. 4: Components of 2D Haar wavelet transform

3.5 CONNECTED COMPONENT ANALYSIS

Next stage in testing phase is connected component analysis [10][11]. To identify each and every object clearly, labelling is done. Then, region properties are used to characterize every labelled region in the

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 3, Issue 1, ISSN No. 2455-2143, Pages 53-60
Published Online May 2018 in IJEAST (http://www.ijeast.com)

label matrix. These connected components are grouped depending on the set condition which says that area of the components that are greater than or equal to 45, eccentricity more than 0.2 and extent greater than 0.01 will be retained. Text region that is extracted from connected component analysis method. Then the trained SVM classifier will group them into text and non text content depending on the training given to the system. After grouping, bounding boxes are drawn for every extracted text area and background will be removed, which yields text localization in the input image.

IV. RESULTS AND ANALYSIS

The experiment is carried out for more than hundred different types of images having shadows, uneven light illumination and for multilingual texts.

4.1 RESULTS FOR INPUT IMAGE WITH MULTILINGUAL TEXTS

Input to the system is RGB image. Here, image that consists of texts in English is considered. Input image is depicted in Fig. 5.

Fig. 5: Input RGB image

Input RGB image selected for testing will be pre-processed i.e., converted to gray scale image and Prewitt edge detection method will be applied on it. and is depicted in Fig. 6.

Fig. 6: Prewitt edge image

Haar wavelet transform is performed on test image to obtain more accurate edge image in order to obtain average (A), horizontal (H), vertical (V) and diagonal (D) components and the same is depicted in Fig. 7.

Fig. 7: Haar Wavelet transformed edge image

Then, each of the Haar wavelet transform components is dilated. Next step is to apply connected components analysis. After this step, components are filtered based on the Area, Eccentricity and Extent that are calculated. Depending on the criteria on which the filtering is based, the components that satisfy will be retained. The image after applying filter is depicted in Fig. 8.

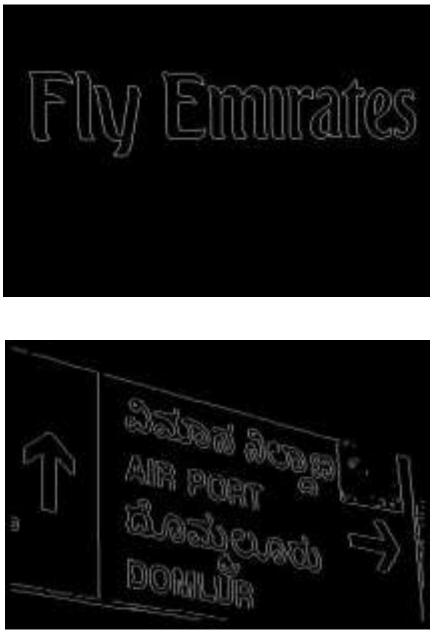

Fig. 8: Connected components after filtering

The components that are remaining after filtering process will then be surrounding boxes which is depicted in Fig. 9.

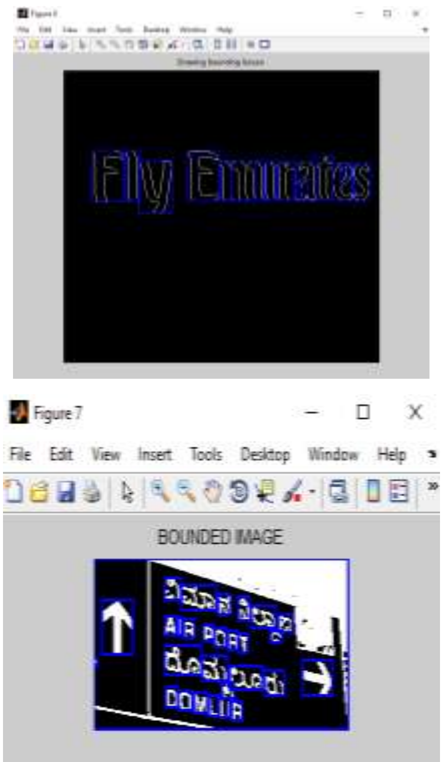

Fig. 9: Drawing bounding boxes

4.2 SYSTEM ANALYSIS

System analysis was carried out by considering about fifty images, both English and Kannada language. Efficiency of the proposed system was determined by calculating of Precision rate, Recall and F-score. The same is depicted in Table. 1.

TABLE. 1: QUANTITATIVE ANALYSIS

Language	Precision in %	Recall in %	F-score
English	92.20	94.00	91.25
English with Kannada	85.21	92.52	87.44

It has been noticed that, about 91.25 % of the images with English text was identified accurately and about 87.44 % of the images with both English and Kannada text was identified precisely.

V. CONCLUSION

Text localization from a natural scene image with complicated background is a challenging and significant problem. Here in this work, a hybrid technique is implemented for text localization that involve deriving potential text areas from an image depending on edge features in wavelet transformed image. For this purpose, Prewitt edge detector is made used to identify strong, sharp edges in training phase. For testing purpose, Haar wavelet transform is applied. Dilation is carried out for the components that are obtained from Haar wavelet. Then, connected components analysis is done and are filtered based on the set criteria to obtain potential text region. Then those filtered components are given to SVM classifier which will group them into text and non text content. For text contents, bounding boxes are drawn around it yielding us text localization of the given image.

VI. ACKNOWLEDGEMENT

Authors would like to thank the principal and staff of JNNCE for their support in this work.

VII. REFERENCES

[1] Seeri Shivananda V.; Pujari J.D.; and Hiremath P.S. (2015). Multilingual Text Detection In Natural Scene Images Using Wavelet Based Edge Features and SVM Classifier in International Journal of Advanced Research in Computer Science and Software Engineering, 2015. (Vol. 5, Issue 11, Pg 81-89).

[2] Shraddha.; Nayak Sankhya N. (2015). Text detection and Character Extraction in Natural Scene Images in International Journal of Emerging Technology and Advanced Engineering, 2015. (Volume 5, Issue 2).

[3] Elakkiya U.; Safa M. (2015). Text Detection in Natural Scene Images in International Journal of Latest Trends in Engineering and Technology, 2015. (Vol. 6, Issue 2, Pg 48-55).

[4] Tsai Sam S.; Chen David; Chen Huizhong; Hsu Cheng Hsin. (2011). Combining Image and Text Features: A Hybrid Approach to Mobile Book Spine Recognition in IEEE Transactions on Image Processing, 2011. (Vol. 21).

[5] Kadam Poonam B.; Desai Lathika R. (2014). A Hybrid Approach to Detect and Recognize Texts in Images in IOSR Journal of Engineering, 2014. (Vol. 4, Issue 7, Pg 13-19).

[6] Seeri Shivananda V.; Pujari J.D.; Hiremath P.S. (2016). Text Localization and Character Extraction in Natural Scene Images using Contourlet Transform And SVM Classifier in Image, Graphics And Signal Processing, 2016. (Pg 36-42).

[7] Pan Yi Feng; Hou Xinwen; Liu Cheng Lin. (2011). A Hybrid Approach to Detect and Localize Texts in Natural Scene Images in IEEE Transactions on Image Processing, 2011. (Vol. 20, Pg 800-813).

[8] Kumuda T,; Basavaraj L. (2016). Hybrid Approach to Extract Text in Natural Scene Images in International Journal of Computer Applications, 2016. (Vol. 142, Pg 18-22).

[9] Kadam Poonam B,; Desai Lathika R. (2013) A Hybrid Approach to Detect and Recognize Texts in Images in International Journal of Research in Computer and Communication Technology, 2013. (Vol. 2, Issue 7,Pg 405-410).

[10] Mohanabharathi R,; Surender K,; Selvi C. (2013). Detecting and Localizing Color Text in Natural Scene Images using Region Based and Connected Component Method in International Journal of Modern Engineering Research, 2013. (Vol. 3, Issue 1, Pg 331- 335).

[11] Geetha N,; Samundeeswari E. S. (2014). Image Text Extraction and Recognition using Hybrid Approach of Region Based and Connected Component Methods in International Journal of Engineering Research and Technology, 2014. (Vol. 3, Issue 6, Pg 1195-1200).

[12] Seeri Shivananda V,; Pujari J. D.; Hiremath P.S.; (2015). Multilingual Text Localization in Natural Scene Images using Wavelet Based Edge Features and Fuzzy Classification in International Journal of Emerging Trends and Technology in Computer Science, 2015. (Vol. 4, Issue 1, Pg 210-218).

[13] Thaheera H.; Samthum Abdul A. (2014). A Hybrid Model to automatically extract Text in International Journal for Scientific Research and Development, 2014. (Vol. 2, Issue 6, Pg 458-461).

[14] Khatib Tahani,; Karajeh Huda,; Mohammad Hiba.; Rajab Lama. (2015). A Hybrid Multilevel Text Extraction Algorithm in Scene Images, 2015. (Vol. 10, Pg 105-113).

[15] Zhao Zhenyu,; fang Cong,; Lin Zhouchen. (2015). A Robust Hybrid Method for Text Detection in Natural Scenes by Learning Based Partial Differential Equations in Elsevier, 2015. (Pg 23-34).

[16] Muchhadiya Priyanka. (2015). Detection and Localization of Texts from Natural Scene Images: A Hybrid Approach in International Multidisciplinary Research Journal, 2015. (Vol. 2, Issue 7).

[17] Gite Soniya A,; Hingoliwala H.A. (2015). An Efficient and Robust Method for Text Detection in Low and High Resolution Natural Scene Images in International Journal on Recent and Innovation Trends in Computing and Communication, 2015. (Vol. 3, Issue 7).

[18] Liao Wen Hung,; Wu Yi-Cheih. (2015). An Integrated Approach for Multilingual Scene Text Detection in International Journal of Computer Information Systems and Industrial Management Applications, 2015. (Vol.8 Pg 33-41).

[19] Singh Pooja. (2016). Scene Text Detection using the Regression Tree Technique in International Research Journal of Engineering and Technology, 2016. (Vol. 3, Issue 3, Pg 1678-1682).

[20] Kaur Amandeep,; Bala Manju,; Pooja. (2016). Text Recognition Past, Present and Future in International Journal on Recent and Innovation

Trends in Computing and Communication, 2016. (Vol. 4, Issue 4, Pg 507-516).

[21] Jhod Govil,; Sharma Prabhat. (2017). A Hybrid Approach for Detection and Recognition of Traffic Text Sign MSER and OCR in International Journal of Innovative Research in Computer and Communication Engineering, 2017. (Vol. 5, Issue 7, Pg 13071- 13076).

[22] Islam Rashedul,; Islam Md. Rafiqul,; Talukder Kamrul Hasan. (2016). An Approach to Extract Text Regions from Scene Image in International Conference on Computing, Analytics and Security Trends, 2016.

[23] Modi Hiral,; Parikh M.C. (2017). A Hybrid Approach for Optical Character Verification in National Conference on Advances in Computer Science Engineering and Technology, 2017. (Pg 180-184).

[24] Khandait S,; Khandait P,; Jambhular P. (2017). Scene Text Detection and Language Translation using SWT and EBMT in Advances in Intelligent Systems Research, 2017. (Vol. 137, Pg 547-555).