Text Extraction from Live Captured Image with Diversified Background using Edge Based & K-Means Clustering

Anuj Singh
Department of Electronics and Communication Engineering
S.S.G.B.C.O.E.T, Bhusawal, Maharashtra, India

A.S.Bhide
Professor, Department of Electronics and Communication Engineering
S.S.G.B.C.O.E.T, Bhusawal, Maharashtra, India

Preeti Singh
Assistant Professor, Computer Science & Engineering Department
BBD Group of Institutions, Lucknow, U.P., India

Abstract - The proposed system highlights a novel approach of extracting a text from image using K-Means Clustering. Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Recent studies in the field of image processing 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. As the commercial usage of digital contents are on rise, the requirement of an efficient and error free indexing text along with text localization and extraction is of high importance. The proposed system has broader scale of consideration of input image with much complicated backgrounds along with consideration of sliding windows. For much accuracy, morphological operation is included to accurately distinguish the text and non-text area for better text localization and extraction. The experimental result was compared with all the prior significant work in text extraction where the results show a much robust, efficient, and much accurate text extraction technique.

Keyword: Text Extraction, Discrete Wavelet Transform, K-Means Clustering, Morphological Operations

I. INTRODUCTION

Text data present in images contain useful information for automatic explanation, indexing, and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging in the computer vision research area. The proposed methods compare two basic approaches for extracting text region in images: edge-based and connected-component based. The algorithms are implemented and evaluated using a set of images that vary along the dimensions of lighting, scale and orientation. Accuracy, precision and recall rates for each approach are analyzed to determine the success and limitations of each approach. Text information extraction consists of 5 steps [1]: detection, localization, tracking, extraction and enhancement, and recognition (OCR). In case of scene text particular focus is set on extraction. This step is done on previously located text area of image and its purpose is segmentation of characters from background that is separation of text pixels from background pixels. Text extraction strongly affects recognition results and thus it is important factor for good performance of the whole process. Text extraction methods are classified as threshold based and grouping-based. First category includes histogram-based thresholding [2], adaptive or local thresholding [3] and entropy-based methods. Second category encompasses clustering-based, region based and learning-based methods.

The proposed work will introduce novel text extraction techniques with k-Means clustering. The system also introduces morphological operation like dilation and erosion for segregation of text and non-text regions for better accuracy. The rest of this paper is organized as follows. We discuss related work in Section II. The research methodology is discussed in Section-III. Implementation and Results is described in Section-IV and finally conclusion is described in Section-V.

II. RELATED WORK

Syed Saqib Bukhari [5] presents a new algorithm for curled textline segmentation which is robust to above mentioned problems at the expense of high execution time. His approach is based on the state-of-the- art image segmentation technique: Active Contour Model (Snake) with the novel idea of several baby snakes and their convergence in a vertical direction only. Samuel Dambreville [6] has combined the advantages of the unscented Kalman Filter and geometric active contours to propose a novel method for tracking deformable objects. Chen Yang Xu [7] has introduced a new external force model for active contours and deformable surfaces, which we called the gradient vector flow (GVF) field. The field is calculated as a diffusion of the gradient vectors of a gray level or binary edgemap. Wumo Pan E.t.al [8] has proposed a novel approach to detect texts from scene images captured by digital cameras. The system converts the text detection issue to a contour classification problem by means of the topographic maps, and performs shape classification by exploiting the over-complete and sparse structure in the shape data.

Fabrizio e.t. al [9] has presented a text localization technique which was considered to be efficient in the difficult context of the urban environment. The system uses a combination of an well-organized segmentation procedure based on morphological operator and a configuration of SVM classifiers with a variety of descriptors to estimate regions that are either text or non-text area. The system is competitive but generates many false positives Baba [10] has proposed a novel approach for text extraction by analyzing the textural evaluation in general scene images. The work has introduced a hypothesis that texts also have equivalent charecteristics that differentiates them from the natural background. The researcher has estimated spatial difference of texture to achieve the distribution of the degree of likelihood of text region. Aghajari [11] propose an approach to automatically localize horizontally texts appearing in color and complex images. The text localization algorithm achieved a recall of 91.77% and a precision of 96%. Hrvoje e.t. al [12] propose new method for scene text extraction in HSI color space using modified cylindrical distance as homogeneity criterion in region growing algorithm. The work has also introduced Solution for seed pixel selection based on horizontal projection. Jayant e.t. al [13] present a novel method for extracting handwritten and printed text zones from noisy document images with mixed content. We use Triple-Adjacent-Segment (TAS) based features which encode local shape characteristics of text in a consistent manner. The experiment was tested with only similar types of text present in page. The system also lags different scripts testing. Sumit [14] has presented a technique for using soft clustering data mining algorithm to increase the accuracy of biomedical text extraction. The development of the proposed algorithm is of practical significance; however it is challenging to design a unified approach of text extraction that retrieves the relevant text articles more efficiently. The proposed algorithm, using data mining algorithm, seems to extract the text with contextual completeness in overall, individual and collective forms, making it able to significantly enhance the text extraction process from biomedical literature.

The text detection algorithm is also based on color continuity [15]. In addition it also uses multi-resolution wavelet transforms and combines low as well as high level image features for text region extraction. The text finder algorithm proposed is based on the frequency, orientation and spacing of text within an image [16]. Texture based segmentation is used to distinguish text from its background. Further a bottom-up 'chip generation' process is carried out which uses the spatial cohesion property of text characters. The chips are collections of pixels in the image consisting of potential text strokes and edges. The results show that the algorithm is robust in most cases, except for very small text characters that are not properly detected. Also in the case of low contrast in the image, misclassifications occur in the texture segmentation.

A focus of attention based system for text region localization has been proposed by Liu and Samarabandu [17]. The intensity profiles and spatial variance is used to detect text regions in images. A Gaussian pyramid is created with the original image at different resolutions or scales. The text regions are detected in the highest resolution image and then in each successive lower resolution image in the pyramid.

The approach [18] used utilizes a support vector machine (SVM) classifier to segment text from non-text in an image or video frame. Initially text is detected in multi scale images using edge based techniques, morphological operations and projection profiles of the image. These detected text regions are then verified using wavelet features and SVM. The algorithm is robust with respect to variance in color and size of font as well as language.

III. RESEARCH METHODOLOGY

The goal of the paper is to implement, test, and compare and contrast two approaches for text region extraction in natural images, and to discover how the algorithms perform under variations of lighting, orientation, and scale transformations of the text. The algorithms are from Liu and Samarabandu [17] and Gllavata, Ewerth and Freisleben [18]. The comparison is based on the accuracy of the results obtained, and precision and recall rates. The technique used is an edge-based text extraction approach [16,17], and the technique used is a connected-component based approach [18].

In order to test the robustness and performance of the approaches used, each algorithm was first implemented in the original proposed format. The algorithms were tested on the image data set provided by Xiaoqing Liu and Jagath Samarabandu, as well as another data set which consists of a combination of indoor and outdoor images taken from a digital camera. The results obtained were recorded based on criteria such as invariance with respect to lighting conditions, color, rotation, and distance from the camera (scale) as well as horizontal and/or vertical alignment of text in an image. The experiments have also been conducted for images containing different font styles and text characters belonging to language types other than English.

A. Edge Based Text Region Extraction

The basic steps of the edge-based text extraction algorithm are given below, and diagrammed in Figure1. The details are explained in the following sections

a. Create a Gaussian pyramid by convolving the input image with a Gaussian kernel and successively down-sample each direction by half.
b. Create directional kernels to detect edges at 0, 45, 90 and 135 orientations.
c. Convolve each image in the Gaussian pyramid with each orientation filter.
d. Combine the results of step 3 to create the Feature Map.
e. Dilate the resultant image using a sufficiently large structuring element (7x7) to cluster candidate text regions together.
f. Create final output image with text in white pixels against a plain black background.

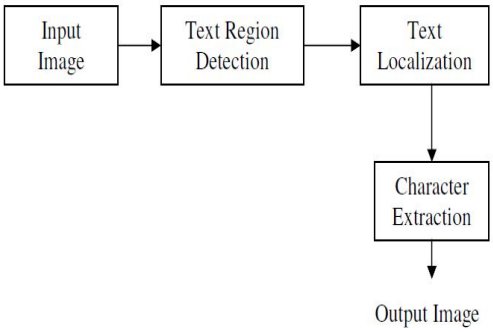

Figure 1 Basic Block diagram for Edge Based Text Extraction

This section corresponds to Steps 1 to 4 of Fig. 1. Given an input image, the region with a possibility of text in the image is detected [16, 17]. A Gaussian pyramid is created by successively filtering the input image with a Gaussian kernel of size 3x3 and down sampling the image in each direction by half. Down sampling refers to the process whereby an image is resized to a lower resolution from its original resolution. A Gaussian filter of size 3x3 will be used as shown in Figure1 each level in the pyramid corresponds to the input image at a different resolution. A sample Gaussian pyramid with 4 levels of resolution is shown in Figure 3. These images are next convolved with directional filters at different orientation kernels for edge detection in the horizontal (0°), vertical (90°) and diagonal (45°, 135°) directions. The kernels used are shown in Figure 5

Figure 2 Default filter returned by the fspecial Gaussian function in Matlab. Size [3 3], Sigma 0.5

Figure 3 Sample Gaussian pyramids with 4 levels

Figure 4 Each resolution image resized to original image size

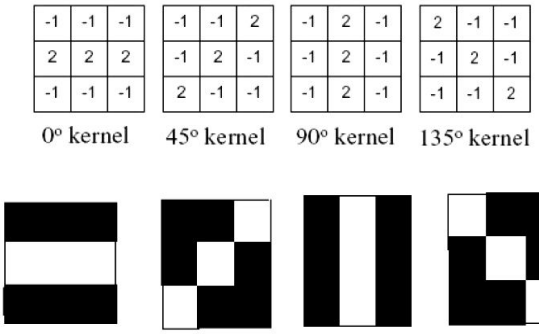

Figure 5 the Directional Kernels

After convolving the image with the orientation kernels, a feature map is created. A weighting factor is associated with each pixel to classify it as a candidate or non candidate for text region. A pixel is a candidate for text if it is highlighted in all of the edge maps created by the directional filters. Thus, the feature map is a combination of all edge maps at different scales and orientations with the highest weighted pixels present in the resultant map.

The common OCR systems available require the input image to be such that the characters can be easily parsed and recognized. The text and background should be monochrome and background-to-text contrast should be high [18]. Thus this process generates an output image with white text against a black background [16, 17]. A sample test image [16, 17] and its resultant output image from the edge based text detection algorithm are shown in Figures **6** (a) and 6 (b) below.

Figure 6 (a) Original image (b) Result

B. K-Means Clustering

The k-means is basically a clustering algorithm which partitions a data set into cluster according to some defined distance measure. One of the significant tasks in machine learning is to comprehend images and extracting the valuable details. In this direction of analyzing data within the image, segmentation is the first phase to estimate quantity of the object present in an object. K-means clustering algorithm is an unsupervised clustering protocol [16] which categorizes the input data points into multiple types based on their inherent distance from each other. The

protocol considers that the data features create a vector space and tries to locate normal clustering in them. The K-means function is given in (1).

[*mu, mask*]=kmeans (ima, k) **(1)**

where mu is the vector of class means, mask is the classification image mask, *ima* is the color image and *k* is the number of classes. The points are clustered around centroids in eq. (2) which are obtained by minimizing the objective [18].

Let ***m* = max (*ima*)+1 ,**

then **mu = {(1:k) * m} / (k+1)** **(2)**

The maximum function shown above is the maximum value in the in *ima* matrix which represents the colored image in order to achieve the maximum value of the content colors where the color values are revealed as a unit value for all pixel. This stage is done to explicitly used for estimating the histogram. The working principle of the k-means clustering algorithm in the proposed system is as discussed below.**(i)** The histogram of intensities which should highlight estimates of pixels in that specific tone is estimated as shown below(3) where, n = total estimates of observations k = total estimates of tones. The quantity of the pixels is estimated by the m_i which has equivalent value. The graph created with the help of this is only the alternative way to represents histogram. **ii.** The centroid with k arbitrary intensities as in eq. (2) should be initialized. **iii.** The following steps are iterated until the cluster labels of the image do not alters anymore. **iv.** The points based on distance of their intensities from the centroid intensities are clustered. **v.** The new centroid for each of the clusters is evaluated. **Σ==*kiimn*1** **(3)**

C. Morphological Operation

The morphological operations like dilation and erosions are used for better approach of refining text region extraction. The non-text regions are removed using morphological operations. Various types of boundaries like vertical, horizontal, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. But, it is also known that the identified region of text consists of all these boundary and region information can be the area where such types of boundaries will be amalgamated. The boundaries with text are normally short and are associated with one other in diversified directions. The proposed system has deployed both dilation and erosion for associating separated candidate text boundaries in every detail constituent sub-band of the binary image.

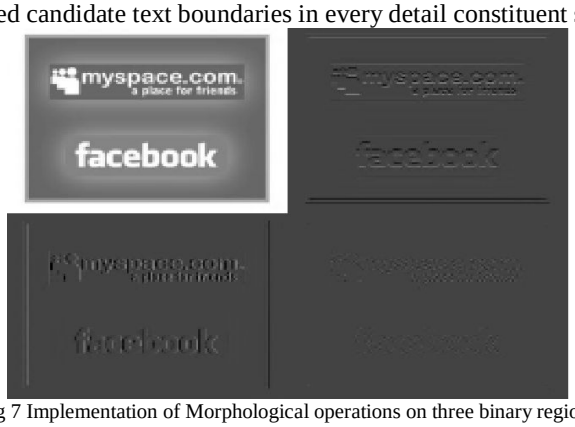

Fig 7 Implementation of Morphological operations on three binary regions

Finally, the morphological operations like dilation and erosion is designed exclusively to fit use-defined input of text based image with various types of characteristics.

IV. IMPLEMENTATION AND RESULTS

The framework is designed in Matlab in 32 bit system 1.8 GHz with dual core processor where total of 150 different types of images are considered for the experiment. The basic graphics video display card of DIAMOND AMD ATI Radeon is used for experimenting on both OS of Windows Vista and Windows 7. The implementation also considers images with single text, multiple text, text with different sizes of fonts, text with complex and simple background, text with different languages. The input image binarised to grayscale which is then subjected to discrete wavelet transform. The system then subjects the processed image into k-means clustering protocol. Morphological operation like erosion and dilation is deployed in order to remove the entire unwanted non-text region which can be confused with the text regions sometimes. Finally text localization and extraction takes place as shown in the results below.

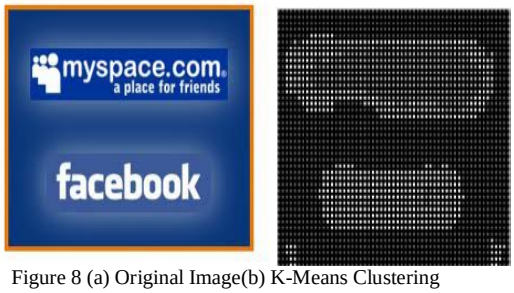

Figure 8 (a) Original Image(b) K-Means Clustering

(c) Erosion/ Dilation

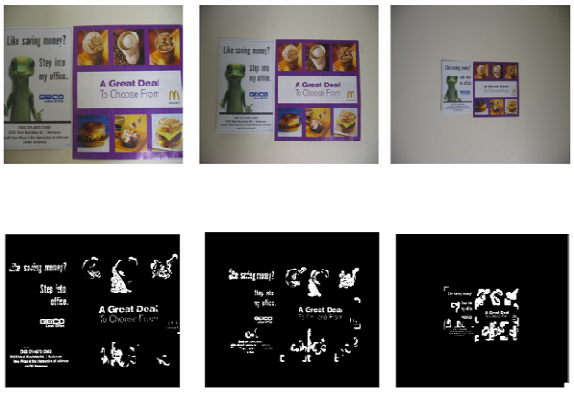

Figure 9 (Row 1) Original indoor images at three different scales (Row 2) Results from edge based algorithm

V. CONCLUSION

The proposed system has introduced a novel process of text extraction considering multiple cases of image with its textual contents. The system has been implemented using k-means clustering algorithm. It also deploys methodology of sliding window for reading sub-bands of high frequency. The results obtained by each algorithm on a varied set of images were compared with respect to precision and recall rates. In terms of scale variance, the connected component algorithm is more robust as compared to the edge based algorithm for text region extraction. In terms of lighting variance also, the connected component based algorithm is more robust than the edge based algorithm. In terms of rotation or orientation variance, the precision rate obtained by the connected component based algorithm is higher than the edge based, and the recall rate obtained by the edge based is higher than the connected component based The average precision rates obtained by each algorithm for the remaining test images are similar, whereas the average recall rate obtained by the connected component algorithm is a little lower than the edge based algorithm. Thus, the results from the experiments indicate that in most of the cases, the connected component based algorithm is more robust and invariant to scale, lighting and orientation as compared to the edge based algorithm for text region extraction. For the edge based algorithm, the overall precision rate is 47.4% and recall rate is 75.09%. For the connected component based algorithm, the overall precision rate is 50.10% and recall rate is 73.42%.

Morphological operations like dilation and erosion has been introduced finally to refine the text and non-text region appropriately. For more realistic and robust results, the 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. The system is also evaluated with major research results in the past for conventional text extraction

approach and is found to be potential for more accurately extracting text information. The future work will be to extending the similar concept of extracting text from video with higher accuracy.

REFERENCES

[1] A.W.M. Smeulders, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.

[2] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 34–58.

[3] Syed Saqib Bukhari, Coupled Snakelet Model for Curled Textline Segmentation of Camera-Captured Document Images, Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, 2009

[4] Samuel Dambreville, Tracking deformable objects with unscented Kalman filtering and geometric active contours, American Control Conference, IEEE, 2006

[5] Wumo Pan, T.D. Bui, C.Y. Suen, Text detection from natural scene images using topographic maps and sparse representations, IEEE, 2009

[6] J. Fabrizio, M. Cord, B. Marcotegui. Text Extraction from Street Level Images, CMRT09 - CityModels, Roads and Traffic 2009. Paris, France.

[7] Baba, Y.[Yoichiro], Hirose, A.[Akira], Spectral Fluctuation Method to Extract Text Regions in General Scene Images,: A Texture-Based Method to IEICE(E92-D), No. 9, September 2009, pp. 1702-1715

[8] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh, A Text Localization Algorithm in Color Image via New Projection Profile,The International Multi conference of Engineers and Computer Scientist, Vol-2, 2010

[9] Hrvoje Dujmić, Matko Šarić, Joško Radić, Scene text extraction using modified cylindrical distance, NNECFSIC'12 Proceedings of the 12th WSEAS international conference on Neural networks, fuzzy systems, evolutionary computing & automation, World Scientific and Engineering Academy and Society (WSEAS), 2011

[10] Sumit Vashishta, Yogendra Kumar Jain, Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 4, 2011

[11] K.C. Kim, H.R. Byun, Y.J. Song, Y.W. Choi, S.Y. Chi, K.K. Kim and Y.K Chung, *Scene Text Extraction in Natural Scene Images using Hierarchical Feature Combining and verification*, Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04), IEEE.

[12] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, *TextFinder: An Automatic System to Detect and Recognize Text in Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.

[13] Xiaoqing Liu and Jagath Samarabandu, *A Simple and Fast Text Localization Algorithm for Indoor Mobile Robot Navigation*, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5672, 2005.

[14] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, *Fast and Robust text detection in images and video frames*, Image and Vision Computing 23, 2005.

[15] Qixiang Ye, Wen Gao, Weiqiang Wang and Wei Zeng, *A Robust Text Detection Algorithm in Images and Video Frames*, IEEE, 2003.

[16] Xiaoqing Liu and Jagath Samarabandu, *An Edge-based text region extraction algorithm for Indoor mobile robot navigation*, Proceedings of the IEEE, July 2005.

[17] Xiaoqing Liu and Jagath Samarabandu, *Multiscale edge-based Text extraction from Complex images*, IEEE, 2006.

[18] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, *A Robust algorithm for Text detection in images*, Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis, 2003.