

Image Text Extraction and Recognition using Hybrid Approach of Region Based and Connected Component Methods

Ms. N. Geetha¹

Assistant Professor

Department of Computer Applications
Vellalar College for Women (Autonomous)
Thindal, Erode - 638 012, Tamilnadu, India

Dr. E. S. Samundeeswari²

Associate Professor

Department of Computer Science
Vellalar College for Women (Autonomous)
Thindal, Erode - 638 012, Tamilnadu, India

Abstract— Image mining concerns the extraction of implicit knowledge, image data relationship or other patterns not explicitly stored in the images. Text in images is one of the powerful sources of high-level semantics. If these text occurrences could be detected, segmented, and recognized automatically, they would be a valuable source of high-level semantics for indexing and retrieval. The proposal in this work entitled as “Image Text Extraction and Recognition Using Hybrid Approach of Region Based and Connected Component Methods” has been developed to detect, extract and recognize the text regions and the system is based on efficient edge detectors, connected component methods and optical character recognition. Text detection and extraction in images is important for content based image analysis. This problem is challenging due to the complex background, the non-uniform illumination, and the variations of text font, size and line orientation. The proposed method in this work develops an efficient text extraction and recognition methods that utilizes the concept of morphological operations using MATLAB. Existing text extraction methods – edge based and connected components produce better results when applied separately. But these methods produce more false positives. So it is proposed to take advantage of both methods and combine these methods in the proposed system. The result shows that the proposed methodology yields better results than the other two methods.

Keywords—Text region detection, Clustering, Binarization, Segmentation, Recognition.

I. INTRODUCTION

With the increasing use of digital image capturing devices, such as digital cameras, mobile phones and PDAs, content-based image analysis techniques are receiving intensive attention in recent years. Among all the contents in images, text information has inspired great interests, since it can be easily understood by both human and computer, and finds wide applications. The need for image mining is high in view of the fast growing amounts of image data.

Extracting text from images and video for indexing and retrieval purposes is a promising solution because most text has certain common characteristics that allow the

development of robust algorithms possible. These common characteristics include high contrast with the background, uniform color and intensity, horizontally aligned, text characters are in the foreground, background and text may be ambiguous, background and text are sometimes reversed, possesses certain frequency and orientation information, shows spatial cohesion—characters of the same text string (a word, or words in the same line) are of similar heights, orientation, and spacing.

Fully automatic text extraction from images has always been a challenging problem. The difficulties arise from variations of text in terms of character font, size, orientation, texture, language and color, as well as complex background, uneven illumination, shadows and noise of images.

Experiments show that applying conventional OCR technology directly leads to poor recognition rates. Therefore, efficient detection and segmentation of text characters from the background is necessary to fill the gap between image and video documents and the input of a standard OCR system. Optical Character Recognition (OCR) is a method to locate and recognize text stored in an image and convert the text into a computer recognized form and a uniform representation such as ASCII or Unicode. OCR systems can only deal with printed characters against clean backgrounds and cannot handle characters embedded in shaded, textured or complex backgrounds.

II. LITERATURE SURVEY

To better understand Image Text Extraction and Recognition, it is useful to review and examine the existing research works in literature. Therefore, recent approaches and methodologies used for extracting text from images have been discussed.

Wu et al. [1997] used a set of Gaussian derivative filters to extract texture features from local image regions. With the corresponding filter responses, all image pixels are assigned to one of the three classes (“text”, “non-text” and “complex background”), and then k-means clustering and morphological operators are used to group text pixels into text regions.

Chen et al. [2004] extracts texts in video images by detecting horizontal and vertical edges with a canny filter and smoothing the edge maps to extract candidate text lines with morphological operators. Several gray and gradient features are employed to verify candidate text lines with a support vector machine (SVM) classifier.

Zhu et al. [2007] first used nonlinear local binarization algorithm to segment candidate connected components. Several types of component features, including geometric, edge contrast, shape regularity, stroke statistics and spatial coherence features, are then defined to train an adaboost classifier for fast coarse-to-fine pruning of non-text components.

Sunil Kumar *et.al* [2007] proposed a novel scheme for the extraction of textual areas of an image using globally matched wavelet filters with fisher classifiers for text extraction from document images and scene text images. A clustering-based technique has been devised for estimating globally matched wavelet filters using a collection of ground truth images. The algorithm of text extraction scheme is extended for the segmentation of document images into text, background, and picture components which include graphics and continuous tone images.

Pan et al. in [2011] use a text region detector to estimate probabilities of the text position and scale information. This detector is based on histogram of oriented gradients (HOG) and a waldbost cascade classifier on image pyramids. The information extracted from each scale is merged into a single text confidence map and text scale map. Secondly, the gray-level image is segmented using the Niblack's local binarization algorithm and a connected component analysis is carried out with a conditional random field (CRF) model to assign candidate components as text and non-text by considering both unary component properties, such as width or height, and binary contextual component relationships, such as spatial distance or overlap ratio.

When studying the related work on this issue, it has been found that the text extraction and recognition can be designed by the following steps of processes such as pre-processing, text region detection, text localization and extraction and text recognition. Text recognition from images is a well known problem and is possible using methods like edge detection and connected component methods. It is decided to use hybrid approach by combining these two methods.

The proposal in this work is the extension of Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu *et.al* [2011].

III. SYSTEM DESIGN

Either Region based methods or Connected Component (CC) based methods alone cannot detect and localize texts very well. Specifically, region-based methods can extract local texture information to accurately segment candidate components while CC-based methods can filter out non-text components and localize text regions accurately. To overcome the above difficulties, a hybrid approach to extract and recognize texts in images is developed by taking advantages of both edge based and connected component based methods.

The main aim of this system is to design the software that extracts the textual information present in images. It processes images based on their pixels values to identify potential text blocks and recognize the text information. The text detection, extraction and recognition phase of the proposed method uses an efficient edge-based clustering and segmentation approach and template matching technique. The processing step of the proposed system is shown in Fig.1.

Image text extraction and recognition system using hybrid approach of region based and connected component methods receives an input in the form of a still image or a sequence of images. The images can be in compressed or un-compressed, gray scale or color image with text. Text extraction and recognition in images mainly consists of five steps. The first one is to find text region in original images. Then the text needs to be separated from background. Next step, a binary image has to be produced (for example, text is white and background is black). Then, segment the text characters in the binary image. And finally, recognize the segmented characters.

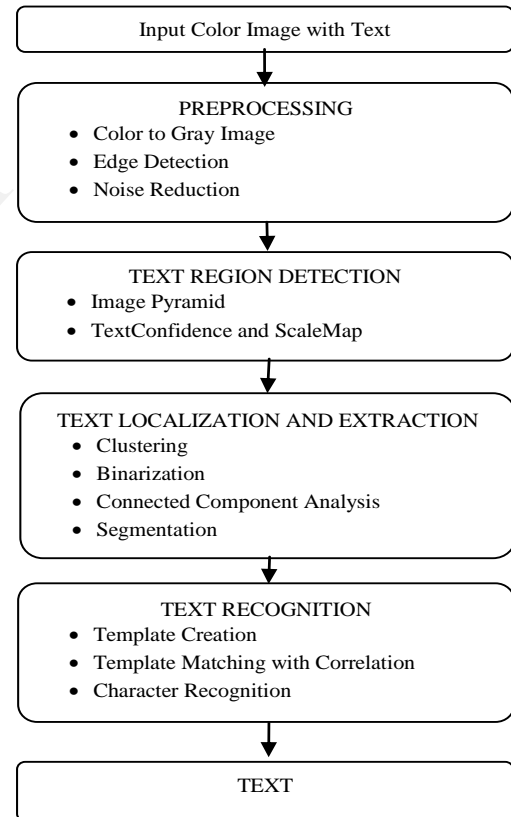


Fig.1. Steps of the proposed system

A. Pre-Processing

An image is the input to the system. The input color image is converted into gray image. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some for further processing.

In the proposed system, the canny edge filter gives the better edge detection in images. Edge detection process is responsible for detecting possible edges and boundaries in an

image. It is based on the fact that text has boundaries. It helps in filtering out the irrelevant data from the image, while maintaining the basic structural features in the image.

Noise reduction is usually the first process and the noise filter reduces noise at the cost of smoothing the image and hence softening the edges. Median filtering is better for removing outliers without reducing the sharpness of the image.

B. Text Region Detection

Initially, the original color image is converted into a gray level image, on which the image pyramid is built up with nearest interpolation to capture text information of different scales. The image pyramid decreases both size and resolution of an image and this is useful for speedup of detection process. The text confidence value of the pixel and the scale value are calculated by weighted arithmetic and geometric mean of corresponding pyramid pixel values.

C. Text Localization And Extraction

Text extraction is a critical step as it sets up the quality of the final recognition result. To extract and utilize local text region information, a text region detector is designed to estimate the text confidence based on which candidate text components can be segmented and analyzed accurately. Extraction can be grouped in two broad categories: clustering and binarization.

K-Means algorithm is a clustering technique, which classify pixels in an image into K number of clusters. The algorithm takes a 2 dimensional image as input. The main advantages of this algorithm are its simplicity and low computational cost, which allow it to run efficiently on large data sets. The process is as follow

- (i) Randomly choose number of clusters K.
- (ii) Compute the histogram of pixel intensities.
- (iii) Randomly choose K pixels of different intensities as centroids.
- (iv) Centroids are calculating by mean of pixel values in a region.
- (v) Now, compare a pixel to every centroid and assign pixel to closest centroid to form a cluster.
- (vi) When all pixels have been assigned, initial clustering has been completed.
- (vii) Recalculate position of centroids in K clusters.
- (viii) Repeat step 5 & 6, until centroids no longer move.
- (ix) Image separated into K clusters.

K means clustering algorithm is used to form candidate text regions by grouping text pixels. Based on the histogram of intensity values, higher gray scale of histograms are considered as text and other one as non text. Connected component classifier is used to estimate the text confidence. This is useful for strong classification of text location.

After text detection process, the text region can be localized well. But it is still not suitable for recognition because of the embedded complex background. To this end, a binarization process is needed to solve this problem. It segments the text character from the background. The output is a binary image that contains text candidate regions. The precise location of text in image can be indicated by bounding boxes, the text still needs to be segmented from the background to facilitate its recognition.

D. Text Recognition From Images

Text recognition performs character recognition on the binarized text image. The extracted text images can be transformed into plain text using optical character recognition technique. The process of optical character recognition (OCR) is segmentation, correlation and classification. In the first stage, OCR crops each character in the binary image. In the correlation phase OCR matches the cropped characters with the templates. During the classification process, it recognizes the text in binary image, if the cropped character matches with the template. In the template matching process, the characters will be identified as letters and the binary image will be converted to text.

IV. RESULTS AND DISCUSSIONS

The accuracy of text recognition from images depends on the text extraction. The problem of correct segmentation of isolated characters is the most difficult process in text extraction. A simple solution is proposed to this problem based on efficient edge and connected component based segmentation and optical character recognition techniques.

The text regions have been successfully extracted irrespective of the text font and size using the proposed system. The proposed method has provided a comprehensive model of text extraction and recognition in images.

In order to evaluate the performance of the proposed method, 20 images from a variety of sources such as logos, book covers, printed advertisements and text effects have been tested on the system.

The proposed system is applied to extract the text regions from background and recognize the characters from extracted text regions. Different stages of system output of a sample advertisement image are shown in Fig. 2-a to 2-h



Fig.2-a.

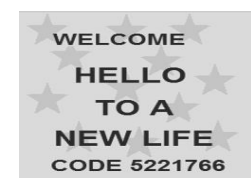


Fig.2-b.



Fig.2-c.

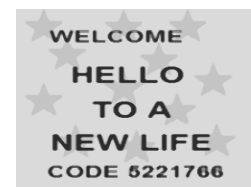


Fig.2-d.



Fig.2-e.



Fig.2-f.



Fig.2-g.

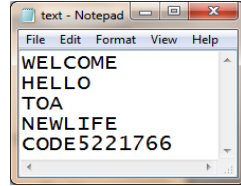


Fig.2-h.

Fig.2-a. Sample Input Color Image Fig.2-b. Gray Image Fig.2-c. Canny Edge Detected Image Fig.2-d. Median Filtered Image Fig.2-e. Text Region Detection Fig.2-f. Binary Image Fig.2-g. Text Extraction Image Fig.2-h. Recognized Text File

The existing Edge based and Connected Component based approach has been used for text detection and localization. These two existing methods increase false positives and processing time, etc. The proposed method combines these two methods to overcome the existing problems. To compare the two existing methods Edge based method and CC based method with the proposed system, the original image taken from the image database is shown in Fig. 3-a. The Final text extracted image using Edge based method, CC based method and Proposed method are shown in Fig. 3 b, c, and d respectively.



Fig.3-a IMAGE



Fig.3-b EDGE BASED



Fig.3-c CC BASED



Fig.3-d. PROPOSED

Hence, from the results, it is shown that proposed method is found to be better than the existing edge based and connected component based method for various images.

Performance Measures

Metrics used to evaluate the performance of the system are Precision, Recall rates and F-Score. Precision and Recall rates have been computed based on the number of correctly detected characters in an image in order to evaluate the efficiency and robustness of the system and the Metrics are as follows:

- False positives (FP) / False alarms are those regions in the image which are actually not characters of a text, but have been detected and recognized by the algorithm as text regions.
- False negatives (FN)/ are those regions in the image which are actually text characters, but have not been detected and recognized by the algorithm.
- Correctly detected and recognized characters are True Positives (TP).
- Incorrectly detected and recognized characters are True Negatives (TN).

Precision rate (P) is defined as the ratio of correctly detected and recognized characters to sum of correctly detected and recognized characters plus false positives. It is represented in equation as,

$$PrecisionRate = \frac{TP}{TP + FP} * 100\%$$

Recall rate (R) is defined as the ratio of the correctly detected and recognized characters to sum of correctly detected and recognized characters plus false negatives. It is represented in equation as,

$$RecallRate = \frac{TP}{TP + FN} * 100\%$$

F-score is the harmonic mean of recall and precision rates. It is represented in equation as,

$$F - Score = \frac{2 * P * R}{P + R}$$

Accuracy (A) is defined as the ratio of the sum of correctly and incorrectly detected and recognized characters to sum of correctly and incorrectly detected and recognized characters, false positives and false negatives. It is represented in equation as,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Precision and recall rates are useful as measures to determine the accuracy of proposed algorithm in locating correct text regions and eliminating non-text regions and recognizing correct text characters from images. Precision and recall rates, f-score and accuracy for different images used in this system are analyzed to determine the success and limitations. The comparison is based on the accuracy of the results obtained with precision and recall rates and F-Score. The average performance for the two existing methods, Edge based and Connected Component are listed in the Table 1 and 2 and that of the proposed system is shown in Table 3.

TABLE 1: Performance measures of the Edge based method

IMAGES	PRECISION RATE (%)	RECALL RATE (%)	F-SCORE (%)	ACCURACY (%)
1	100	100	100	100
2	41.18	100	58.33	47.37
3	100	100	100	100
4	93.33	100	96.55	93.75
5	93.33	100	96.55	94.12

6	66.67	100	80	90.91
7	100	93.33	96.55	96.15
8	71.43	100	83.33	71.43
9	100	100	100	100
10	40	100	57.14	45.45
11	100	100	100	100
12	50	100	66.67	54.55
13	100	100	100	100
14	100	100	100	100
15	100	100	100	100
16	100	100	100	100
17	100	100	100	100
18	100	100	100	100
19	100	100	100	100
20	100	100	100	100

The proposed method is compared with the two existing methods Edge based and Connected Component method and the average performance measures of comparisons are shown in Table 4. The average precision rate of the proposed system is 97.67 %, which means that there is less number of false positives. The system produces average recall rate 99.81%, as only some relatively weak text regions in the images are missed.

TABLE 2 : Performance measures of the Connected Component based method

IMAGES	PRECISION RATE (%)	RECALL RATE (%)	F-SCORE (%)	ACCURACY (%)
1	100	100	100	100
2	58.33	100	73.68	64.29
3	100	100	100	100
4	73.68	100	84.85	73.68
5	73.68	100	84.85	73.68
6	90	100	94.74	90.91
7	100	95.45	97.67	96.15
8	62.5	100	76.92	62.5
9	100	100	100	100
10	66.67	100	80	71.43
11	100	100	100	100
12	33.33	100	50	37.5
13	100	100	100	100
14	100	100	100	100
15	100	100	100	100
16	100	100	100	100
17	100	100	100	100
18	100	100	100	100
19	100	100	100	100
20	100	100	100	100

Hence, from the results in all cases, it is shown that proposed method is found to be better than the connected component method and edge based method for various types of images by providing better precision, recall, f-score and accuracy rates. The comparisons are shown graphically in Fig. 4 and Fig. 5 for Precision rate, and Recall rate and F- Score and Accuracy in column chart and line graph respectively.

TABLE 3: Performance measures of the Proposed method

IMAGES	PRECISION RATE (%)	RECALL RATE (%)	F-SCORE (%)	ACCURACY (%)
1	100	100	100	100
2	100	100	100	100
3	100	100	100	100
4	93.33	100	96.55	93.33
5	93.33	100	96.55	93.33
6	95.24	100	97.56	95.24
7	100	96.15	98.04	96.15
8	100	100	100	100
9	100	100	100	100
10	71.43	100	83.33	71.43
11	100	100	100	100
12	100	100	100	100
13	100	100	100	100
14	100	100	100	100
15	100	100	100	100
16	100	100	100	100
17	100	100	100	100
18	100	100	100	100
19	100	100	100	100
20	100	100	100	100

TABLE 4: Comparisons of average performance measures

METHODS	EDGE BASED	CONNECTED COMPONENT BASED	PROPOSED
MEASURES			
PRECISION	87.79	87.91	97.67
RECALL	99.67	99.77	99.81
F-SCORE	91.76	92.14	98.60
ACCURACY	89.69	88.51	97.47

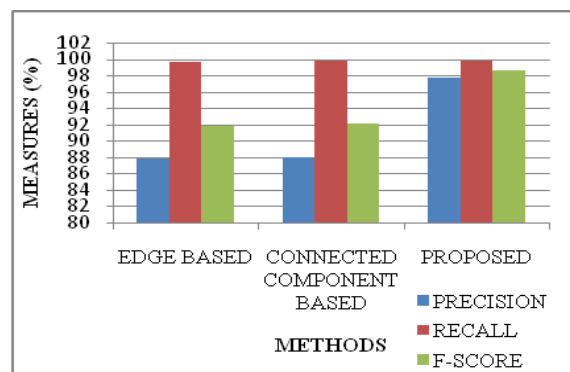


Fig.4. Comparisons of precision and recall rates and f-score with three methods

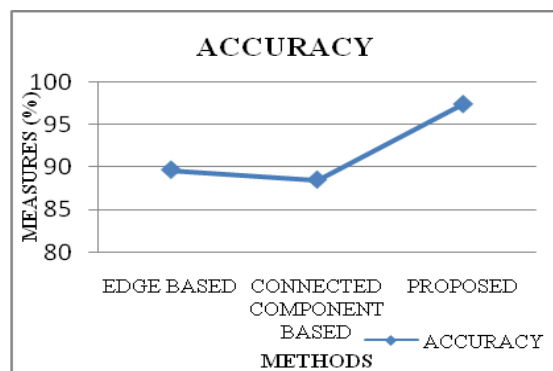


Fig.5. Comparison accuracy line graph

The proposed method has been tested on various types of images and the experimental results show that proposed method outperforms the Edge based method and Connected Component method. The results indicate that the method using efficient edge based and CC based techniques has the efficiency to discriminate between text and non text regions of images and recognize the text region characters.

V. CONCLUSION

This methodology has discussed the extraction and recognition of text from images. Text appearing in images is one feature that gives insight into image content. Segmentation is an important issue in the text extraction analysis research area. Text extraction plays a significant role in content based retrieval and storage systems. In the proposed work, text extraction is based on combining efficient edge and connected component techniques and text recognition is based on template matching techniques.

The proposed method has been extensively tested on different types of images. The results were promising, almost all the text regions could be extracted from images and the proposed method have been compared with the existing methods like edge based methods and the result shows that the proposed method performs better than existing methods. This proposed method enables the text enhancement at various fonts and sizes and improves the binarization of text images, which can significantly improve the recognition performance.

On the whole, the system is successful in achieving its objectives and can be utilized for automatic detection, extraction and recognition of text from images.

REFERENCES

- [1] Antani D. Crandall R. Kasturi, Robust Extraction of Text in Video, Proceedings of the International Conference on Pattern Recognition (ICPR'00) 2000
- [2] Basavaraj Amarapur and Nagaraj Patil ,Video Text Extraction From Images For Character Recognition, IEEE, Ottawa, May 2006
- [3] Chen D. T., Odobez J.-M., and Bourlard H., Text detection and recognition in images and video frames, Proceedings of Conference on Pattern Recognition, 2004
- [4] Chen X. L., Yang J., Zhang J., and Waibel A., Automatic detection and recognition of signs from natural scenes, IEEE Transactions on Image Processing, Jan. 2004
- [5] Chen X. R. and Yuille A. L., Detecting and reading text in natural scenes, IEEE Conference on Computer Vision and Pattern Recognition, 2004
- [6] Gonzalez R. C, Woods R. E, Eddins S. L, Digital Image Processing Using MATLAB, Pearson Prentice Hall, Second Edition, 2004
- [7] Gonzalez R. C, Woods R. E., Digital Image Processing, Addison-Wesley Publishing Company, Inc., Second Edition, 1993
- [8] Jung K., Kim K. I., and Jain A. K., Text information extraction in images and video: A survey, Conference on Pattern Recognition, 2004
- [9] Matti Pietikainen and Oleg Okun, Text Extraction From Grey Scale Page Images by Simple Edge Detectors, Machine Vision and Intelligent Systems Group, Infotech Oulu and Dept. of Electrical Engineering
- [10] Michael R. Lyu, *Fellow, IEEE*, Jiqiang Song, *Member, IEEE*, and Min Cai, A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction, IEEE Transactions On Circuits And Systems For Video Technology, VOL. 15, NO. 2, FEBRUARY 2005
- [11] Niblack W., An Introduction to Digital Image Processing, Strandsberg Publishing, 1985
- [12] Pan Y.-F., Hou X. W., and Liu C.-L., Text localization in natural scene images based on conditional random field, International Conference on Document Analysis and Recognition, 2009
- [13] Rainer Lienhart, *Member, IEEE*, and Axel Wernicke, Localizing and Segmenting Text in Images and Videos IEEE Transactions On Circuits And Systems For Video Technology, VOL. 12, NO. 4, APRIL 2002
- [14] Sunil Kumar, Rajat Gupta, Nitin Khanna, Santanu Chaudhury, and Shiv Dutt Joshi, Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model, IEEE Transactions on Image Processing, Vol. 16, No. 8, August 2007
- [15] Wu V., Manmatha R., and Riseman E.M., Finding Text in Images, In Second ACM International Conference on Digital Libraries, 1997
- [16] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, A hybrid approach to detect and localize texts in natural scene images, IEEE Transactions on Image Processing, Vol.20, No. 3, March 2011
- [17] You S. R. and Chen S. Y., Text Extraction from Color Images, Pattern Recognition and Image Analysis Vol. 12 No. 2 2002
- [18] Zhu Kai-hua, QI Fei-hu, Jiang Ren-jie, Xu Li, Automatic character detection and segmentation in natural scene images, Journal of Zhejiang University Science 2007