RESEARCH ARTICLE OPEN ACCESS

Automatic Text Extraction from Natural Scene Images using 3D-Oriented Wavelet Transform

Dr.M. Praneesh [1], Ashwanth.V [2], Febina.N [3], Sai Krishna P K [4]

[1]Assistant Professor, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science
[2]PG Scholar, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science

ABSTRACT

Text Extraction plays a major role in finding vital and valuable information. Text extraction involves detection, localization, tracking, binarization, extraction, enhancement and recognition of the text from the given image. These text characters are difficult to be detected and recognized due to their deviation of size, font, style, orientation, alignment, contrast, complex colored, textured background. Due to rapid growth of available multimedia documents and growing requirement for information, identification, indexing and retrieval, many researchers have been done on text extraction in images. Several techniques have been developed for extracting the text from an image. The proposed methods were based on morphological operators, wavelet transform, artificial neural network, Skeletonization operation, edge detection algorithm, histogram technique etc. All these techniques have their benefits and restrictions. In this research work, we propose a novel framework to extract text regions from scene images with complex backgrounds and multiple text appearances. This framework consists of three main steps: boundary clustering (BC), stroke segmentation, and string fragment classification. In BC, we propose a new bigram-color-uniformity-based method to model both text and attachment surface, and cluster edge pixels based on color pairs and spatial positions into boundary layers. Then, stroke segmentation is performed at each boundary layer by color assignment to extract character candidates. We propose two algorithms to combine the structural analysis of text stroke with color assignment and filter out background interferences. Further, we design a robust string fragment classification based on 3D-OWT features. The features are obtained from feature maps of gradient, stroke distribution, and stroke width. The proposed framework of text localization is evaluated on scene images, born-digital images, broadcast video images, and images of handheld objects captured by blind persons.

Keywords — Clustering, Artificial Neural Networks, edge detection, 3D-OWT

I. INTRODUCTION

With the recent advances in digital technology, more and more databases are multimedia in nature, containing images and video in addition to the textual information. The research on text extraction from images has been growing recently [1]. Many methods have been proposed based on edge detection, binarization, spatial-frequency image analysis and mathematical morphology operations. All these systems make evident that the text areas cannot be perfectly extracted from the image because natural scenes consist of complex objects, sometimes highly textured, buildings, trees, window frames and so on, giving rise to false text detection and misses. The first step in developing our text reading system is to address the problem of text detection in natural scene images [2]. The written text provides important information and it is not an easy problem to reliably detect and localize text embedded in natural scene images [8]. The size of the characters can vary from very small to very big. The font of the text can be different. Text present in the image may have multiple colors. The text may appear indifferent orientation. Text can occur in a complex background. And also the textual and other information captured is affected by significant degradations such as perspective distortion, blur, shadow and uneven lighting. Hence, the automatic detection and segmentation of text is a difficult and challenging problem. Reported works have identified a number of approaches for text localization from natural scene images. The various approaches are categorized as connected component based, edge based and texture based methods. Connected component based methods use bottom up approach to group smaller components into larger components until all regions are identified in the image. A geometrical analysis is later needed to identify text components and group them to localize text regions. Edge based methods focus on the high contrast between the background and text and the edges of the text boundary are Identified and merged. Later several heuristics are required to filter out non-text regions. But, the presence of noise, complex background, and significant degradation in the low resolution natural scene image can affect the extraction of connected components and identification of boundary lines, thus making both the approaches inefficient[3]. This paper presents a method for image preprocessing based on Shannon’s definition of information Entropy. The approach is generally applicable to any image. The basic concept is that the background remains informatively poor, whereas the objects carry relevant information. This method preserves the details, highlights edges, and decreases random noise.

The problem of Automatic Character Recognition can be stated as below.

> "*Given a set of M objects, which are divided into N non-intersecting subsets, each representing a single character and to each character there exists a description X in the form of a multidimensional vector representing features, the problem of recognition involves finding a match to N_i to an alphabet or digit, where i is the i^{th} image subset.*"

Thus problem is to design an algorithm that provides optimal solution, which efficiently decides the class to which the character belongs. Various techniques have been proposed which can be grouped into three main categories, namely,

(i) Structural analysis
(ii) Neural networks.
(iii) Template matching or correlation method

Structural analysis uses a decision tree to assess the geometric features of the character's contour. This method is tolerable to poor quality of the shape of the characters, but its results are unreliable with complex natured images. As most of Indian vehicle number plates are complex in nature, usage of this method might not give accurate results. Neural networks are methods based on training and learning process rather than programming. While learning to recognize a recurring pattern, the network constructs a model that adapt to unique features of the characters. This method has a proven track record with character recognition. The main drawback of using neural networks is the time complexity involved during training. While using with ANPR systems, the neural network has to be trained whenever a new pattern is discovered, which is very frequent. Hence, this method, eventhough, is very popular in OCR applications is not used for ANPR.

The template matching is the method that is designed mainly to recognize characters in scanned documents. This method takes each character of the plate and attempts to match it to a set of predefined standards. One drawback encountered while using template matching method is that it is susceptible to noise, that is, the accuracy of the result depends on the quality of the image. In the present research work, the template matching method is used during character recognition as this method has less implementation complexity and is very fast. To solve the problem of noise handling, a preprocessing phase is included, where a Denoising method is applied to improve the quality of the image.

II. RELATED WORKS

[2] Used genetic neural network, morphology and active contours for ANPR. In their algorithm, first the input image is divided into several virtual regions sized 10x10 pixels. In the next step, several performance algorithms were applied within each virtual region. Algorithms such as edge detection, histograms and binary thresholding were used. The output after region enhancement is used as input to the genetic neural network, which provides the initial selection of the probable situation of the number plate. Further refinement was applied using active contours to fit the output tightly to the number plate. With a small and well chosen subset of images, the system was able to deal with a large variety of images with real world characteristics. According to Chitrakala Gopalan, Manjula.D(2008)-The problem of text extraction from different kinds of images such as Scene text images, Caption text images & document images with an unified framework,So the proposed method is to apply a variation of Contourlet transform on images to decompose it into set of directional sub bands with texture details capture in different orientations at various scales[1].

NobuoEzaki, Marius Bulacu Lambert Schomaker(2004) - Proposed four character-extraction methods based on connected components. And they tested the effectiveness of the methods on the ICDAR 2003 Robust Reading Competition data. The performance of the different methods depends on character size. In the data, bigger characters are more prevalent and the most effective extraction method proves to be the sequence: Sobel edge detection, Otsu binarization, connected component extraction and rule-based connected component filtering[2].

S. A. Angadi , M. M. Kodabagi - A methodology to detect and extract text regions from low resolution natural scene images is presented. 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[3].

G. Sahoo, Tapas Kumar., et al - Proposed a set of sequential algorithms for text extraction and enhancement of image using cellular automata are proposed. The image enhancement includes gray level, contrast manipulation, edge detection, and filtering[5].

JiSoo Kim, SangCheol Park,et al-Proposed three text extraction methods based on intensity information for natural scene images. The first method is composed of gray value stretching and binarization by an average intensity of the image. This method is appropriate to extract texts from complex backgrounds. The second method is a Split and Merge approach which is one of well-known algorithms for image segmentation. The third one is a combination of the two. Experimental results show that the proposed approaches are superior to conventional methods both in simple and complex images[7].

G. Rama Mohan Babu, P. Srimaiyee-Proposes an algorithm which is insensitive to noise, skew and text orientation. It is free from artifacts that are usually introduced by thresholding using morphological operators[6].

Hidden Markov Model (HMM) is a doubly stochastic process, with an underlying stochastic process that is not observable (hence the word hidden), but can be observed through another stochastic process that produces the sequence of observations. An HMM is called discrete if the observations are naturally discrete or quantized vectors from a codebook or continuous if these observations are continuous. HMMs have been proven to be one of the most powerful tools for modelling speech and later on a wide variety of other real-world signals. These probabilistic models offer many desirable properties for modelling characters or words. One of the most important properties is the existence of efficient algorithms to automatically train the models without any need of labelling pre-segmented data.

A Neural Network (NN) is defined as a computing structure consisting of a massively parallel interconnection of adaptative "neural" processors. The main advantages of neural networks lies in the ability to be trained automatically from examples, good performance with noisy data, possible parallel implementation, and efficient tools for learning large databases. NNs have been widely used in this field and promising results have been achieved, especially in handwriting digit recognition. The most widely studied and used neural network is the Multi-Layer Perceptron (MLP). Such an architecture trained with back-propagation is among the most popular and versatile forms of neural network classifiers and is also among the most frequently used traditional classifiers for handwriting recognition. Other architectures include Convolutional Network (CN), Self-Organized Maps (SOM), Radial Basis Function (RBF), Space Displacement Neural Network (SDNN), Time Delay Neural Network (TDNN), Quantum Neural Network (QNN), and Hopfield Neural Network (HNN).

III. METHODOLOGY

In the present research work, enhancement of images are termed as 'Preprocessing Techniques' and consist of two stages.
Stage 1: Implement Denoising method to remove the noise generated during acquisition of images.
Stage 2: Analyze the performance of various edge detection algorithms on complex images.

Edge detection play an important role as a pre-processing step in many image processing applications, particularly in object classification and recognition systems. An edge is defined as the boundary between two regions with relatively distinct gray level properties. Information about edges in an image helps to identify the contours of an image and help to retrieve regions enclosed by those contours. An 'edge' image represents a higher level of abstraction (i.e. less information to process) and edges are features invariant to absolute illumination (as opposed to color information). Edge detectors are algorithms which perform edge detection. Edge detectors have gained popularity because the classifiers are usually over-fitted to one particular lighting / intensity condition and edge detection is fairly robust to changing illumination conditions.

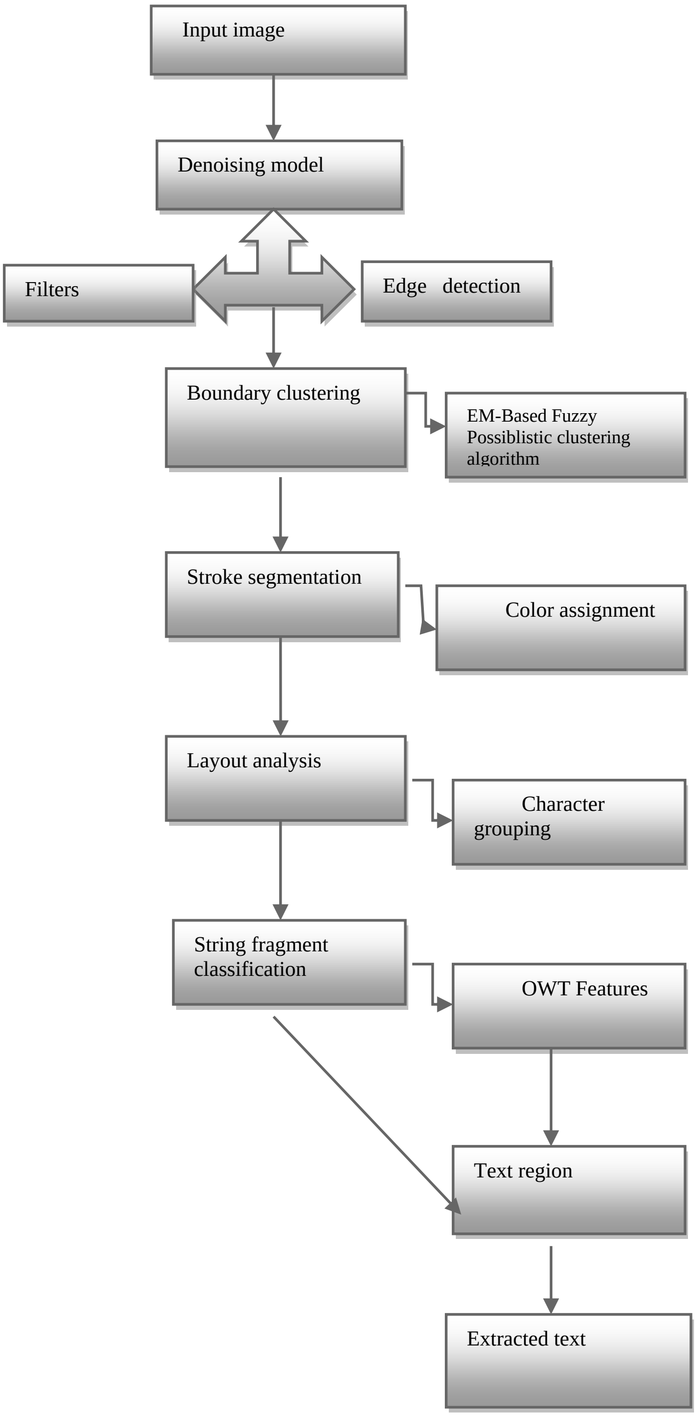


Figure-1 System Architecture

The result of applying an edge detector to an image is a set of connected curves that indicate the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus, applying an edge detector to a car image significantly reduces the amount of data to be processed and filters out non-relevant information, while preserving the important structural properties of an image. Edge detection produces an edge map that contains important information about the image. If the edge detection step is successful, the subsequent task of interpreting and recognizing the information contents of the number plate from the original car image may be substantially simplified. Both the stages described alone are considered as pre-processing in the present research work. These techniques are called pre-processing because they are normally carried out before the real analysis and manipulations of the image data. The main objective is to correct the distorted or degraded image data to create a more faithful representation of the real image

This research work, thus proposes an improved variant of the model, by replacing the median filter with a filter that is more suitable to remove noise from camera captured images. The filters considered to replace median filters are

(i) Gaussian filter

(ii) Wavelet filter

All the three filters selected have been successfully exploited to remove noise from images. The anisotropic diffusion filter is improved by the use of 4th order PDE, followed by any one of the three noise removal techniques. Thus, three new hybrid models are proposed.

1. 4th Order PDE based Anisotropic Diffusion Filter + Gaussian Filter + BayesShrink (AGB Model)
2. 4th Order PDE based Anisotropic Diffusion Filter + Median Filter + BayesShrink (AMB Model)
3. 4th Order PDE based Anisotropic Diffusion Filter + BayesShrink (AB Model)

Proposed Approach

OWT is used for sake of simplicity and efficiency. And it simply consists of applying the lifting steps of a 1D wavelet transform in the direction of the image contours. So image is filtered along with direction. This result is better energy compaction in the lowest subband, compared to the separable wavelet. Transform has similar complexity as the separable wavelet transform while providing better energy compaction and staying critically sampled, which makes it a good candidate for compression applications.

A three dimensional wavelet transform can be accomplished by performing three separate one dimensional transforms. First the image is filtered along the x-dimensional and decimated by two. And it is followed by filtering the sub image along the y-dimension and decimated by two. Finally, we have split the image into four bands denoted by LL, HL, LH and HH after one-level decomposition.

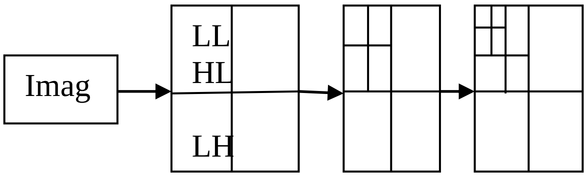


Figure -2: Level One, Two and Three Decomposition

In mathematical terms, the operation or low pass filtering in the inner product between the signal and the scaling function $(\emptyset)$ as shown in the equation 1 whereas the differencing operation or high pass filtering in the inner product between the signal and the wavelet function $(\emptyset)$ as show in the equation 2

$$c_j(k) =< f(t), Q_{j,k}(t) >= \int f(t), Q_{j,k}(t)dt$$

$$d_j(k) =< f(t), \varphi_{j,k}(t) >= \int f(t), \varphi_{j,k}(t)dt$$

The scaling function or the low pass filter is defined as

$$\emptyset_{j,k}(t) = 2\frac{j}{2\emptyset}(2^j t - k)$$

The wavelet function or the high pass filter is defined as

$$\varphi_{j,k}(t) = 2\frac{j}{2\varphi}(2^j t - k)$$

where j denotes the discrete scaling index, k denotes the discrete translation index.

The reconstruction of the image is carried by the following method. First, will up sample by a factor of two on all the four sub bands in each dimension. The sum of four filtered sub bands to reach the low-low sub band at the next finer scale. And then repeat this process until the image is fully reconstructed. Application of DWT divides an image into four sub bands, which arise from separable applications of vertical and horizontal coefficients.

The LH, HL and HH sub bands represents detailed features of the images, while LL sub band represents the approximation of the image. To obtain the next coarse level the LL subband can further be decomposed, thus resulting in the 2-level wavelet decomposition. The level of decomposition performed is application reliant

```
1.  For row=1 to height
        For col= 1 to width
        edgepixel[row]=edgepixel[row]+edge(row,col)

2.  For row=1 to height
        If edgepixel[row]>Tmin
            mean[row]=mean(); //of edge pixel positions
            variance[row]=variance( ) //of edge pixel positions

3.  For row=1 to height
        Find the set of continuous rows satisfying
            variance[row]> maximum variance (Vxmax).

4.  For each band,
        set top= starting row
        set bottom= ending row.
```

Step 1 Algorithm

```
1.  For each band,
        Calculate minimum and maximum values of µx (µxmin and
µxmax)
        Calculate maximum value of vx (vxmax).

2.  For each band,
        Set left = (µxmin - vxmax)
        Set right = ( µxmax + vxmax).

3.  For each band,
        Draw box having diagonal corners (left, top) and (right,
bottom)
```

Step 2 Algorithm

```
1.  For each bounding box in second stage
        From left to right

            Find first prominent vertical edge having
                height > predefined minimum height (Hmin)
            if found
                set new_left=left

        From right to left

            Find first prominent vertical edge having
                height > predefined minimum height (Hmin)
            If found
                set new_right=right

2.  Draw box with left-top and right-bottom corners' coordinates as
(new_left, top) and (new_right, bottom)

3.  Among these new bounding boxes, the overlapped or very close
    bounding boxes are merged to get common bounding box. This
    case actually occurs in case of multi line character set number
    plate.
```

Step 3 Algorithm

IV. RESULTS AND DISCUSSION

The experiments were conducted in three stages. The first stage analyzed the performance of the localization. The second stage analyzed the performance of the preprocessing algorithms that are used to enhance the natural image. The third stage analyzed extraction of the text. To analyze the performance of the various algorithms used in the proposed system, the following performance metrics were used. F-Measure used to evaluate the Filter based feature selection methods. F-Measure is the harmonic mean of the precision and recall. Precision is defined as the number of documents with Keyword *i* in cluster *j* divided by number of documents with Keyword *i* in cluster *j*, and recall is defined as the number of documents with Keyword *i* in cluster *j* divided by number of documents with Keyword *i* .The formula for the corresponding Precision, Recall and F-Measure is shown in the following equations.

$$Precision\,(i,j) = \frac{n_{i,j}}{n_j}$$

$$Recall\,(i,j) = \frac{n_{i,j}}{n_i}$$

$$F - Measure\ = 2\ x\ \frac{(Precision\ x\ Recall)}{(Precision\ +\ Recall)}$$

: Number of documents with Keyword *i* in cluster *j*.

: Number of documents with Keyword *i*.

: Number of documents with Keyword *i* in cluster *j*.

An overall value for F-Measure is calculated by taking the weighted average of all values for F-Measure.

$$F - Measure\ = \sum_i \frac{n_i}{N} max\ _j F(i,j)$$

The Execution time for all the images are tested and compared with 3D-OWT based features which are used. Execution time is calculated using tic toc method. To compare the compression performance, execution time parameter is used.

Table-1 Results for component image

Methods	precision	recall	f-measure	execution time	accuracy
Existing Method	0.67	0.78	0.56	0.75	0.78
proposed method	0.73	0.79	0.76	0.45	0.89

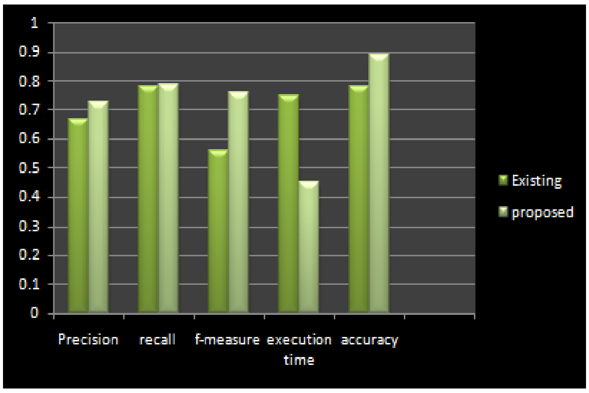


Figure-3: graph results for component image

Table-2 Results for comic image

Methods	precision	recall	f-measure	execution time	accuracy
Existing Method	0.57	0.78	0.56	0.75	0.75
proposed method	0.73	0.78	0.76	0.43	0.88

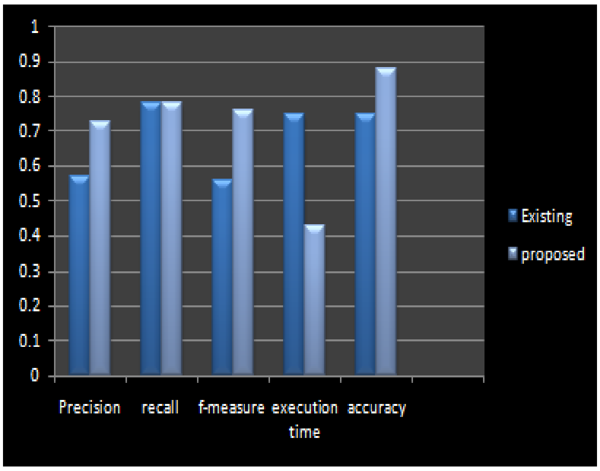


Figure-4: graph results for comic image

Table-3 Results for car image

Methods	precision	recall	f-measure	execution time	accuracy
Existing Method	0.60	0.74	0.65	0.75	0.77
proposed method	0.78	0.82	0.76	0.35	0.89

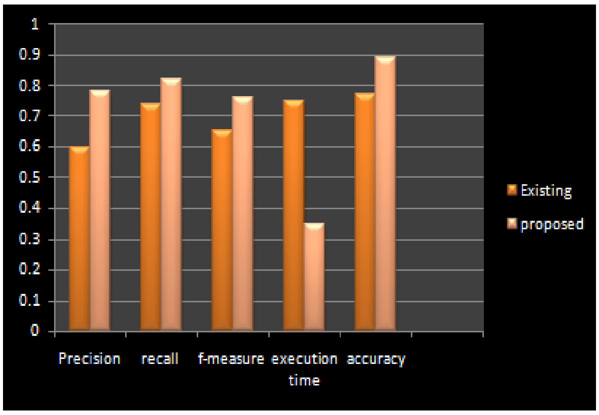


Figure-5: graph results for car image

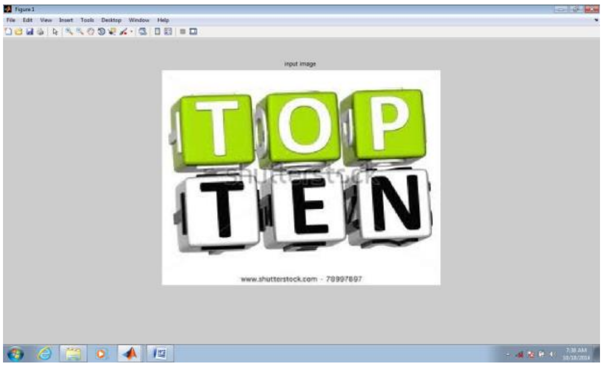


Figure 1
File Edit View Insert Tools Desktop Window Help
input image
TOP
TEN
www.shutterstock.com · 78997897

Figure-6 Input Image

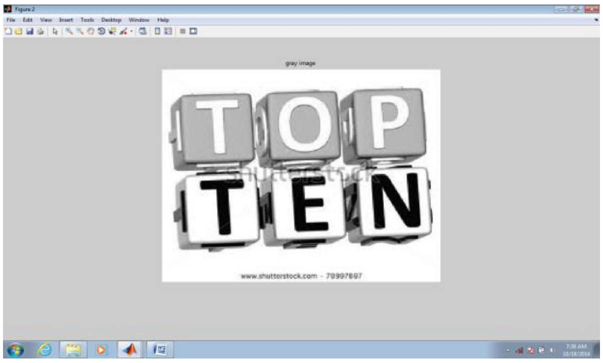


Figure-7 Gray scale Image

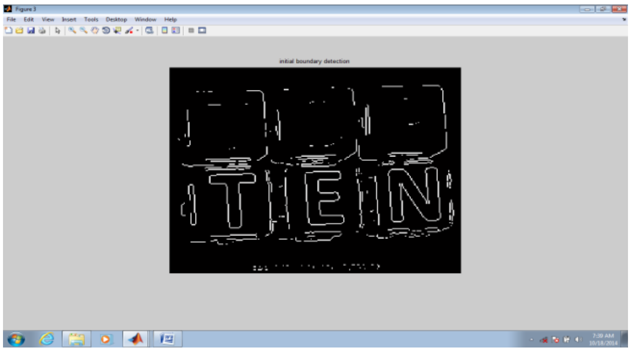


Figure-8 Initial Boundary Condition

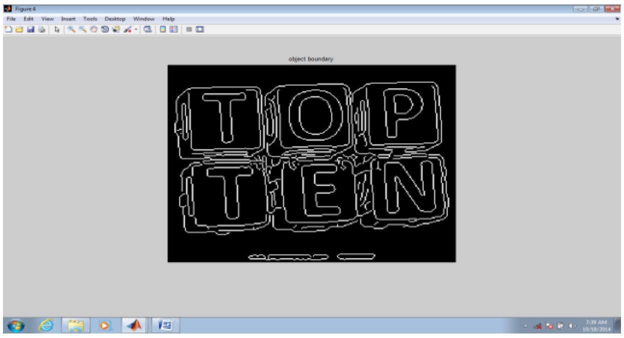


Figure-9 Object Boundary

Figure-10 Text Localization Process

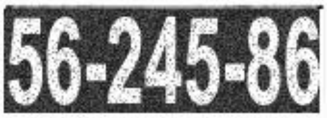


Figure-11 Extracted Text

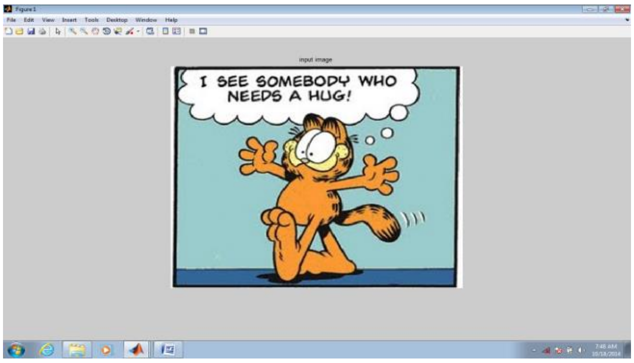


Figure-12 Input Image

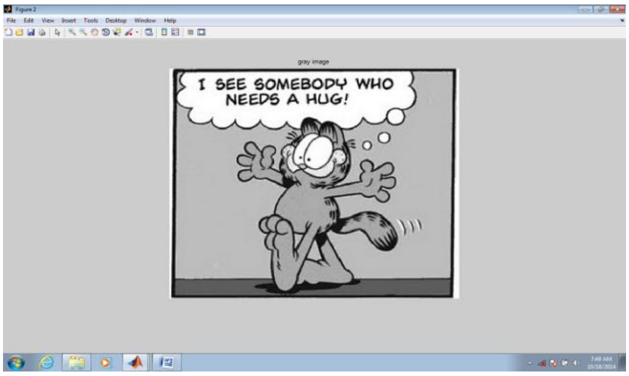


Figure 2
File Edit View Insert Tools Desktop Window Help
gray image
I SEE SOMEBODY WHO NEEDS A HUG!

Figure-13 grayscale Image

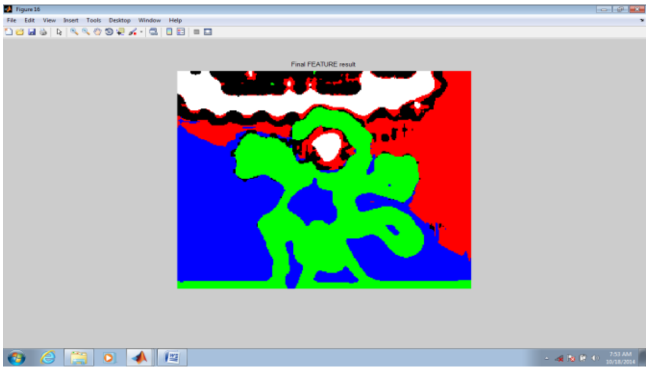
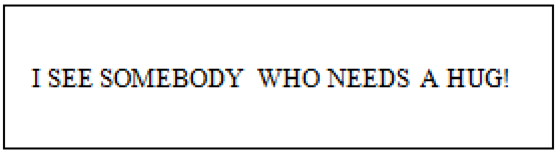


Figure-14 OWT Process



I SEE SOMEBODY WHO NEEDS A HUG!

CONCLUSION

In this paper, we propose a novel framework to extract text regions from scene images with complex backgrounds and multiple text appearances. This framework consists of three main steps: boundary clustering (BC), stroke segmentation, and string fragment classification. In BC, we propose a new bigram-color-uniformity-based method to model both text and attachment surface, and cluster edge pixels based on color pairs and spatial positions into boundary layers. Then, stroke segmentation is performed at each boundary layer by color assignment to extract character candidates. We proposed two algorithms to combine the structural analysis of text stroke with color assignment and filter out background interferences. Further, we design a robust string fragment classification based on 3D-OWT features. The features are obtained from feature maps of gradient, stroke distribution, and stroke width. The proposed framework of text localization is evaluated on scene images, born-digital images, broadcast video images, and images of handheld objects captured by blind persons. Our proposed algorithm attains 94% accuracy in the form of component and critical images.

REFERENCES

1. Chitrakala Gopalan and Manjula.D "Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images" International Journal of Computer Science and Engineering 2:4 2008.

2. Nobuo Ezaki**,** Marius Bulacu et al., "Text Detection from Natural Scene Images:Towards a System for Visually Impaired Persons" vol. II (ICPR 2004).

3. S. A. Angadi and M. M. Kodabagi "A Texture Based Methodology for Text Region Extraction from
Low Resolution Natural Scene Image". vol. II (ICPR 2004).

4. Jan Urban, Jan Vaneˇk et al., "Preprocessing of microscopy images via Shannon's entropy".

5. G. Sahoo1,Tapas Kumar et al.., "Text Extraction and Enhancement of Binary Images Using Cellular Automata" vol 6(3), August (2009).

6.Rama Mohan Babu.G,srimaiyee.p"Text Exraction From Heterogeneous image using Mathematical Morphology" JATIT 2005 – 2010.

7. JiSoo Kim, SangCheol Park,et al "Text locating from natural scene images using image intensities"16 January 2006.

8. M. Praneesh, Dr. D. Napoleon, 2020, Extraction of Text from Compound Images using Ridler Calvard Based Median Filtering Algorithm, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NSDARM – 2020 (Volume 8 – Issue 04).

9. Napoleon, D., et al. "An efficient modified fuzzy possibilistic c-means algorithm for segmenting color based hyperspectral images." *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012)*. IEEE, 2012.

10. Kumar, A., Umurzoqovich, R. S., Duong, N. D., Kanani, P., Kuppusamy, A., Praneesh, M., & Hieu, M. N. (2022). An intrusion identification and prevention for cloud computing: From the perspective of deep learning. *Optik*, *270*, 170044.

11. Napoleon, D., et al. "Self-organizing map-based color image segmentation with fuzzy C-Means clustering and saliency map." *International Journal of Computer Application* 3.2 (2012): 109-117.

12. Praneesh, M., and R. Jaya Kumar. "Novel approach for color based comic image segmentation for extraction of text using modify fuzzy possibilistic c-means clustering algorithm." *Int J Comput Appl IPRC* 1 (2012): 16-18.

13. Boonsatit, N., Rajendran, S., Lim, C. P., Jirawattanapanit, A., & Mohandas, P. (2022). New adaptive finite-time cluster synchronization of neutral-type complex-valued coupled neural networks with mixed time delays. *Fractal and Fractional*, *6*(9), 515.

14. Napoleon, D., Praneesh, M., Sathya, S., & SivaSubramani, M. (2012). An efficient numerical method for the prediction of clusters using k-means clustering algorithm with bisection method. In *Global Trends in Information Systems and Software Applications: 4th International Conference, ObCom 2011, Vellore, TN, India, December 9-11, 2011. Proceedings, Part II* (pp. 256-266). Springer Berlin Heidelberg.