ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 10(17), DOI: 10.17485/ijst/2017/v10i17/114415, May 2017*

Text Extraction and Recognition from the Normal Images using MSER Feature Extraction and Text Segmentation Methods

Nitin Sharma* and Nidhi

Department of IT, UIET, Panjab University, Chandigarh – 140413, India;
sharma.niti243@gmail.com

Abstract

Image mining is concerned with the extraction of contained information, image information connection or other patterns not clearly stored in the images. Text in images is one of the dominant features and its extraction is a big task. If this type of text could be segmented, detected, extracted and recognized automatically, than it would be a precious source of high-level retrieval process. In the research work, text extraction and recognition from the normal images using MSER feature extraction and text segmentation methods has been developed to detect the text regions and the system is based on efficient optical character recognition process. Text extraction and recognition from the normal images is important for content based image analysis. This problem is challenging due to the complex background of images, reflection of light in images and shadow portion presented in images. The proposed technique in this work develops a well-organized text extraction and recognition methods that utilizes the concept of morphological operations using digital image processing. Existing text extraction method, namely, region based method produces enhanced results when applied on the normal images. The advantage of segmentation for the feature extraction of text region is proposed in the system.

Keywords: Binarization, Morphological Operations, MSER Feature Extraction, Recognition and Optical Character Recognition, Segmentation, Text Region Detection

1. Introduction

Text Extraction[1] is connected with the appropriate text data extraction from image collection. Most of the information these days is presented either on photographers/videos form or in paper form. Numbers of applications are there through which the text extraction[2] is helpful. Applications may include, multimedia systems, geographical information system, libraries and information retrieval system. The main aim is the finding of image regions of text that are useful for the OCR (optical character recognition).

Image text has useful information that can help the user to acquire the proper idea of image. Extracting the image is significant for many applications. Because of variability in size, character font, text directions and complex background presence, the character recognition becomes the difficult task. Following are some of the properties of text[3], namely, color, size, edge and motion.

*Author for correspondence

1. Color: Color intensity also has an impact on the text extraction. The process of extraction becomes more effective and simple if the color of the characters would be same.
2. Variation in the text size can be a problem and this can be reduced with the appropriate data through the procedure of text region detection.
3. Edge: These are text consistent features.
4. Motion: Motion is applied to videos with text and it is known as text movement either horizontally/vertically.

The input is received in the form of images and significant text is the output. Images can be colored/gray scale, compressed/uncompressed. Text information extraction is divided into detection, enhancement and extraction, tracking, OCR and localization.

The process of text detection is for determining the text in frame by estimating the text accessible confidence in regions of native images by classification. To decide the text location in image and to generate the bounding boxes in text is the process of text localization. Text tracking can be used to cut back the time period for text localization. The image extracted and enhanced should be converted into binary image before sending it to OCR system. The segmentation of components from the background is the stage of text extraction. Text extraction enhancement is required as an outcome for text regions with less resolution and noise. Later, the images of extracted text are distorted in plain text after applying the OCR system. To convert the printed text into editable text, concept of OCR (Optical Character recognition) method is used. The System allows the machine to identify the text repeatedly. It is a technology for functioning like human reading ability. Conversion of files/texts or captured images by digital camera, scanned paper documents into searchable and editable data. The main benefit of OCR is that it encodes information into a format which is machine-readable as well as human-readable, whereas barcodes and 2D symbols are simply machine-readable[4].

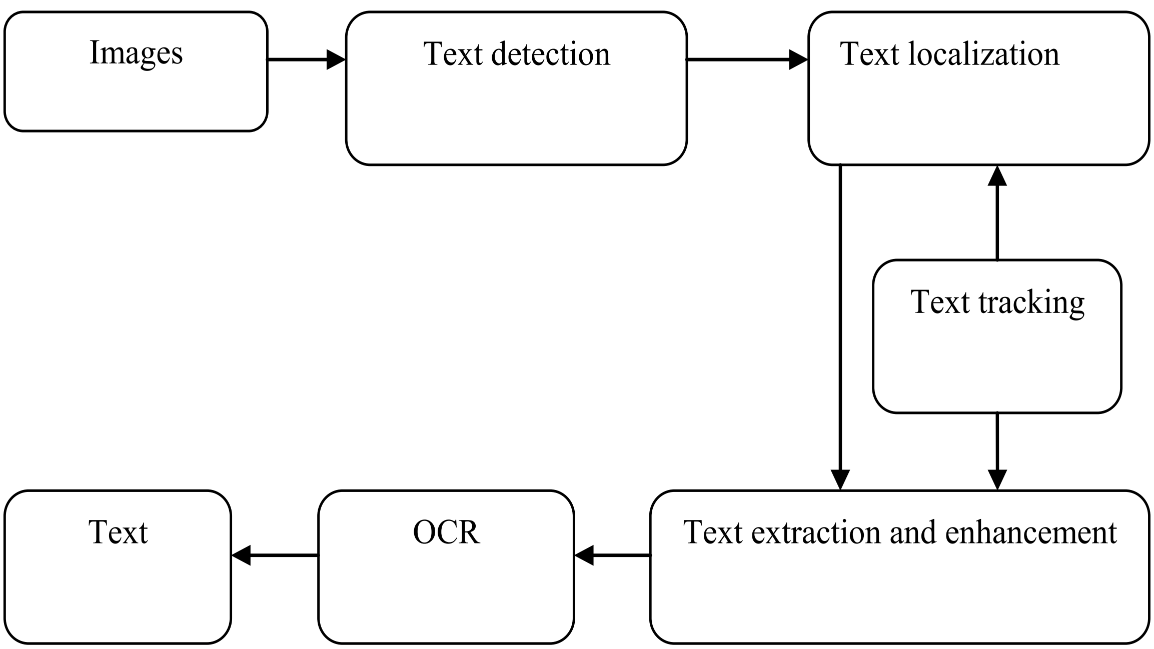


Figure 1. Text extraction system.

2. Techniques for Feature Extraction

By using the image features obtained from the equivalent images taken from varied viewpoints, image stitching is achieved. The feature based techniques extract the distinctive features from every image and match the features meant for establishing an inclusive correspondence. Following are the feature extraction techniques:

2.1 SURF (Speeded Up Robust Features)

SURF algorithm[5] is an enhancement of SIFT (Scale invariant feature transform) algorithm. This technique uses a 'fast multi scale hessian key point detector' for finding the key points. The typical edition of SURF is faster than the SIFT and as per the survey of authors, it is more robust than the other feature extraction techniques. SURF algorithm[6] has three main stages, namely, Interest point detection, local neighborhood and matching.

For detection[7], the technique uses square shaped filters for Gaussian smoothing approximation. If the integral image is utilized than the filtering of the image with a square would be faster.

$$R(x,y) = \sum_{k=0}^{x}\sum_{l=0}^{y} K(k,l)$$

SURF[8] utilizes the Hessian determinant for scale selecting and is calculated by Lindeberg. With a point s=(x,y) in image K, the 'hessian matrix' at s point and scale is,

$$G(q,\sigma) = \begin{matrix} M_{x,x}(q,\sigma) & M_{xy}(q,\sigma) \\ M_{y,x}(q,\sigma) & M_{yy}(q,\sigma) \end{matrix}$$

Where are considered as second order derivative of 'gray scale image'.

The points of interest can be found out at varied number of scales. The images are smoothed with Gaussian filter and sub sampled for getting the high level of pyramid. So, different floors/stairs are calculated for various measures:

$$\sigma_{approx} = Currentfiltersize * \frac{Scaleofbasefilter}{Sizeofbasefilter}$$

Descriptor is used for providing unique and robust explanation for features of image by explaining the intensity distribution for pixels. For achieving the rotational invariance, the orientation of point of interest needs to be finding out. For describing the regions for the point, extraction of square regions takes place that centralized on the point of interest and oriented with the orientation. The matching pairs are finding out by comparing the descriptors form the varied number of images.

2.2 SIFT (Scale Invariant Feature Transform)

SIFT[9] was proposed by David Lowe that has the capacity to distinguish and depict neighborhood picture elements successfully. The essential SIFT estimation includes the following stages:

Scale-space local extreme detection is used for building Gaussian scale space that can be finished by adjustable scale 2D Gaussian operator. The Gaussian image difference is later obtained by lessening the scale in every octave[10].

After establishing the key point competitor, if the low difference or off chance is localized in the edge then it is uprooted as this is not recognized dependably with minimum perspective variety or less changes. Utilization of two thresholds takes place for prohibiting the low contrast points and excluding the edge points.

In the step of orientation assignment, very keypoint assigned more orientation dependent on the directions of local image gradient. The step can be achieved by rotation invariance taken as key point descriptor that can be shown relative to the orientation and later achieves the image rotation invariance.

Key-point[11] is identified as the local minimum / maximum of the result of the Gaussian difference on the scale. The low-contrast candidate points and edge response points along the edges that are later discarded. Thereafter, each pixel in Gaussian difference is compared with its nine adjacent pixels at the same scale and nine corresponding neighboring pixels at each adjacent scale to obtain a Gaussian difference image. If the pixel value is the largest or smallest value among all the compared pixels, it is selected as a key[12].

2.3 BRISK (Binary Robust Invariant Scalable Keypoint)

The inherent difficulty in extracting the right features from the image is to balance two challenging objectives, namely, high-quality descriptions and low computational requirements. Perhaps, the most relevant work to solve this problem is SURF which has been shown to achieve robustness and speed, but according to the survey, BRISK[13] achieves comparable best quality at less computational time. In a nutshell, this work proposes a new method for generating key points from images by using a Scaling Space Key Detection that uses the salience criteria to identify points of interest in the image and dimension. In order to improve the efficiency of the calculation, key points are detected in the octave layer and the middle layer of the image pyramid. The position and scale of each key are obtained in a continuous domain by quadratic function fitting. Application at the Neighborhood of each key is a point located on a suitably scaled concentric circle to retrieve the gray value: the local intensity gradient is processed to determine the characteristic feature orientation. Finally, a directional BRISK[14] sampling pattern is used to obtain the pairwise brightness comparison results that are combined into a binary BRISK descriptor. Once generated, the BRISK key is efficiently matched due to the binary nature of the descriptor. Due to the emphasis on computational efficiency, BRISK also leverages the speed savings provided by the SSE instruction set widely supported by today's architectures. The syntax in MATLAB for brisk is:

$$points\ =\ detectBRISKFeatures(I)$$

Example:

$$points\ =\ detectBRISKFeatures(I, Name, Value)$$

2.4 MSER (Maximum Stable Extreme Region)

MSER[15,16] is used as a method of spot detection in an image. This technique was proposed by author Matas to find the correspondence between image elements from two images with different viewpoints. This method of extracting a combined number of corresponding image elements facilitates wide baseline matching, and it results are better stereo matching and object recognition algorithms[17].

A. Candidate connection component extraction method in ICDAR 2011 and ICDAR2013 competition in the first place, and achieved good performance. If all pixel sets have intensity values greater or less than their outer boundaries, the Extreme Value Region (ER) is well defined as the connected component region on the image. An extremism region is maximally stable when it has a minimum value and the maximum stable extrema region can be considered as a virtually invariant local binarization over a wide range of thresholds. The MSER method is very effective for multi-scale detection and for low-quality connection components with near-linear complexity.The two polarities of the connection components are extracted based on the MSER method, firstly, a black MSER on a white background and secondly, a white MSER on a black background. Although, the MSER method achieves promising recall and even extract most candidate connected components in low quality images and has low precision noise. Due to the MSER structure of the root tree, it is possible to efficiently trim the repeated MSER by searching all MSER sequences. A similar MSER pruning method is used to discard the components of the repetitive noise MSER[18] connection.

B. Noise Candidate Removal Trimming algorithm can effectively remove duplicate MSER. However, any non-text objects, such as bicycles, leaves, etc., can produce noise MSER-connected component areas. A classifier fusion strategy between the enhancement and convolution neural networks is used to filter the noise candidate

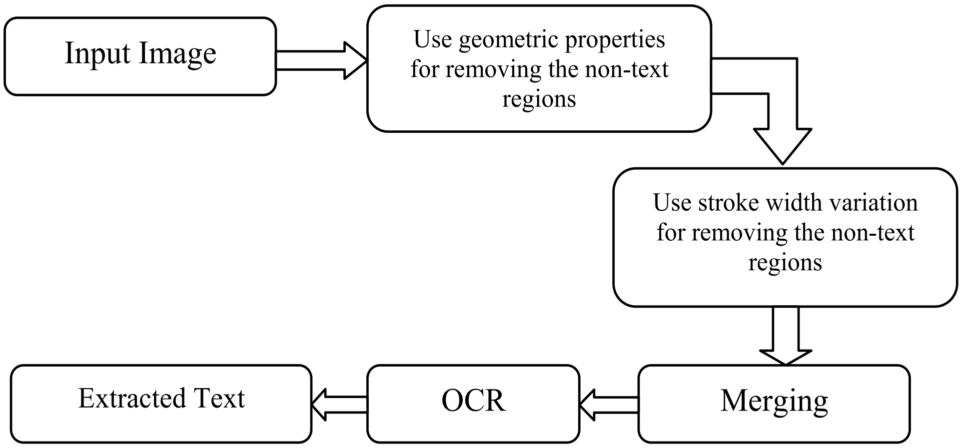


Figure 2. MSER algorithm flowchart.

connectivity components. First, the candidate MSER[19] is classified into text and non-text using a fast enhancement classifier with hand-made geometric features. The non-text MSER is discarded as noise, which can effectively remove a large amount of noise MSER at a small computational cost. Text candidates with high recognition certainty are retained as candidates, and the text is identified with low recognition probability through the CNN, which can learn advanced features to robustly recognize the text MSER from non-text outliers. Finally, the negative noise can be removed and text would be saved.

For using the MSER algorithm[20], the usage of common attributes taken place that are defined below:

1. There should be more edges in the image text.
2. Text width should be more than height.
3. Text is measured in the form of size.
4. Basic text has special texture, except the followed text which is irregular.

Figure 2 is for MSER algorithm[21] and showing the basic flow of the algorithm. It detects the text regions for detecting some of the non text regions. For removing the non-text regions, use of geometric properties takes place. The concept of Stroke width variation would be used if geometric properties are unable to remove non-text regions in that particular image.

3. Working Methodology

To check the effectiveness and accuracy of proposed text extraction and recognition from the normal images using MSER feature extraction with text segmentation technique, the proposed work is simulated. The methodology with algorithms of proposed work is given below:

Step 1: Design and develop a proper GUI of proposed text extraction and recognition from the normal images.
Step 2: Develop a code to upload test normal image for the text extraction and recognition.
Step 3: Apply pre-processing on uploaded image for testing. In pre-processing step, we apply some basic process like Binarization, filtering, resizing, conversion etc. to make the uploaded image useful in simulation.
Step 4: Develop a code for the region detection using the morphological operations and find only text region.

Step 5: Develop a code for the feature extraction from the extracted region of pre-processed image using MSER feature extraction technique. The MSER extraction implements the following steps:

- Remove threshold of intensity from black to white from the binarised image.
- Find out the connected components which are Extremely Regions of image.
- Set a threshold when an extremely region of binarised image is Maximally Stable and suitable for the region extraction.
- Approximate an extracted region.
- Save those regions as feature sets of image.

Step 6: After that, develop the code for the image segmentation on the basis of MSER feature region and segment text from the image and match with the dataset and recognized the letters.

fori=1: length of regions
Segmented_Text(i)=Regions(i)
If feature of Segmented_Text matched with Database Text
Recognized_Text(i)=[ABCDEFGHIJKLMNOpqrstuvwxyzabcdefghijklmnopqrstuvwxyz0123456789.’;”;`~!@#$%^&*()+_-<>?/\|{}]
Save Recognized_Text
end(if)
end(for)

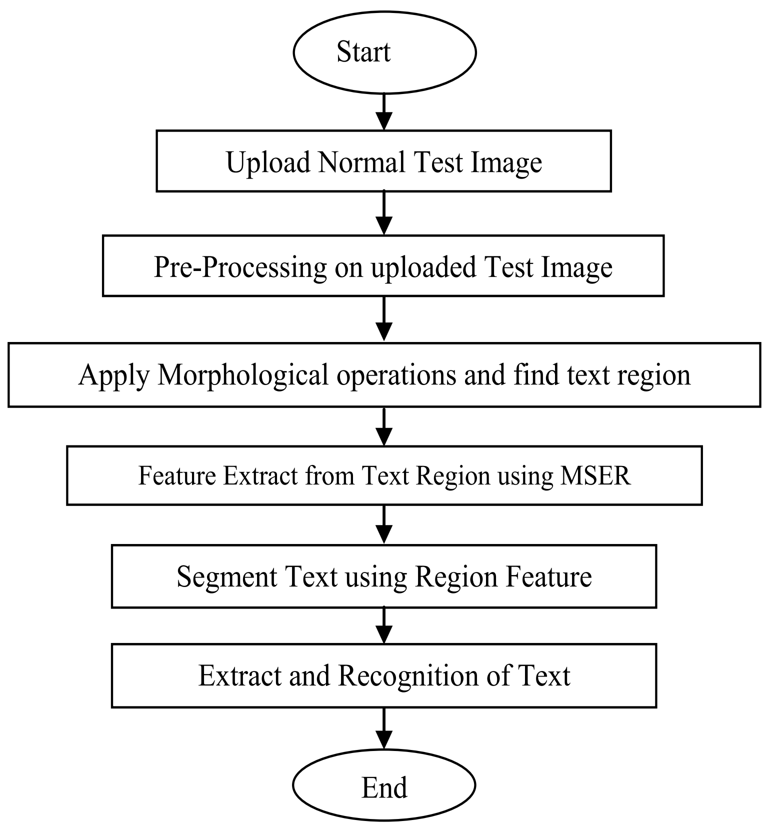


Figure 3. Flowchart of proposed work.

Step 7: After that saves and display the recognized text in the result window of proposed work.

Figure 3 describes the flowchart of proposed text extraction and recognition from the normal images. In this proposed work, we use the MSER feature extraction technique to find out the feature based on the region of the uploaded image. For the simulation of proposed work, the testing is done on several images according to the flowchart. Primarily, we upload the test image and apply the pre-processing on the uploaded image to make a better image. After that, the feature of uploaded image and segment text region is being found out and the extracted text on the basis of their feature is recognized and the recognized results are saved.

4. Result and Simulation

This segment describes the result and simulation of proposed work for text extraction and recognition from the normal images using the MSER feature extraction technique. In proposed work, text region based segmentation is used to segment the text region from the original

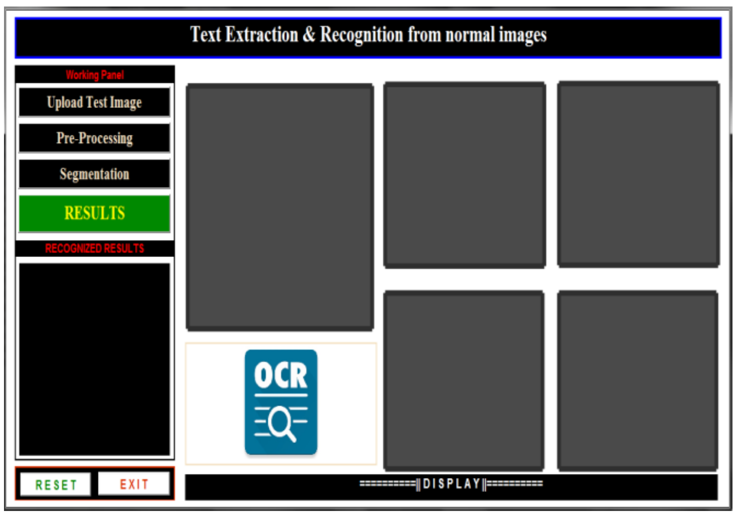


Figure 4. Main figure window.

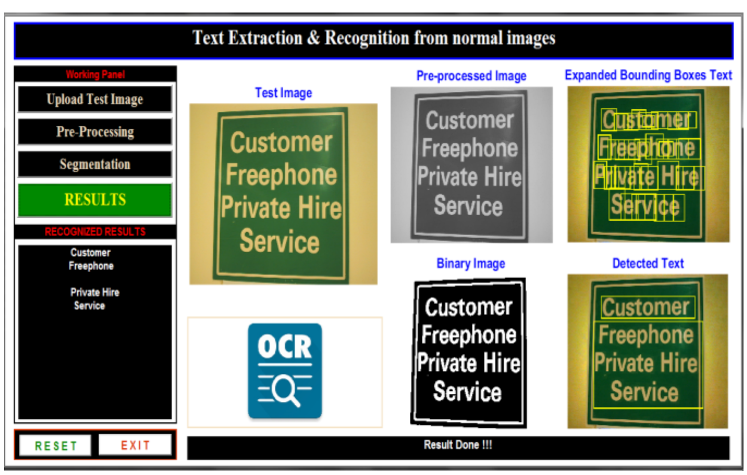


Figure 5. Main figure window with testing.

Figure 6. Extraction and recognition results (a) Original test image (b) Pre-processing of original test image (c) Binarization of pre-processed image (d) Text segmentation and (e) Text extraction

image and on the basis of text region; feature extraction is applied to extract the feature from the image.

From Figure 4 shows the simulation window of proposed work with particular work place. In above figure there are four pushbuttons that are used to simulate the proposed model to extract and recognized the text from the normal images.

Above figure shows the simulation window of proposed work with a test image. In above figure, all pushbuttons are used for simulation of the proposed work in stepwise process. Firstly, the text image is uploaded and the preprocessing is applied and after that the text from pre-processed image using the text segmentation process is extracted. After the simulation, we can see the recognized text in the RESULT panel from the uploaded image and on the basis of recognition; we find the accuracy of proposed work.

Above figure shows the simulation of proposed work with some test images. In above figure;

(a) Original test image which is uploaded in simulation window for text recognition

(b) Pre-processing of original test image

(c) Binarization of pre-processed image to apply morphological operations

(d) Text Segmentation and,

Table 1. Simulation results with accuracy

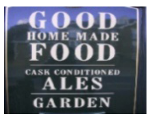
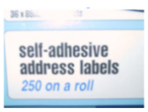
S. No.	Images	Real Text	Recognized Text	Accuracy (%)
1		Customer Freephone Private Hire Service	Customer Freephone Private Hire Service	100
2		GOOD HOME MADE FOOD CASH CONDITIONED ALES GARDEN	$1 GOOD HOME MADE FOOD 1LES GARD8N	58
3		36X89 self-adhesive address labels 250 on a roll	36X88i self adhesive address labels 250 on a rol1	97
4		ARENA LEISURE SERVICES LIMTED Registered Office	ARENA LE1SURE SERVICES LIMTED Registered Office00	92
5		LLOYDS BANK	LLOYDS BANK	100
6		Footpath To Colchester and Greenstead	Footpath To Colchester and Greenstead	100
7		Khalsa Care Foundation Sikh Gurdwara	C' is it Khalsa Care Foundation T Sikh Gurdwara ***T;9	100

Table 1 Continued

8		YOU ARE HERE	YOU ARE HERE	100
9		RED LION YARD SHOPS HIGH STREET	RED YARO L1ON SHOPS H1O8 STREET	85
10		Peacocks	Pe8cock8	75

(e) Text Extraction, are shown.

Form the above observations, we find out the average accuracy of proposed work for text extraction and recognition from the normal images. At first, the original color image is converted into a gray level image. The image dimension decreases both size and resolution of an image and this is useful for speedup of extraction and recognition process. The text assurance value of the pixel and the scale value are calculated by geometric and morphological steps. An original image with text is the input to the proposed system. The input color original image is converted into gray image. The aim of pre-processing is an improvement of the original image information that suppresses unwanted noise or enhances some for further processing or binarization of gray image.

The proposed experiment show that applying traditional OCR technology directly leads to poor recognition rates in case of distorted image. Hence, efficient extraction and segmentation of text characters from the background is necessary to fill the gap between images. Optical Character Recognition (OCR) is a technique to locate and recognize text stored in an original image and convert the extracted text into a computer recognized appearance. OCR systems can only agreement with printed texts alongside clean backgrounds and cannot handle texts embedded in shaded, textured or complex backgrounds. From Table 1, the main aim of this proposed work is to propose the module that extracts the textual information present in images and recognized the text. It processes images based on their pixels values to classify the text blocks and recognize the text information present in images. The text segmentation, detection, extraction and recognition stage of the proposed work uses an efficient feature extraction approach and segmentation approach and template matching technique to recognize the text. The processing step of the proposed system is shown in Figure 3.

From Figure 7 the comparison of accuracy between proposed and previous work is represented in the form of graph. In graph, X-axis denotes the number of images which are included in the proposed work and Y-axis denotes the accuracy of proposed work in percentage. In above figure, blue color line graph represents the proposed

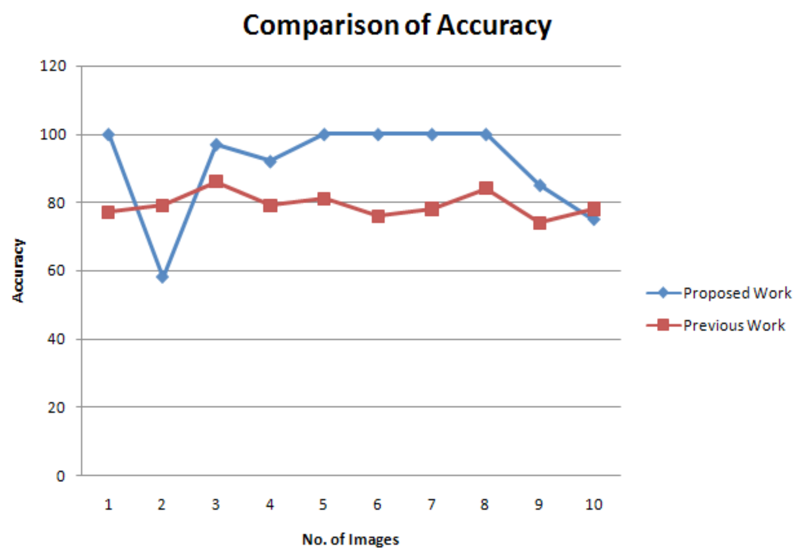


Figure 7. Accuracy of proposed work.

work and the red color line graph represents the previous work. Form the above graph, it has been observed that the average percentage of accuracy is more than 90% with normal images and mostly 100% accuracy is achieved for the database images for proposed work and it is better than the previous work.

5. Conclusion

This proposed work has discussed the extraction and recognition of text from normal images. Text appears in normal images is one feature that gives insight into image content. Text segmentation is a significant issue in the text extraction and recognition analysis research area. Text extraction and recognition plays a significant role in content based retrieval and storage systems. In the proposed work, text extraction and recognition is based on combining efficient segmentation and connected component techniques within the image and text recognition is based on template matching techniques based on OCR system. The proposed technique has been extensively tested on different types of normal images. The results were promising, almost all the text regions could be extracted and recognized from uploaded images and the proposed technique have been compared with the existing methods and the outcome shows that the proposed technique performs better than existing techniques. The accuracy of proposed work is more than 90% and for most of the case recognition result is near about 100%.

6. References

1. De Jesus M, Guimaraes SJF, Do Patrocinio Jr ZKG. Video text extraction based on image regularization and temporal analysis. IEEE International Symposium on Multimedia, Dana Point CA: 2011. p. 305–10. CrossRef.
2. Patel R, Mitra SK. Extracting text from degraded document image. Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna: 2015. p. 1–4. CrossRef.
3. Devi GG, Sumathi CP. Text extraction from images using gamma correction method and different text extraction methods — A comparative analysis. International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai: 2014. p. 1–5. CrossRef. PMid:24561215

4. Kumuda T, Basavaraj L. Detection and localization of text from natural scene images using texture features. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai: 2015. p. 1–4. CrossRef.
5. Zhou Z, XiaowenOu, Xu J. SURF feature detection method used in object tracking. International Conference on Machine Learning and Cybernetics, Tianjin: 2013. p. 1865–8.
6. Juan S, Qingsong X, Jinghua Z. A scene matching algorithm based on SURF feature. International Conference on Image Analysis and Signal Processing, Zhejiang: 2010. p. 434–7.
7. Cai P, Wang, Liang YH. Fast image stitching based on improved SURF. IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanchang: 2016. p. 411–6.
8. Qian S, Chan-juan L, Hai-lin Z, Ying L, Tong-tong C. SURF Feature Description Method of Color Image Based on Quaternion. 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), Krakow: 2015. p. 507–11. CrossRef.
9. Harkat H, Elfakir Y, Bennani SD, Khaissidi G, Mrabti M. Ground penetrating radar hyperbola detection using Scale-Invariant Feature Transform. International Conference on Electrical and Information Technologies (ICEIT), Tangiers: 2016. p. 392–7. CrossRef.
10. Kabbai L, Azaza A, Abdellaoui M, Douik A. Image matching based on LBP and SIFT descriptor. IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15), Mahdia: 2015. p. 1–6. CrossRef.
11. Narhare AD, Molke GV. Trademark Detection Using SIFT Features Matching. International Conference on Computing Communication Control and Automation, Pune: 2015. p. 684–8. CrossRef.
12. Borg NP, Debono CJ, Zammit-Mangion D. A single octave SIFT algorithm for image feature extraction in resource limited hardware systems. IEEE Visual Communications and Image Processing Conference, Valletta: 2014. p. 213–6. CrossRef.
13. Leutenegger S, Garitachli M, Siegwart RY. BRISK: Binary Robust Invariant Scalable Keypoints. IEEE International Conference on Computer Vision 978-1-4577-1102-2/11/$26.00 c 2011. IEEE.
14. Li, Li H, Söderström U. Scale-invariant corner keypoints. IEEE International Conference on Image Processing (ICIP), Paris: 2014. p. 5741–5.
15. Leutenegger S, Chli M, Siegwart RY. BRISK: Binary Robust invariant scalable keypoints. International Conference on Computer Vision, Barcelona, 2011. p. 2548–55. CrossRef.
16. Tao L, Jing X, Sun S, Huang H, Chen N, Lu Y. Combining SURF with MSER for image matching. IEEE International Conference on Granular Computing (GrC), Beijing: 2013. p. 286–90. CrossRef.
17. Geng Z, Zhuo L, Zhang J, Li X. A comparative study of local feature extraction algorithms for Web pornographic image recognition. 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing: 2015. p. 87–92. PMid:26136499 PMCid:PMC4535587
18. Kimmel R, Zhang C, Bronstein A, Bronstein M. Are MSER Features Really Interesting?. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011 Nov; 33(11):2316–20. CrossRef.
19. Ali A, Pal R. Detection and extraction of pantograph region from bank cheque images. 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida: 2016. p. 498–501. CrossRef.
20. Adlinge G, Kashid S, Shinde T, Dhotre VK. Text Extraction from image using MSER approach. International Research Journal of Engineering and Technology (IRJET), 2016. p. 2453–7.
21. Mammeri A, Boukerche A, Khiari EH. MSER-based text detection and communication algorithm for autonomous vehicles. IEEE Symposium on Computers and Communication (ISCC), Messina: 2016. p. 1218–23. CrossRef.