

A Hybrid Method for Text Extraction from Mosaiced Image of Text Dominant Video

P Nagabhushan¹, Vimuktha Evangeleen Jathanna²

^{1,2}Department of Studies in Computer Science
University of Mysore, India

Abstract— One method of building a text file from text dominant video, is by extracting the text from each individual frame and then resorting to a post processing method which overcomes the difficulty of the presence of common text portions in adjacent frames. Towards alleviating the post processing difficulty, an alternative is to mosaic the adjacent video frames and extract the text from the mosaiced image. In this paper we propose a hybridized approach based on edge clustering and connected components for text extraction and localization from mosaiced image generated from text dominant video.

Keywords— Text dominant video, image mosaicing, mosaiced image, edge, clustering.

I. INTRODUCTION

The present day mobile phones, smartphones, PDAs, IPADS are integrated with digital cameras that have the ability to capture high quality images. The trend of using these cameras has also facilitated to capture wide variety of images and videos containing different contents like event, scene, text, and so on. These integrated digital cameras because of simple click and capture mechanism, high resolution, portability, and handiness, are also employed to capture text rich documents available at any source as single static image or video. This has eventually replaced traditional scanners which could scan target document available at that source and provided that such a document could be feed able into the scanner.

A text dominant video captured through digital cameras contains huge amount of text captured from different sources like text books, journals, newspaper, pages. The text from these text dominant videos if harnessed efficiently provides a much truer form of text information for content retrieval and content understanding. One method of extracting and localizing text from text dominant videos is to extract text frame by frame. i.e as in [1] we know video is made up of static images called Frames, and we can process each frame considering it as a single static image. In frame by frame text extraction, the text processing algorithms are applied to each frame, the recognized results are stored in individual notepad files and at the end, all the notepad file contents are merged into a single notepad file. The cost of localizing and recognizing the text information from each individual frame as in case of frame by frame text extraction becomes computationally expensive and this involves the post processing of updating the contents of the notepad with

the arrival of textual information from every next frame. Hence, an alternate method is to mosaic each individual frames and then extract the text from the mosaiced image in one go.

In [2] and [29] the authors have exploited the property of temporal and spatial redundancy present in text dominant video and proposed a method to mosaic the overlapping text portions present in the adjacent frame. The mosaicing approach is useful in case of text rich documents containing different text properties like multiple lines, multiple columns and text with smaller fonts. Mosaicing of adjacent frames is also useful when there exist different camera movements like horizontal right to left track movement, top to bottom/ bottom to top pedestal or in hybrid scan patterns as discussed in [2] and [29].

This paper aims at producing a text file from the mosaiced text image which resulted from mosaicing of the adjacent frame contents of a text rich document video proposed in [2] and [29]. We describe an approach for detecting, localizing, extracting and recognizing text from the mosaiced document video image using edge based clustering and connected components. Figure 1 and figure 2 depict the mosaiced image from the different text rich document video.

The paper is organised as follows: in section II a brief coverage of related literature is given. In section III we have discussed the complexities present in the mosaiced images. . In section IV the text processing in mosaiced images are discussed. In section V a detailed description of the proposed Hybrid approach is provided. In Section VI the experimentation results, performance analysis and execution time for the proposed hybrid approach are discussed. In section VII we conclude the paper.



Figure 1 Mosaiced image from Lak_LTRtrack1 video [2]

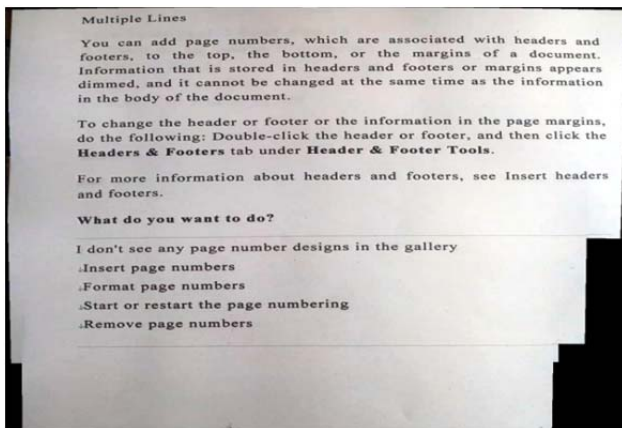


Figure 2 Mosaiced image from Mult_PTb1 video[29]

II. LITERATURE SURVEY

As stated in [1],[14], there are several ongoing research projects on camera based document analysis which includes text detection, extraction, enhancement, recognition and its applications. A brief appraisal on the literatures is given below

Nagabhushan P et al in [2] and [29] proposed a vertical and horizontal strip based mosaicing technique based on SIFT for track movement and pedestal movement videos. The reference frame were matched initially with the other adjacent frames and then the vertical strips were created. The false matches from SIFT were fitted using RANSAC and a simple transformation blending function was proposed

P. Nagabhushan et al. [3] proposed a hybrid approach that combines connected component analysis and texture feature analysis. Canny detector is used to get the possible text regions. Connected component analysis is used to get the candidate text regions. Due to background complexity non-text regions are also present. So to remove non-text regions texture feature analysis is done. The proposed method can handle document images with complex background effectively.

Christian Wolf et.al. in [4] proposed a method for content based image and video retrieval in multimedia documents. The algorithm for this approach used morphological post processing to detect the text and a measure of accumulated gradients by multiple frame integration for artificially embedded text in videos

Many approaches for text extraction by mobile cameras from natural scene have been proposed. A focus mechanism using a hint on the location of target text was adapted by Egyul Kim et.al. [5]. This method adapts pixel sampling and a mean shift algorithm within the focused section and a HCL distance measure is applied to compare target seed color.

Fast and robust text detection in images and video frames was proposed by Qixiang Ye et.al. [6] It is beneficial for real-time video content analysis in complex background. In this method the wavelet energy feature was calculated to locate all possible text pixels and a density growing

method was used to connect the pixels. Finally an SVM classifier was used to identify true text regions. The above stated method lack in finding multiple line of text regions.

Low resolution character recognition for video text images was proposed in [7] by Ataru Ohkura et.al. The approach used subspace method for recognition of characters which were integrated from multiple low resolution characters by super resolution technique. The method lacks in using the structure of characters for low resolution character recognition.

A new approach for video text detection was proposed by Min Cai et.al.[8], which is based on invariant feature analysis such as edge strength, edge density and horizontal distribution. Non-text regions are eliminated using low threshold. A coarse-to-fine detection is applied to locate the text component efficiently. The proposed method is efficient to handle different contrast, font size, and complex background.

Many research works have been carried out on Video retrieval and indexing based on textual queries [9]-[12], A supervised learning method based on color and edge information is used to detect text regions in these cases.

Different approaches for text processing in video frames with complex background were proposed by Fang Liu et.al.[13] and Shuicai Shi et. al.[15] in which density based histogram segmentation and block change rate approach for clustering candidate characters by integrating spatial connectivity and color feature of character pixels is employed.

A method for segmenting and recognizing text embedded in video and images was proposed in [17] by Jean-Marc and Datong Chen. This method used multiple segmentation of same text region producing multiple hypotheses based on Markov random field and eliminating background by connected component analysis and by enforcing a more stringent constraint called GCC by using ID-median operator.

An edge based text verification approach for content retrieval for video caption text was proposed by Jian Zhang in [18]. The verification was based on the investigation of the relation between the candidate blocks and the neighbor areas. The proposed method detects and removes background edges from both background and text blocks and also refines the text boundaries.

A sequential Monte-Carlo video segmentation method was proposed in [19] by Datong Chen et.al., which used a probabilistic algorithm for segmenting and recognizing text of any gray scale value embedded in video sequences based on practical filter by setting weighted samples and refines it by random sampling under temporal Bayesian framework.

Silvio Ferreria et.al. in [20] proposed a technique to isolate Text areas by applying texture segmentation technique with fuzzy estimation to compute text orientations of mobile camera text applied in a frame work for layout analysis.

An automatic video text detection, localization and extraction approach in videos was proposed by Chengjun Zhu et.al [21]. In this approach Support Vector Machine was employed to partition video blocks into text and non

text regions. Both artificial text and scene text were considered in this approach. Also due to the employment of correlation of videos using special symbol, detection of false regions were effectively reduced

It is evident based on the literatures that lot of related works have been carried out to extract and recognize text from document, natural scene and artificially embedded texts present in videos and images. The problem we are proposing is a unique problem based on text extraction and recognition from mosaiced images. The detailed procedure for extraction and recognition of text is given in next sections.

III. COMPLEXITIES IN PROCESSING MOSAICED IMAGES

Mosaicing of adjacent video frames refers to the process of stitching common text portions to form one large image composite. In case of text rich document video, the mosaiced image is formed from minimum of 2 to maximum of at least 40-50 adjacent frames that are resulted after frame selection, frame registration and other pre-processing stage proposed in [2][29]. The resultant mosaiced image contains certain intricacies which affect the image quality. Figure 1 is a classic example depicting the intricacies present in the mosaiced images. The complexities present in the mosaiced images are

1. Seam line: seam lines are the stitching lines resulted from low cost affine based translation blending function representing each / part of (strip based method proposed in [2] and [29]) adjacent image frame. The seam lines exhibit strong edges and while applying text processing algorithms these edges are also detected. Hence, elimination or suppressing it becomes a pre-processing problem during text extraction. Figure 1 depicts the seam lines present in the mosaiced image.
2. Illumination changes: during frame stitching process the variations in lighting during video acquisition affects the quality of mosaiced images.
3. Color shifts: this occurs due to lens distortion and camera motion during image acquisition and the mosaiced image appears as if it is captured from different sources.

IV. TEXT PROCESSING IN MOSAICED IMAGES

The text rich mosaiced image resulted from frame mosaicing process of text dominant video proposed in [2], is utilized to detect, extract and recognize text to produce a text file. As illustrated in the literature survey, lot of research work has been done based on - texture, connected component and edge component to localize and extract the text. The process of text detection from any camera captured images can be performed using edge based, region based or texture based approaches. Text detection and extraction finds text and non text regions in an image. Text localization is done by finding the connected components in the separated text regions with bounding boxes. The recognition is done with the help of available OCR. The same procedure can be applied on the mosaiced images by altering the conventional text processing approaches due to the intricacies present in the mosaiced images.

The conventional text processing approaches like projection based, connected component based, edge based segmentation cannot be applied directly on the mosaiced

images to extract and localize text regions, due to the complexities present in the mosaiced images that are discussed in the previous section.

Horizontal projection technique cannot be applied since the projection of mosaiced image may be cylindrical, rectilinear, or slant due to different camera movement. Connected component based method assumes that character in an image has uniform color while mosaiced images exhibit different color shifts. In case of edge based method, it assumes that character are made up of edge strokes hence retrieval of edges is important constraint in edge based method. In case of mosaiced images, we cannot only utilize either Edge based methods or Component (CC) based methods to detect and localize texts due to several artifacts like mosaicing noise, seam lines, color shifts. This is because, the 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, we have used hybrid approach based on a combination of edge, clustering and connected components that can localize and extract text from mosaiced images. This results in a unique way of handling mosaiced images. The edge based clustering initially finds edges using Sobel operator and the detected edges based on eccentricity, similar color and intensity histograms are grouped together. In this way accurate text regions are identified. Then connected components are applied on the clustered image to filter text and non-text regions. The detailed description of the proposed approach is given in the next section.

V. PROPOSED HYBRID APPROACH

The text detection, extraction and recognition phase for the proposed hybrid method uses both edge-based clustering and connected component based methods. The images are processed based on the pixel values to identify potential text blocks and recognize the text information. The text detection and extraction method is performed efficiently based on edge-based clustering and segmentation approach. The proposed text processing system for extracting text from mosaiced image using hybrid approach consists of six steps as depicted in the figure (3).

A. Pre-Processing

The aim of pre-processing is to improve the mosaiced image by suppressing unwanted distortions and enhancing the image for further processing. The pre-processing stage in the hybrid method includes 3 stages.

1. Color mosaiced image to gray level conversion
2. Noise removal using median filter
3. Image Cropping
4. Edge Detection

1) Color Mosaiced Image to gray level conversion

As the first step of pre-processing color image is converted into gray image. There are many functions in image processing to convert the true color image to the grayscale

intensity image by eliminating the hue and saturation information while retaining the luminance. The gray images are easier to process and edge detection is simpler. One method to convert the true color image to gray image is by extracting the red, green and blue channels and then using the weighted average as given in equation 1 and 2.

$$\left. \begin{aligned} \text{redChannel} &= \text{color}(:, :, 1) \\ \text{greenChannel} &= \text{color}(:, :, 2) \\ \text{blueChannel} &= \text{color}(:, :, 3) \end{aligned} \right\}$$

$$5.1 \quad \text{gray} = .299 * \text{redChannel} + .587 * \text{greenChannel} + .114 * \text{blueChannel} \quad 5.2$$

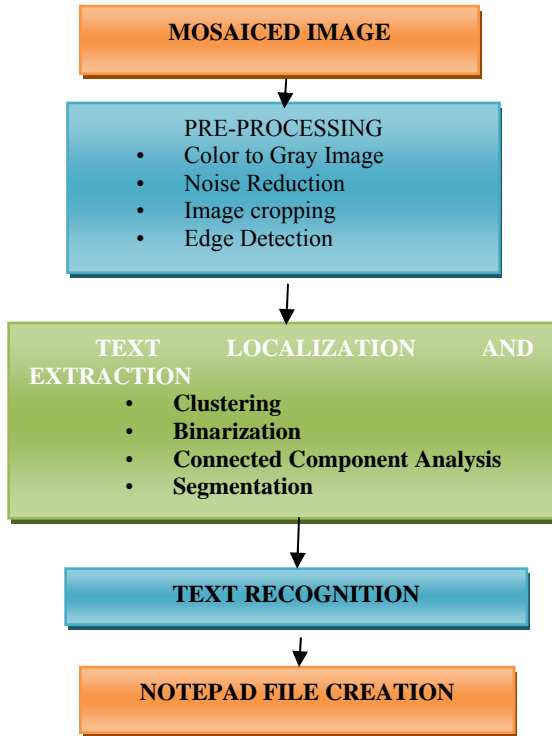


Fig.3 Block diagram of Proposed Hybrid Approach.

2) Noise Reduction

The aim of pre-processing is to improve the image quality by suppressing unwanted distortions and enhance the input image for further processing. In case of mosaiced images other than Gaussian noise and signal noise from acquisition devices, there exist seam lines due to stitching inspite of high cost blending function used in [2] and [29]. We consider the seam lines also as noise and suppress it by applying noise removal or noise reduction algorithm. The noise filter reduces noise and seam lines by smoothing the image and hence softening the edges. We have used Median filtering for removing noisy outliers without minimizing the sharpness of the image. The **median filter** is a nonlinear digital filtering technique, often used to remove noise. Noise reduction is a typical pre-processing step to improve the results of edge detection and text extraction. Median filtering is used to preserve edges while removing noise



Fig 4 Noise Suppressed Mosaiced Image

3. Image Cropping

The mosaiced image exhibits the null black space around the mosaiced image due to projection and blending. This region interferes with the text processing approach based on color clustering. Hence this region is cropped by applying watershed shed segmentation and a cropping functions defining an area around the text region. The resultant mosaiced image is given in Figure 5



Fig.5 Cropped Mosaiced Image

4. Edge Detection

Once the seam lines and other noise are suppressed, the proposed system utilizes the Sobel edge filter for extracting text. Edge detection helps in detecting possible edges and boundaries in a mosaiced image based on the fact that text has boundaries. This helps in filtering out the irrelevant data from the image, while maintaining the basic structural features in the image. The text from the mosaiced image can be detected accurately using Sobel edge map. The Sobel detector contains two coordinates namely, x-coordinate and y-coordinate. The x-coordinate will do increasing “right-direction” and the y-coordinate will do increasing “down-direction” for detecting the text in the image. The Sobel edge detector performs 2D convolution operation and then finds approximate gradient magnitude. This edge detector uses 3 x 3 convolution masks for detecting text in the image. The mask slides over the whole image and square of pixels manipulate at a time to detect the text. By using this detector, the text from the mosaiced image is detected and noise is removed slightly. The result of edge detection is given in figure.



Fig.6 Edge detected mosaiced image before noise removal and cropping

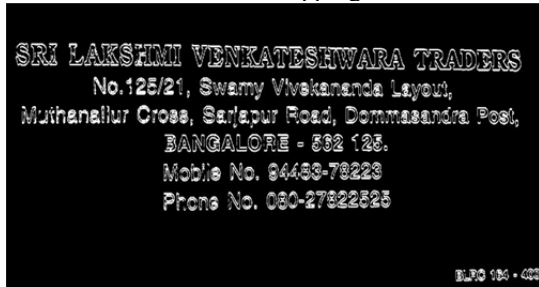


Fig.7 Edge detected mosaiced image after noise removal and cropping

B. Text Localization and Extraction

In any text processing system text extraction is a critical step as it sets up the quality of the final recognition result. To identify foreground text characters and background / non text regions from the mosaiced image we have considered using K-means clustering. K-means clustering groups the features based on distance measure. Generally, the mosaiced images contain color shifts due to lens distortions, illumination changes and camera movement. This color shift also gets reflected on the gray image with different gray values on both foreground and background. The source images considered in [2] and [29] generally contains white background with black characters. The mosaiced image is not easy to segment due to the intricacies present. Hence we have combined 3 features together. The first feature we have considered is the color in RGB color space. The color histogram for an image is constructed by quantizing the colors within the image and counting the number of pixels of each color. The histogram based method is very suitable for mosaiced image retrieval because they are invariant to geometrical information in images, such as translation and rotation.

The second is to calculate intensity of the image in gray level edge image and the third is the eccentricity i.e is the shape feature of the character edge. Depending on these three features 2 clusters are created one representing foreground that is the text and the other representing background based on the similarity score calculated using Euclidian distance measure. The three features with K-means clustering accurately segments and extract text and non text region. Then a connected component classifier can be used to classify text and non text regions. The algorithm to extract color feature, edge and intensity is as given below Algorithm to extract the features of mosaiced images after pre-processing –

Step 1 : Start

Step 2 : Calculate the eccentricity of the edge image.

Step 3 : Calculate mean of each color space i.e. red, green and blue

Step 4 : Calculate the intensity histogram of gray levels

Step 5 : Label the image.

Step 6 : Store eccentricity, intensity and mean of each color space as vector in the feature matrix.

Step 7 : Store label of image and index of image as vector in other matrix.

Step8 : Stop

The combined features are given to the clustering algorithm to group the similar feature. A brief note on the working of K-means clustering is given in the following sub section.

1) K-Means Clustering

The k-means is basically a clustering algorithm which partition a data set into cluster according to some defined distance measure [27][28]. K-means clustering algorithm is an unsupervised clustering protocol [27] K-means clustering is a technique that partitions the objects into K mutually exclusive clusters, by maximizing the inter cluster distances and minimizing the intra cluster distances. Each cluster is described by its centroid. However, the k-means algorithm is highly dependent upon the value of k i.e. number of clusters in the partition. So the number of k has to be defined before clustering. The optimum number of clusters may vary from image to image. The K-means function is given in (1).

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \tag{1}$$

where there are k clusters S_i , $i = 1, 2, \dots, k$ and μ_i is the centroid or mean point of all the points $x_j \in S_i$.

The working principle of the k-means clustering algorithm has two phases. In the first phase we select k centers randomly. We have selected $k = 2$ that represent foreground character and background. In the next phase we measure each data feature to the nearest center. We have used Euclidian distance measure to determine the distance between each data feature and the cluster center. The Euclidian distance measure is as given below

$$D(X_i, Y_i) = \sqrt{\sum_{l=1}^m (X_l - Y_l)^2} \tag{2}$$

When the entire data features are included in any cluster, the initial group is formed. Again the same process is performed to recalculate the cluster. The new centroid for each of the clusters is evaluated using

$$\mu_i := \frac{\sum_{i=1}^m 1\{c(i) = j\} x^{(i)}}{\sum_{i=1}^m 1\{c(i) = j\}} \tag{3}$$

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over all the data features, j iterates over all the centroids and μ_i are the centroid of data feature. The following steps are iterated until the cluster labels of the image do not change anymore.

The algorithm is as given below

// Algorithm for K-means

Step 1: Randomly choose number of clusters K.

Step 2: Compute the data features and store it in a data

matrix.

- Step 3: Randomly choose K data features.
- Step 4: Centroids are calculated for data features.
- Step 5: Now, compare feature to every centroid and assign pixel to closest centroid to form a cluster using Euclidian distance formula.
- Step 6: When all features are assigned, initial clustering has been completed.
- Step 7: Recalculate position of centroids in K clusters.
- Step 8: Repeat step 5 & 6, until centroids no longer move.
- Step 9: Image separated into K clusters.

Figure 8 represents the result from k-means clustering. The blue color depicts the background and red represents the detected characters.

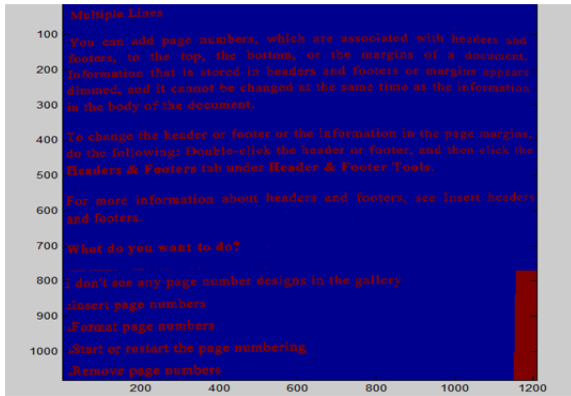


Fig 8. Result from K-means Clustering

3) Text Extraction

After text detection process, the text region from the mosaiced image can be localized. A connected component classifier is applied on the clustered image to localize and extract text region.. The precise location of text in image can be indicated by bounding boxes. The figure below depicts the result of segmentation.

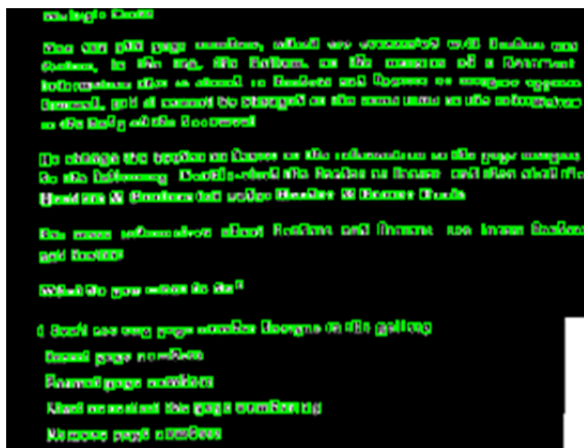


Fig.9 Text Extracted Image from Mosaiced Image.

3) Text Recognition

The extracted text is recognized using the commercially available Tesseract OCR engine to recognize the text. The efficiency of the proposed hybrid approach is around 97%. The performance analysis is done in the next section. The recognized text is stored in the text file.

VI. EXPERIMENTS

The software module was implemented using Matlab 2014(a). We tested reading English printed text containing single, multiple lines, multiple columns with different font styles size and spacing of text which we had created for the experimentation purpose. Different camera movements like horizontal track, vertical pedestal and hybrid scan patterns were also considered. The performance analysis and the execution time for the proposed approach are discussed.

1) Performance Measures

Performance is analysed for the mosaiced text rich document video frames. Metrics used to evaluate the performance of the block 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:

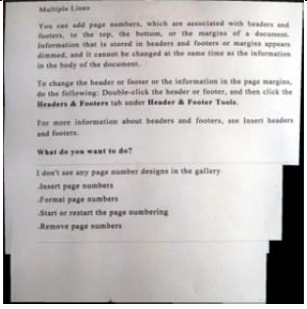
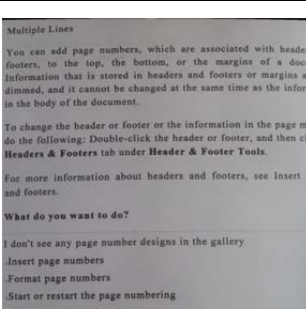
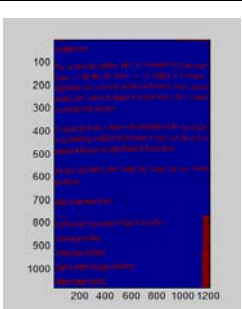


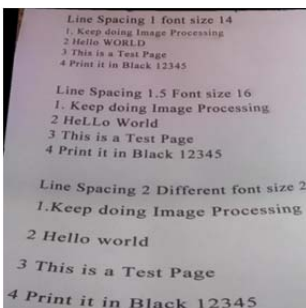
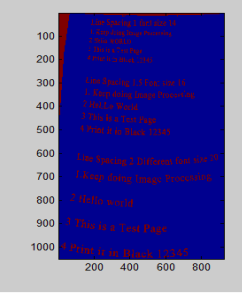

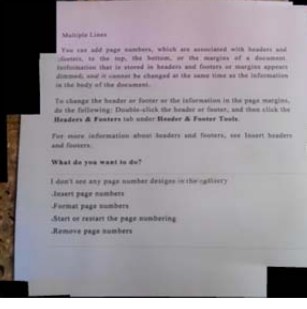
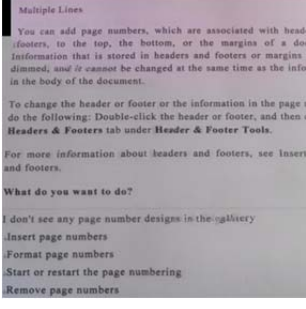
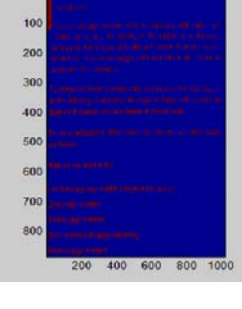

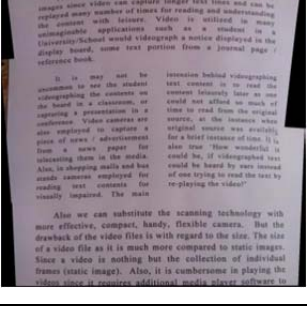
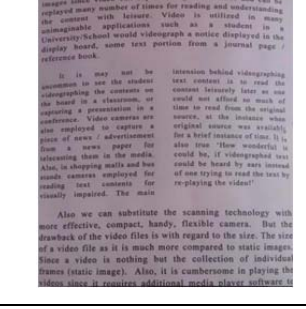
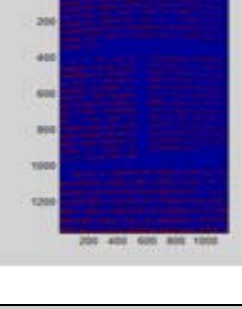
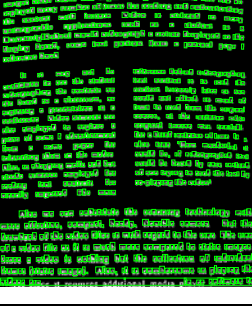
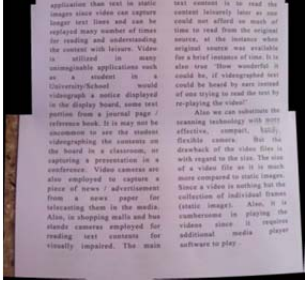

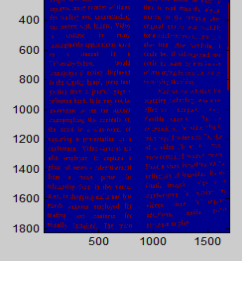
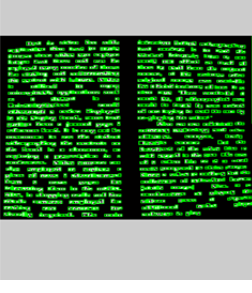
- False positives (FP) : False alarms are those regions in the image which are actually not characters of a text, but have been detected and recognized by the algorithm as text regions.
- False negatives (FN) : are those regions in the image which are actually text characters, but have not been detected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

TABLE I
PERFORMANCE MEASURES OF HYBRID METHOD FOR MOSAICED IMAGE

MOSAICED IMAGES	PRECISION RATE (%)	RECALL (%)	F-SCORE (%)	ACCURACY (%)
Mos1	100	100	100	100
Mos2	100	100	100	100
Mos3	100	100	100	100
Mos4	93.33	100	96.55	93.33
Mos5	93.33	100	96.55	93.33
Mos6	95.24	100	97.56	95.24
Mos7	100	96.15	98.04	96.15
Mos8	100	100	100	100
Mos9	100	100	100	100
Mos10	71.43	100	83.33	71.43

The accuracy of the proposed approach based on the accuracy value from the table is around 97%. We have analysed the results for the mosaiced image resulted from the method proposed in [2]. The complexities like how mosaic image is created are discussed in detail in [2] and [29]. The 10 mosaiced images we have selected from [2] and [29] to extract text using hybrid approach had different types of texts properties like multiple lines, multiples columns and different camera movement. The result of the proposed approach is given in Table 2.

TABLE 2 RESULTS OF HYBRID METHOD FOR MOSAICED IMAGE

Sampl e ID	Mosaiced Image	Cropped Image	K-means Clustered image	Text extracted Image
Mos1				
Mos2				
Mos3				
Mos4				
Mos5				

The proposed approach was executed using Matlab 2014(a) on Intel Core I5 processor, 4.00 GB RAM with 1.60 processor speed. A system clock was set on at the beginning of the m-file for both the approaches and execution time was recorded. The graph below shows the execution time for the proposed approach.

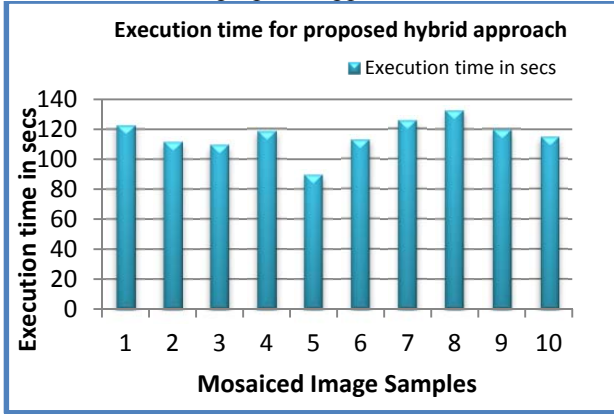


Fig 10 Execution time for the proposed hybrid approach

VII. CONCLUSIONS

In this paper we have proposed a hybrid approach based on edge clustering and connected components to extract the text from mosaiced text document images. The proposed algorithm is 97% efficient in recognizing the text. The recognized result is stored in the text file. The execution time, performance analysis of the proposed approach is also discussed in this paper.

REFERENCES

[1] K. Jung, K.I. Kim, A.K. Jain, "Text information extraction in images and video: a survey", *Pattern Recognition*, 37(2004) 977-997.

[2] Nagabhushan P, Vimuktha Evangeleen Jathanna. "Mosaicing of Text Contents From Adjacent Video Frames". *International Journal of Machine Intelligence*. Vol 3, Issue 4, 2011.

[3] P. Nagabhushan, S. Nirmala(2009) ,"Text Extraction In Complex Color Document Images For Enhanced Readability",*Intelligent Information Management*, pp: 120-133.

[4] Christin Wolf, Jean-Michel Jolion, and Françoise Chassaing. "Text localization, enhancement and binarization in multimedia documents". In *Int. Conf. on Pattern recognition*, pages 1037–1040, Quebec City, Canada., August 2002

[5] Egyul Kim, SeongHun Lee, JinHyung Kim. "Scene Text Extraction using Focus of Mobile Camera". In *10th IEEE Int. Conf. on Document Analysis and Recognition*, pages 166–170 -2009

[6] Qixiang Ye, Qingming Huang, Wen Gao, Debin Zhao. "Fast and robust text detection in images and video frames". *Int. Conf. on Image and vision Computing*, 23(2005) 565- 576

[7] Ataru Ohkura, Daisuke Deguchi, Tomokazu Takahashi, Ichiro Ide, Hiroshi Murase. "Low -resolution Character Recognition by Video-based Super Resolution". In *10th IEEE Int. Conf. on Document Analysis and Recognition*, pages 191-195 – 2009

[8] Min Cai, Jiqiang Song, Michael R.Lyu. "A New Approach for Video Text Detection". Department of Computer Science and Engineering, The Chinese University of Hong Kong.

[9] Lifang Gu. "Text Detection and Extraction in MPEG Video Sequences". *CBMI'01, Brescia, Italy*, September 19-21, 2001

[10] C.V.Jawahar, Balakrishna Chennupati, Balamanoah Paluri, Nataraj Jammalamadaka. "Video Retrieval Based on Textual Queries", *International Institute for Information Technology, Hyderabad*.

[11] Hichem Karray, Mohammed Salah, Adil.M.Alimi. "TEVI: Text Extraction for Video Indexing". *University of Sfax, Tunisia*.

[12] R. Lienhart, Indexing and retrieval of digital video sequences based on automatic text recognition, *Technical Report 6/96*, University of Mannheim, 1996.

[13] Fang Liu, Xiang Peng, Tianjiang Wang, Songfeng Lu. A Density Based Approach for Text Extraction in Images'. *IEEE Transaction*, 978-1-4244

[14] Trung Quy Phan, Palaiiahnakote Shivakumar, Chew Lim Tan "A Laplacian Method for Video Text Detection". In *10th IEEE Int. Conf. on Document Analysis and Recognition*, pages 66–70 -2009

[15] Shuicai Shi,Tao Cheng, Shibin Xiao,Xuequiang LV. Text Processing in Video Frames with Complex Background. *International Forum on Information Technology and Applications*, *IEEE pages 450-455-- 2009*

[16] D. Doerman, J. Liang, and H. Li, "Progress in camera-based document image analysis," *Proceeding International Conference of Document Analysis and Recognition*, pp. 606–616, 2003.

[17] Datong Chen and Jean-Marc Odobez. "Robust Video Text Segmentation and Recognition with Multiple Hypotheses", *IEEE, Proc of International Conference on Image Processing*, vol [2], *Pages -433-435 —2002*

[18] Jing Zhang, Dmitry Goldgof, R.Kasturi. "A New Edge Based Text Verification Approach for Video", *Department of Computer Science and Engineering, University of South Florida*.

[19] Datong Chen and Jean-Marc Odobez. "Sequential Monte Carlo Video Text Segmentation", *IEEE, Proc of International Conference on Image Processing*, vol [1] —2003

[20] Silvio Ferreira, Vincent Garin, Bernard Gosselin. "A Text Detection Technique applied in the Framework of a Mobile Camera-Based Application" *Pages – 133-139, TCTS Labs , Belgium*

[21] Chengjun Zhu, yuanxin Quyang, Lei Gao, Zhenyong Chen, Zhang Xiong."An Automatic Video Text Detection, Localization and Extraction Approach". *School of computer Science and Technology, Beihang University, SITIS - Pages 166-175-2006*

[22] S.Antani, U. Gargi, D. Crandall, T. Gandhi, and R. Kasturi." Robust Extraction of text in video". *IEEE Transaction*, *Pages-831-834— 2000*

[23] Seiichi Uchida, Hiromistu Miyazaki, Hiroaki Sakoe." Mosaicing-by-Recognition texts captured in multiple video frames" *Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan*

[24] Xian-Sheng Hua, Pei Yin, Hong-Jiang Zhang. "Efficient Video Text Recognition Using Mutiple Frame Integration" *Microsoft Research Asia, Depatment of Computer Science and Technology, Tsinghua University*

[25] Chitrakala Gopalan, D.Manjula. "Counterlet based Approach for text Identification and Extraction from Heterogeneous Textual Images".

[26] G. Rama Mohan Babu, P. Srimaiyee, A.Srikrishna(201), "Text Extraction From Heterogeneous Images Using Mathematical Morphology",*Journal Of Theoretical And Applied Information Technology*,Vol.16,No.1,pp 39-47.

[27] S K Gupta, K Sambasiva Rao and Vasudha Bhatnagar, K-means Clustering Algorithm for Categorical Attributes, *DataWarehousing and Knowledge Discovery Lecture Notes in Computer Science*, 1999

[28] Hassana Grema Kaganami, Zou BeiJi, M Sami Soliman, Optimal Color Image Enhancement Using Wavelet and K-means Clustering, *International Journal of Digital Content Technology and its Applications*. Volume 5, Number 1, January 2011

[29] Mosaicing of Text Contents from Consecutive Frames in Pedestal Shot Videos, P. Nagabhushan, Vimuktha Evangeleen Jathanna, *International Journal of Engineering Research & Technology(IJERT)*, Volume. 4 - Issue. 07 , July - 2015