International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 04, July 2016

Recognition and Diminution of abnormal textual content from video by automatique en peinture

[1]Punde Amruta Arun, [2]Prof. Amune Amruta

[1,2]Department of Computer Engineering, G. H. Raisoni Collage of Engineering and Management, Ahmednagar, University of Pune, India.

[1]amruta.a.punde@gmail.com, [2]amruta.amune@raisoni.net

Abstract **A video textual content diminution theme primarily involves two main stages; foremost, associate automatic video text recognition and second, efficient video restoration once the text removal. Automatique en peinture (Automatic In-painting in François) used for effective restoration of the video once text removal. In some case there we want of an automatic approach to induce eliminate unwanted matter contents from a video. The proposed system at first processes the given video victimization FFMPEG library and that we can regenerate the video by using Audio stream and Video Stream in specific. Inside the primary stage, we have to extract audio stream and video stream, which we will keep in temporary location. We will extract each frame of given video stream, apply method each frame/image to recognize text and detect the text from frames. We use OCR to extract text from all the frames. In second stage the in-painting advanced approach applied to reconstruct the missing data and finally we will create the video without textual content using processed frames. The experimental results demonstrate the effectiveness of both our proposed video text detection approach and the video in-painting technique, and consequently the entire automatic video text removal and restoration process.**

Keywords —FFMPEG Library; In-painting; OCR (Optical Character Recognition); text detection; text recognition; SWT (Stroke width transform); CC (connected components).

I. INTRODUCTION

The texts sometimes seem in video sequence as logos, subtitles, captions or banners. The informative embedded texts are often for the most part found within the news and alternative common TV broad castings. Though texts offer extra data, not all of them area unit necessary as they will block important portions of a video. Consider the case, for example, when indirect advertisement is not permitted but it is already included within a frame sequence in the form of a caption, textual advertisement running at the bottom of TV shows. Hence, there should be a way to erase the unwanted text from the video. This motivates the need of an automatic approach to remove undesired textual contents from a video. Automatique en peinture method of filling the missing regions of an image from the surrounding parts is known as Digital Image In-painting. The digital Image In-painting has numerous applications like restoration of broken recent printing and recent images, error recovery of images and videos. The two stage framework for automatic video text removal: 1) detect and remove embedded video texts and 2) fill-in their remaining regions by appropriate data. The operations basically are performed by splitting the video into various frames and considering each as a separate image [1], [3]. The separate stages of automatic text detection and removal/extraction after splitting video into various frames are as follows [3]:

1) Text Detection: Determining the presence of text location in each frame.
2) Text localization: Deciding the exact location of text in each frame.
3) Text tracking: Integrate location of text across the adjacent frames of video.
4) Text extraction: segmentation of text component from background.
5) Text removal: removing the text from the frame.

The textual content will be formed into plain text and removed and then fill the regions using appropriate way of in-painting method. Image in-painting is employed to recover the broken image and fill regions that are missing in original image. It restores the loss of information and reconstructs image looks as a natural image. All the previous methods had been not really perfect for automatic detection and removal of text from video. And so a new approach of automatic inpainting for video text detection and removal proposed as a great milestone in the area of automation by A. Mosleh [1].

In existing system technique, basically the operations are performed by splitting the video stream into frames and considering each frame as a separate image as an input for the stream of operations. The text are detected in each frame over here by generating edge map, and connected components (CC) are generated by using stroke width transform (SWT) and unsupervised clustering for detecting only text regions and rejecting other. The text objects are tracked using CAMSHIFT algorithm and filtering of noncaption text takes place. In the last step of inpainting after extraction of text from video, 3D bandlet transform is used as bandlet transform effectively represent image geometry, and exploits spatio temporal regularities to perform regularities to perform a regularization to address the restoration. In existing system video inpainting method, each caption is considered as volume and so smooth video without text is generated without further any more operation for maintaining visual consistency.

In this paper we propose a video text recognition approach which consist of multilingual text localization and detection technique and followed by an effective restoration stage. This paper focuses on new approach for recognition and restoration of multilingual text from video. The propose system concentrate on the improved technique for detection and removal of multilingual text from video by using enhance in-painting technique. The first stage means recognition of text will be done by using FFMPEG and OCR algorithm. Second stage diminution of text from video will be done by using Extended version of Criminisi's Algorithm for in-painting an image/frame. The propose system focuses on reducing the time complexity and give better results in respect of quality, also proposed system will be concentrate on multilingual text recognition.

The rest of this paper consisting, Section II reviews the related work. A brief overview of Exiting system is provided in Section III. Then brief overview of proposed system introduce in Section IV. The technique used for recognition and diminution of anomalous textual contents from video by automatic in-painting introduced in Section V. In Section VI experimental results are provided. Finally, Section VII concludes this paper.

II. LITERATURE SURVEY

The two stage framework for automatic video text removal:

1) detect and remove embedded video texts and
2) fill-in their remaining regions by appropriate data.

The operations basically are performed by splitting the video into various frames and considering each as a separate image [1]. The review associated with text detection techniques for single language is provided 1st during this section, then the text detection technique for bilingual text is provided and the brief overview of existing in-painting techniques for video/image is provided.

A. Text Detection Related Work

M. Favorskaya et. al. has projected Intelligent in-painting methodology for text removal from video, that is predicated on texture analysis in spatial domain moreover as in temporal domain of video. The feel synthesis methodology was the foremost fortunate technique for reconstruction of complete components of image. Intelligent in-painting methodology contains a risk to use call rules that management on restoration strategies choice [2].

Y. Chen et. al. projected associate degree automatic text extraction, removal and in-painting in complicated document pictures by decomposing the document image into distinct object planes like textual regions, nontextual objects, background textures, etc. By using knowledge based text extraction and identification the text with totally different characteristic from every plane is detected, then the effective adaptive in-painting neighborhood adjustment theme is applied right away once text removal [4].

D. Chen et. al. proposed general methodology for extracting and recognizing text from any grayscale worth pictures and videos. This methodology divided into 2 main steps as text lines detection then text in these lines recognition. Machine learning ways for text detection encounters difficulties thanks to character size and grey scale variation and serious computation value. During this methodology to beat on top of downside 2 steps localization theme is intended. Opening move quickly find the text lines and allows to normalize the text into distinctive size. Second step verification trained Multi Layer Perceptrons (MLP) or Support Vector Machine (SVM) is applied on background freelance options to get rid of the false alarms. This methodology improves the detection result at lower value as compared with same machine learning tools applied while not normalization. Text recognition

methodology includes the normal character segmentation step followed by OCR algorithmic program at intervals multiple hypothesis frame work. New grey Scale Consistency Constraint (GCC) algorithmic program designed to boost the segmentation results. GCC post processing step was ready to cut back character and word error rates and ready to take away burst like noise that disturbs OCR code. The SVM was a lot of higher and acceptable than the MLP to deal with text texture verification problem [5].

R. Jana et. al. proposed the text recognition of text from image by OCR system. The text recognition of written or written alphabetical text from image is going to be done. OCR converts the written or written characters image and converts into editable text document. The text image is split into separate regions by every line. By feature matching the extracted character and guide system acknowledges actual character used. The technique is based on calculating no. of corner points and utilizing the various properties like object area and convex areas of the image [12].

All the previous ways has not been very good for automatic text detection and removal from video and new approach of automatic inpainting for video text detection and removal planned as a good milestone within the space of automation [1], [3]. All the above methods have been used for text recognition and diminution for single language as English.

B. Multilingual Text Detection Related Work

M. Lyu et. al. projected comprehensive method for detection and localization and extraction of multilingual text. The fundamental 3 ways administrated for text detection as - Edge Detection, Local Thresholding, Hysteresis Edge Recovery. This method capable to handle multilingual text (Chinese and English) because it depends on language-independent characteristics. This method is also robust for the attributes such as - font size, font style, color, background complexity, orientation, etc. For the text detection: the sequential multiresolution dividing into levels from 1 to n and local thresholding of edge map and hysteresis text edge recovery is carried out [6].

L. Gomez et. al. proposed technique based on perception organization through which text emerges as perceptually significant group of atomic objects. Hence human can detect text even in language or script never seen before.

The proposed method built perceptual organization framework that exploits collaboration of proximity and similarity law to create text group hypothesis. This method is totally independent of the language and script in which text appears it can deal with any type of font and text size, no assumption about the orientation of text will be made [7].

The above two techniques deals with multilingual text detection and extraction method from natural scene [6], [7].

C. In-painiting Related Work

The process of filling the missing regions of an image or video from surrounding parts is known as Digital inpainting. In Digital in-painting variety of techniques were proposed by researcher. In-painting is used for restore image where the information loss and reconstruct the image looks as natural.

There are many different types of image inpainting algorithms are compared and made evaluation of these algorithms to provide a comprehensive visualization [8].

- In Texture Synthesis based image in-painting sample of texture is given and goal is produce more of that texture. Simplest solution is to be tile texture sample on rectangular grid of desired size. A rectangular texture is characterized by a primitive element that is regularly placed on a grid or a lattice. Non regular texture is no apparent repeating pattern or local structure but global statistical properties. This algorithm has difficulty with natural images as they are composed of structures in form of edges. This technique is suitable for small subset of in-painting issues but not suitable for a large objects [9].

- The Patrial Differential Equation (PDE) based algorithm is proposed by Marcelo Bertalmio et. al. This is iterative algorithm, it is to continue geometric and photometric information which arrives at border of occluded area into itself. This algorithm good for small missed regions, if the regions are large it take large time and results will not produced good [9].

- The Exemplar based image in-painting contents two basic steps , first priority assignment is done and secondly the selection of best matching patch from the known region, similarity measured by certain metrics and pastes into target patches in the missing region. This method is efficient approach for constructing big target region, means gives better results for the huge missing region in-painting. Exemplar based in-painting gives good results if the missing region consists of simple structure and texture, if there are not sufficient samples in image then it is impossible to reconstruct desired image. It will work for missing region contains only simple structure and texture [9], [10].

- Hybrid in-painting is a combination of both PDE based in-painting and Texture Synthesis based in-painting. This in-painting has two step approach as structure compl followed by texture synthesis. In structure compl segmentation is performed based on the insouciant geometry, colour, texture information on input. The partitioning boundaries are designed to complete segmentation. The next step synthesize texture and colour information in each segment. This technique handle large holes and preserve both texture and structure effectively, computing time is more for the large holes. The missing information in structure component is reconstructed using a structure in-painting algorithm, while texture component is required by an improved exemplar based synthesis [10].

Exemplar based method had major contribution in developing in in-painting field. Main idea in these method is based on copy and paste texture synthesis, the time complexity of these method are high [11].

The Criminisi's Algorithm most referred exemplar based in-painting technique. Criminisi's developed novel technique based on texture synthesis in which filling order is influenced by the linear structured image. In this algorithm the strength of structural and textual synthesis combined to fill missing region in image [11].

V. Alilou et. al. proposed A New Fast Exemplar-Based In-painting algorithm for reconstructing missing parts of an image based on exemplar-matching techniques in which both performance and speed of the algorithm increased. The improvement done in the existing algorithm without modifying algorithm. The technique presented the fast and simple algorithm for in-painting images with the large missing region.

A New Fast Exemplar-Based In-painting algorithm performs well quality of result and faster than previous methods computation time [11].

III. Overview Of Existing System

A. Mosleh et. al. has proposed system presents a two stage framework for automatic video text removal to detect and remove embedded video texts and fill-in their remaining regions by appropriate data (refer to Fig. 1). In the video text detection stage, text locations in each frame are found via novel edge detector which benefits from the geometric features revealed by the bandlet transform and via an unsupervised clustering performed on the connected components (CCs) produced by the stroke width transform (SWT). The text object tracked with Continuously Adaptive Mean Shift (CAMSHIFT) algorithm and filtering of non-caption text takes place. The detected video text regions are removed, and then the video is restored by an in-painting scheme as bandlet based 3D volume regularization algorithm. The proposed video in-painting approach applies spatio-temporal geometric flows extracted by bandlets to reconstruct the missing data [1].

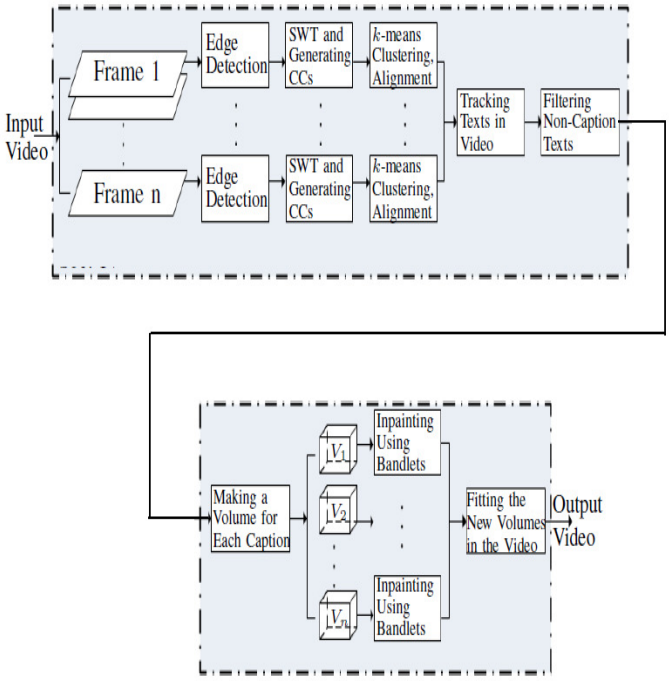

Fig. 1. Main stages of video text detection and removal by A. Mosleh [1]

IV. Overview Of Proposed System

For erase the unwanted text from the video the need of an automatic approach to remove undesired texts from a video. Roughly speaking an automatic video text removal scheme involves two main stages: i) an automatic video text detection, and ii) an effective video completion/restoration after the text removal. There are many existing video text completion techniques rarely cover both of these aspects in a single platform.

In the proposed system framework (refer to Fig. 2). consist of two stages, the first stage is for recognition and diminution of text present in video, the second stage is for the Automatique en peinture after the text removal from video. The proposed system the first stage we will process the video using FFMPEG library and we will extract text embedded as third stream in video. And we will regenerate the video using Audio stream and Video Stream only. This all process can be done using FFMPEG only.

In second stage, we will extract each and every frame of video and we will keep all the frames in temporary location and we will also extract audio stream and that also we will kee temporary location. Now we will process each and e frame/image to find out text or remove the text from images. We use OCR to extract text from all the frames. The in-painting advanced approach applied to reconstruct the missing data and finally we will create the original video using processed frames/images and Audio stream. We will delete all temporary location data.

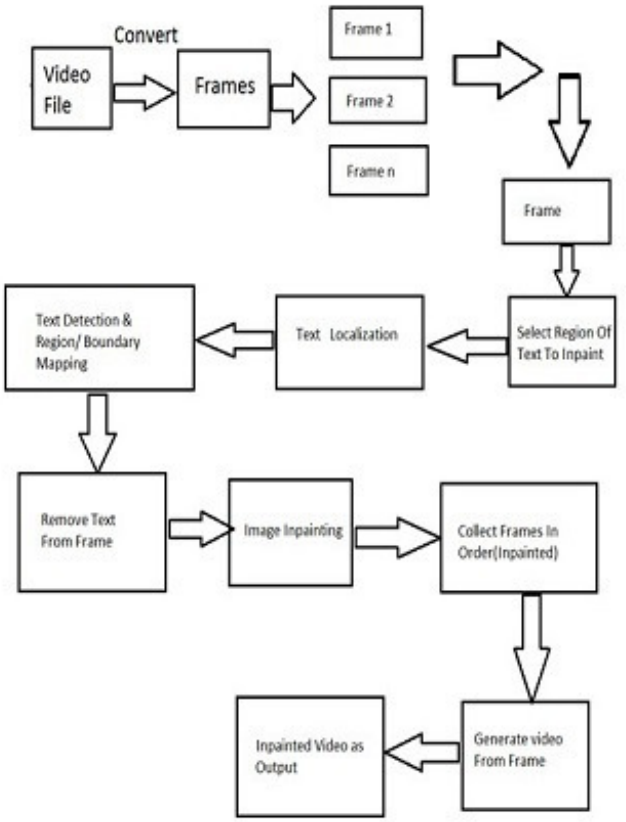

Fig. 2. Proposed System flow of video text detection and removal

V. PROPOSED SYSTEM ALGORITHM

- ***Algorithm of Proposed System***

We apply following steps for recognition and diminution of anomalous textual contents from video by Automatique en peinture:

1. *Select video File*
2. *Convert video file to Frames – n Frames*
3. *Sort out frames which has text region to in paint*
4. *Text localization – Find text areas using OCR*
5. *Text Detection and region / boundary mapping*
6. *Recognize and Remove text from frames*
7. *Apply the in-painting algorithm i.e. Extended version of Criminisi's Algorithm of Enhanced Exemplar Based Approach on frame*
8. *Collect In-painted frames in order*
9. *Generate the video of frames and audio stream, video without textual contents.*

- ***Flow of OCR Algorithm***

Text recognition, even from the detected text lines, remains a challenging problem due to the variety of fonts, colors, the presence of complex backgrounds and the short length of the text strings. Text recognition methodology includes the normal character segmentation step followed by OCR algorithmic program at intervals multiple hypothesis frame work. While text recognizing we will save the extracted text in separate text file. So we can identify the extracted text from the video later and also we will not loss the text which is present in video.

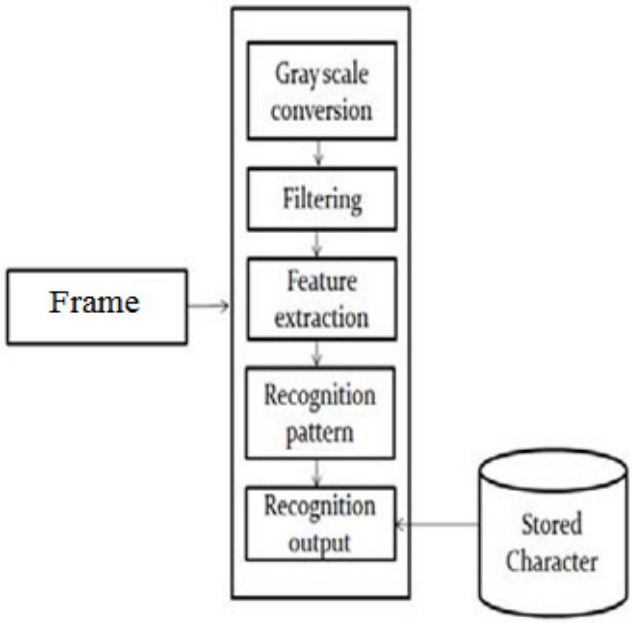

Fig. 3. Flow of OCR Algorithm

- ***Ehanced Criminisi's Algorithm***

Generally an Enhanced Criminisi's algorithm of enhanced exemplar based inpainting approach involves the following steps:

i. *Initialize the target region.*

This is generally performed separately from the inpainting process and requires the use of an additional image processing tool. This is performed by marking the target region in some special colour. Without any loss of generality, let us consider that the colour that the target region will be marked in is green (i.e. R = 0, G = 255, B = 0).

ii. Find the boundary of the target region.

iii. *Select a patch from the region to be in-painted.*

The patch size should be a bit larger than the largest distinguishable texture element in the image. We have used a default patch size of 9 x 9 which can be changed

with the knowledge of the largest texture element in the image. We denote the patch by ψp.

iv. *Find a patch from the image which best matches the selected patch, ψp.*

This matching can be done using a suitable error metric. We use the Mean Squared Error (please refer eq. 1) to find the best matching patch.

$$MSE = \sum \frac{(f_{x,y} - g_{x,y})^2}{N}. \quad (1)$$

where fx,y represents the element of the patch ψp and gx,y represents the elements of the patch for which MSE is to be calculated. N is the total number of elements in the patch.

v. *Update the image information* according to the patch found in the previous step.

VI. Experimental Results

In Existing System the set of videos contains sequences captured from TV, movies and video games. The resolution of each video sequence is 320 × 240. In the implementation of the text detector and the video inpainting scheme, the settings are used[1].

In Proposed System several video sequences are used to evaluate the proposed video text removal method. The set of videos contains sequences captured from TV, cartoon shows, movies and Camera Captured video. The resolution of each video sequence is not restricted; also any type of video can be used for the process. Following are the results we get after the video is processed.

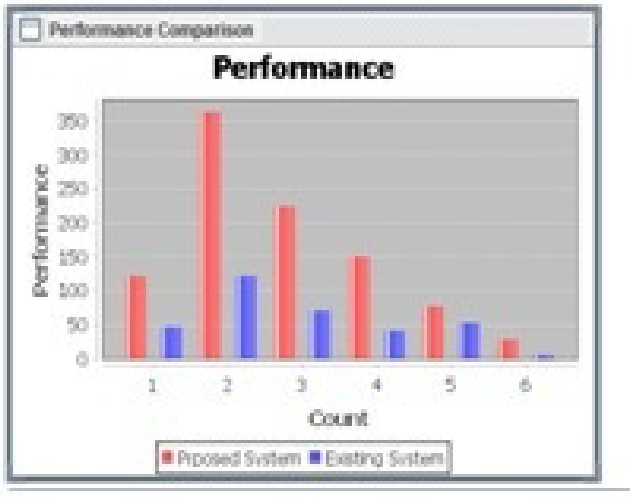

Fig. 4. Performance comparison of Existing and Proposed System.

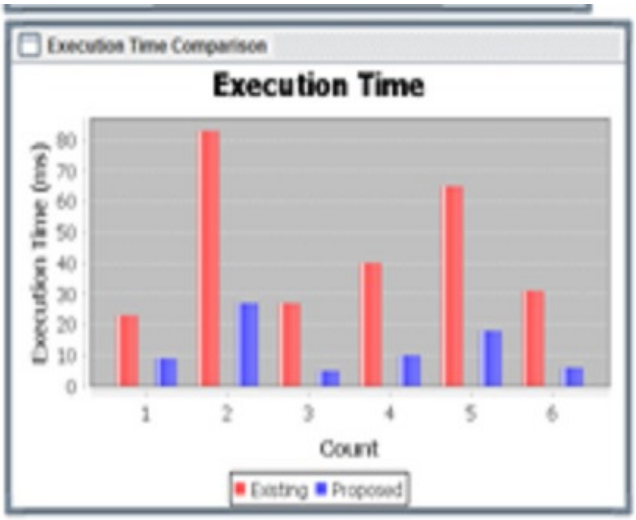

Fig. 5. Execution Time comparison of Existing and Proposed System.

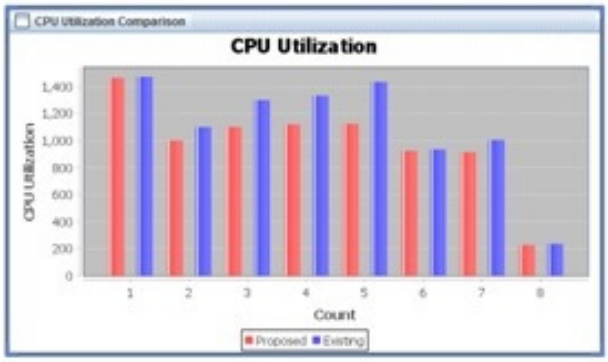

Fig. 6. CPU Utilization comparison of Existing and Proposed System

VII. Conclusion

The existing approaches for Textual content detection and extraction as well as many techniques for in-painting. There were many techniques implemented to detection and removal the textual contents from video by using image in-painting. So, as we have mentioned different techniques to detection, extraction and removal of textual contents from images or video and in-painting of image / video. The Identifier/Locator text separation schemes were developed for one or two languages, but still it takes more time to locate the text, in-painting the missing regions from image/frame.

Existing system for video text detection and removal system having some limits, so the propose a system going under in reduce the time complexity and give better results in respect of quality, also proposed system will be multilingual text detection.

We presented system having two stages, the first stage we processed the video using FFMPEG library and OCR

algorithm. We use OCR to recognize and extract text from all the frames. The second stage used in-painting advanced approach applied to reconstruct the missing data and finally we will create the original video using processed frames/images and Audio stream. The Enhance Criminisi's Algorithm is used for in-painting. The Proposed system will recognize as well as remove the multilingual text (as English and Chinese language text) from the video and the automatic in-painting will be done with the advanced in-painting technique.

ACKNOWLEDGMENT

I would like to take this opportunity to express my heartfelt thanks to my guide Prof. Amruta Amune for her esteemed guidance and encouragement, especially through difficult times. Her suggestions broaden my vision and guided me to succeed in this work. I am also very grateful for her guidance and comments while designing part of my research paper and learnt many things under her leadership.

REFERENCES

[1] A. Mosleh, N. Boouguila, and A. B. Hamza, "An Automatic Inpainting Scheme for Video Text Detection and Removal," in IEEE Transactions on Image processing, vol. 22, no. 11, pp.4460–4472, Nov. 2013.

[2] M. Favorskaya, A. Zotin, and M. Damov, "Intelligent inpainting system for texture reconstruction in videos with text removal," in *Proc. ICUMT*, Oct. 2010, pp. 867–874.

[3] Priyanka Deelip Wagh, and D R Patil, "Text Detection, Extraction and Removal: A Survey," *IJCA Proceedings on National Conference on Emerging Trends in Computer Technology* NCETCT(1):19-23, Dec. 2014.

[4] Yen-Lin Chen, "Automatic Text Extraction, Removal and Inpaiting of Complex Document Images," in International Journal of Innovative Computing, Information and Control, Vol. 8, No.1(A), pp. 303-327, Jan. 2012.

[5] D. Chen, J.-M. Odobez, and H. Bourlard, "Text detection and recognition in images and video frames," *Pattern Recognit.*, vol. 37, no. 3, pp. 595–608, 2004.

[6] M. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 243–255, Feb. 2005.

[7] L Gomez, D Karatzas, " Multi-script Text Extraction from Natural Scenes ", International Conference on Document Analysis and Recognition (ICDAR), 2013.

[8] S. Ravi, P. Pasupathi, S. Muthuumar, N. Krishnan, "Image in-painting techniques - A survey and analysis," Innovations in Information Technology (IIT), 2013 9th International Conference at Abu Dhabi, pp. 36-41, 17-19 Mar. 2013.

[9] Rajul Suthar, Mr. Krunal R. Patel, "A Survey on Various Image Inpainting Techniques to Restore Image", Int. Journal of Engineering Research and Applications, Volume 4, Issue 2(Version 1), pp. 85-88, ISSN: 2248-9622, Feb. 2014.

[10] Khushbu A.Panchal 1, Dr.Mehfuza Holia, "A Survey: Different Technique of Video Inpainting", International Journal of Innovative Research in Computer,Vol. 2 Issue 2,Feb. 2014.

[11] Vahid K Alilou, F Yaghmaee, "Introducing a new fast exemplar-based Inpainting algorithm", Electrical Engineering (ICEE 2014), 22nd Iranian Conference on, pp. 874-878, May. 2014.

[12] Ranjan Jana, Amrita Roy Chowdhury, Mazharul Islam, "Optical Character Recognition from Text Image," International Journal of Computer Applications Technology and Research, Volume 3– Issue 4, pp. 239 - 243, ISSN: 2319–8656, 2014.

