A Survey Paper on Recognition and Diminution of Anomalous Textual Content from Video by Automatique en peinture

Amruta Punde[1], Amruta Amune[2]

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

[2]Professor, 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— Nowadays, A video textual content diminution scheme mainly involves two main stages; first off, associate automatic video text recognition and second, an effective video restoration after the text removal. Automatique en peinture (Automatic In-painting in François) used for effective restoration of the video after text removal. In some case there we need of an automatic approach to get rid of unwanted textual contents from a video. But, there are many techniques available for detecting and removing textual contents as well as for restoration of video. Some of the techniques and papers we are reviewing here in this paper

Keywords— Text detection, edge detection, text extraction, text removal, Optical Character Recognition (OCR), image in-painting, image restore, CAMSHIFT.

I. INTRODUCTION

The texts usually appear in video sequence as logos, subtitles, captions or banners. The informative embedded texts can be largely found in the news and other popular television broad castings. Although texts provide additional information, not all of them are necessary as they may 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 process 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 various applications such as restoration of damaged old printing and old photographs, 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 used to recover the damaged image and fill regions which are missing in original image. It restores the loss of information and reconstructs image looks as a natural image.

In this paper, section II gives brief information about related work already performed for text detection & extraction from video and image. In section III provides analysis of different techniques used for in-painting image and video.

II. PAGE TEXT RECOGNITION AND DIMINUTION : RELATED WORK

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].
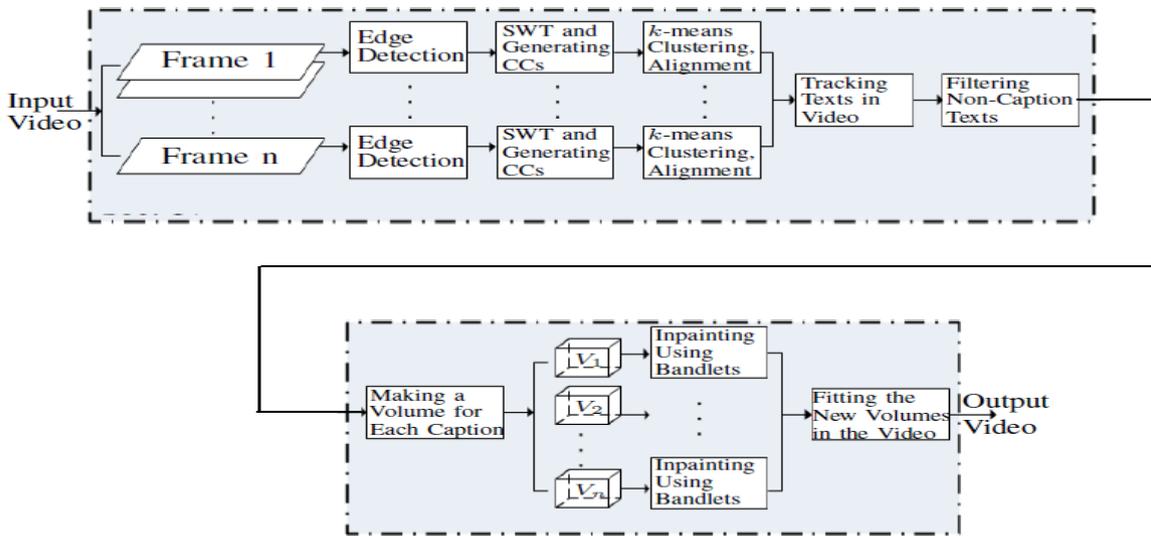

Fig. 1 Main stages of video text detection and removal by A. Mosleh [1]

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. Program realization has intuitive measurement interface, includes all perform that are necessary to reconstruct frames while not essential text. The accuracy by texture synthesis method of reconstruction is 93-95% [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. Algorithmic in-painting and use of encompassing background pictorial information generates the nontext background image [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].

W. Kim et. al. projected the technique for overlay text detection and extraction from video. In this technique 1st transition map is generated then candidate regions area unit extracted and overlay text is detected on the idea of incidence of overlay text in every candidate region and at the tip localization of overlay text region is performed with the assistance of projected overlay text pixels in transition map and immediate step of extraction is administrated. The primary step victimization combination of the modification of intensity and changed saturation a transition map is generated okay with the assistance of this technique. In text detection step the ratio of overlay text region, density of transition pixels and texture based mostly approach area unit basic elements to be operated for text region detection. By victimization no. of various native binary patterns and density of transition pixels in every candidate region the Probability Of Overlay Text (POT) is verify. The overlay text region change is employed so as to require advantage of overlay text continuity between 2 consecutive frames for following frame detection, then color polarity is checked and inverse of intensity is performed to own bright text compared to its encompassing pixels in each module. Last step is adaptive thresholding and changed dam purpose labeling area unit performed and victimization inward filling of background of text region overlay text is extracted [7].

A. Mosleh et. al. proposed text detection methodology supported feature vector generated from connected parts made by Stroke dimension rework (SWT) from image. During this an unsupervised bunch theme for text detection are often used as a connected element primarily based text detection ways. The parts feature vector area unit to separate the text parts from nontextual parts. Supported the alignments of the found text parts a location of the text is decided from image. The accuracy is depends on however exactly Connected Components (CCs) area unit generated in SWT domain meaning edge detection theme. Thus during this the novel bandlet-based edge detection approach used for locating edge locations of texts embedded in varied varieties of pictures and consequently generating SWT values of text from image. During this effectiveness of bandlet in representing resident geometry properties is employed [8].

K. Jung et. al. had done survey of the Text info Extraction (TIE) involves text detection, localization, tracking, extraction, improvement and recognition of text from a given image and TIE problem divided into these 5 sub problems. The text detection referred for determination of presence of text in given frame, unremarkably text detection used for sequence of pictures referred to as frame. Text Localization may be a method of decisive the placement of text in image and generating boxes around the text. Text tracking is performed to cut back the time interval for localization and maintain the integrity of position across adjacent frames. The text extraction is that the stage wherever the text elements area unit segmental from the background. The enhancement of extracted text elements is needed as a result of low resolution and noise. The extracted text can created into plain text by using OCR technology [9].

K. Kim et. al. proposed texture-based technique is developed for detection text in images. First Support Vector Machine (SVM) is employed to investigate matter properties of text; no external text feature extraction module is employed. The SVM directly provides the intensities of raw pixels that form up texture pattern. SVM additionally works well for prime dimensional areas. Next can text regions are known by Continuously

Adaptive Mean Shift (CAMSHIFT) algorithmic rule as results of texture analysis. The SVM and CAMSHIFT are powerful and effective for text detection from any variety of pictures additionally the quick detection is going to be worn out this method. The system also finds the very little text as well as the text from the low contrast images. [10].

J. Gllavata et. al. projected powerful technique for text localization approach for image which may mechanically notice horizontally aligned text with totally different sizes, fonts and colors. The primary wave rework is applied to the image for the classification. The high frequency wave constant distributes the characterized text and nontext space. For distribute the areas provided to projection profile analysis to refine their localization. Finally binary segmented image provided to associate degree OCR engine. This methodology for detection and localization in pictures with advanced backgrounds with differentiating the matter and non-textual region. The additional analysis are going to be text detection approach for videos, use of MPEG motion data for text detection and realize method for whole text detection freely out there OCR system are going to be used from input image to the ASCII text in the end [11].

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, the individual characters by house. When character extraction, text image are calculated by exploitation totally different texture properties as options of various regions, quantitative relation of character space and bell-shaped space. The options like grapheme, upper case, letter, digit, symbols are hold on as guide. By feature matching the extracted character and guide system acknowledges actual character used. The system thought of 3 fonts Arial, Berlin Sans and Lucida fax have thought of as coaching information set, supported this coaching information set totally different no. of characters are tested for OCR. 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].

K. Safronov et. al. proposed new methodology for OCR victimization hierarchic optimization rule. The prevailing ways and algorithms for OCR aren't appropriate for victimization in business. These aren't stable to defects and distortions of recognized characters. The new methodology supported pattern recognition and hierarchic optimization rule. The most ways of character recognition is divided into following teams –
   1. Pattern Systems
   2. Structural Systems
   3. Feature Systems
   4. Nueronal network Systems
The new hierarchic character recognition rule is conferred that uses optimization ways on the premise of patterns with completely different resolutions. The rule isn't smart to the image defects, and produces high recognition accuracy and high rate that employed in industrial systems [13].

All the previous ways has not been very good for automatic text detection and removal from video and new approach of automatic in-painting 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 one language as English.

M. Lyu et. al. proposed comprehensive method for multilingual text detection, localization and extraction. The basic three methods carried out for text detection as –
   1. Edge Detection
   2. Local Thresholding
   3. Hysteresis Edge Recovery
So many methods have been proposed for basically English language text detection, localization, extraction or removal from image and video till date. But if the text of any other language appears in image or video then due to certain reason system fails many times in detecting and removing/extracting those texts from video or image [6]. The comprehensive method for detection and localization and extraction of multilingual text. This method focuses on Chinese and English as well as capable to handle multilingual text because it depends on language-independent characteristics. So proposed method can be applied French, Spanish, Japanese without much adjustment. This method is also robust for the attributes such as – font size, font style, color, background complexity, orientation, etc. in this video sampled at two frames per second and each frame converted into 256 level grayscale image and each frame /image considered as original image for further operations as to detect text, localize and extract text. 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. The text localization coarse-to-fine localization method keeps multiple passes of horizontal and vertical projection to deal with multilingual

text as well as complex layouts. For text extraction adaptive thresholding of gray scale text images and the dam point based inward filling used [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. However the text extraction is usually based on classification of individual region using a priori knowledge for given script or language. The proposed method built perceptual organization framework that exploits collaboration of proximity and similarity law to create text group hypothesis. System assumed that the fact a region can be related to other is central to classifying the region as a text part. The text extraction process is divided into three main steps (refer to Fig. 2) as –
1. Region Decomposition
2. Perceptual organization based analysis
3. Text line Formation

PERCEPTUAL ORGANIZATION BASED ANALYSIS

REGION DECOMPOSITION → Group Hypothesis → Meaningfulness Testing → Evidence Accumulation → TEXT LINE FORMATION & VALIDATION
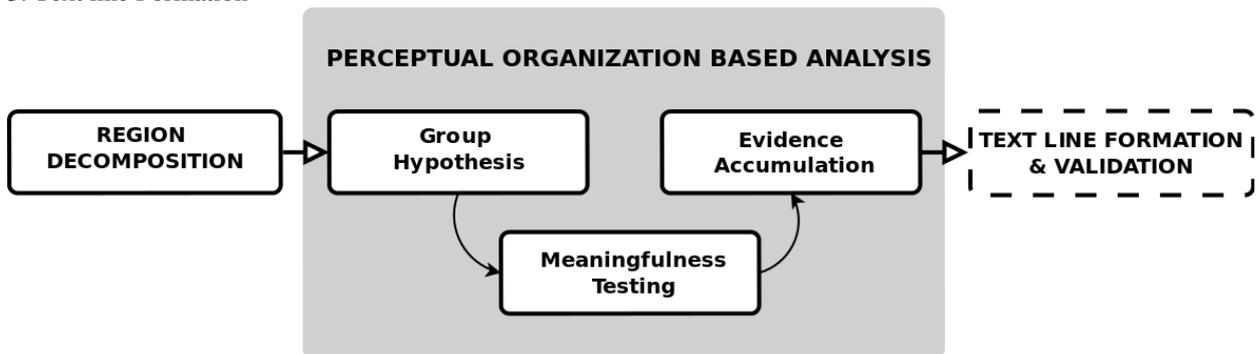
Fig. 2 Text extraction algorithm pipeline [18]

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 [18]. The above two techniques deals with multilingual text detection and extraction method from natural scene [6] [18], totally language independent.

III. IN-PAINTING TECHNIQUES : RELATED WORK

The process of filling the missing regions of an image or video from surrounding parts is known as Digital in-painting. 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. Image in-painting is technique which is used to recover the damaged image and fill regions which are missing in original image possible way. By in-painting the damaged region is to be change in an image.

The video completion is to reconstruct the missing half or holes that ar created by removal or broken of any designated object from video. The main issue in video in-painting is to keep spatial temporal information. Most of researchers try to maintain either spatial consistency or temporal continuity between the frames [16].

The difference between image and video in-painting approaches is in video in-painting the region can be much larger with main focus being the filling in of two dimensional repeating patterns that have associated textures. In image in-painting focus on the removal and completion of unwanted or damaged region in single image using surrounding information [16].

There are many different types of image in-painting algorithms are compared and made evaluation of these algorithms to provide a comprehensive visualization. The algorithms are analysed in both theoretical and practical ways and suitability of these in-painting algorithms over different kinds of applications over different areas. The success of in-painting algorithm lies how well information, colour, shape and structure are propagated into the unknown area. Each method have good renovation ability but problem is with reconstruction of original image, problem is in edges, the sharp edges will cause more errors and take long time for reconstruction [14].
Different approaches for digital image in-painting can be categorised as follows [14] -
1. Texture Synthesis based in-painting
2. Semi automatic and fast Digital in-painting
3. Partial Differential Equation (PDE) based in-painting
4. Exemplar and search based in-painting
5. Hybrid in-painting

270

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 [15].

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 [15].

The Exemplar based image in-painting contents two basic steps , first priority assignment is done and secondly the selection of best matching patch. The Exemplar based approach samples best matching patches from the known region, similarity measured by certain metrics and pastes into target patches in the missing region. Exemplar based in-painting iteratively synthesize the unknown region by the most similar patch in the source region. The method fills structures in the missing region by using spatial information of neighbouring region. The Exemplar based In-painting algorithm includes following four steps -
      1) Initializing the Target Region- Initial missing area are extracted and represented with appropriate data structures.
      2) Computing filling properties
      3) Searching example and compositing
      4) Updating image information
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 [15] [16].

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 completion followed by texture synthesis. In structure completion 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 [16].

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 [17].

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 [17].

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 search process for the best matching patch which minimizes Sum of Squared Differences (SSD) is exceedingly time consuming, the proposed technique improves search strategy without modifying the algorithm. The proposed technique improves search strategy relies on the fact the distance between two patches cannot be smaller than zero, continuation of search process will be superfluous. The improvement done is about color and space. The technique results with improved without modifying the original 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 [17].

IV. CONCLUSIONS

In this way we have studied 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.

Current video text detection and removal system having some limits, so the researchers are currently working on this area and would propose a system going under in reduce the time complexity and give better results in respect of quality, also proposed system will be language independent. The future work will be concentrate on the improved technique for detection and removal of text from video by using enhance in-painting technique.

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]   W. Kim and C. Kim, "A new approach for overlay text detection and extraction from complex video scene," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 401–411, Feb. 2009.

[8]   A. Mosleh, N. Bouguila, and A. Ben Hamza, "Image text detection using a bandlet-based edge detector and stroke width transform," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2012, pp. 63.1–63.12.

[9]   K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," *Pattern Recognit.*, vol. 37, no. 5, pp. 977–997, 2004.

[10] K. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[11] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," in Proc. of International Conference on Pattern Recognition (ICPR), vol. 1, Aug. 2004, pp. 425–428.

[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.

[13] Kirill Safronov, Dr.Ing. Igor Tchouchenkov, Prof. Dr.Ing. Heinz Wörn," Optical Character Recognition Using Optimisation Algorithms ", Institute for Process Control and Robotics (IPR), University of Karlsruhe Karlsruhe, Germany, Workshop on Computer Science and Information Technologies CSIT'2007, Ufa, Russia, 2007.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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