Using Harris Corner Detection and Background Modeling for InPainting Occluded Objects in a Video

IsraaHadi Ali*[1], Mustafa Ehsan M[2]

Information Technology College, Babylon University, Babylon, Iraq
Corresponding Author Email: israa_hadi1968@yahoo.com,

Abstract

In this papera proposed ofan efficient method for video inpainting which can be used to remove a single unwanted object that obscured another wanted object in a period of time in a video scene that captured for two occluded objects only. The proposed system based on statistical modified background modeling and interest point detection for the wanted object. Occlusion is the most important challenge in video inpainting and the proposed method can handle this challenge with a good result and less computational cost.The system consists of the following main steps. Firstly, input the original video then estimate the background of the video. After that detect the foreground objects by subtraction the background. Then applying segment and labeling stage each foreground object. Followed by tracking the selected two objects. Thenanalysis the occluded frames with computing interest point for the wanted object in the non-occluded frames basing on Harris corner detector. The final stage filling the unwanted object pixels by two choices either background model pixels or wanted object pixels.The proposed system focus on interruption cases that may be happened by the unwanted object and blocking another wanted and important object. By applying this methodology, the process of inpainting the unwanted object with the keeping of consistency of video.

Keywords: Video inpainting, Background model, Remove object, Video editing, Video tracking, occlusion.

1. Introduction

Inpainting term indicates the process of refill the holes that produced from the operation of removing a specific region in a video or image [8]. It's considered one of the important topics in computer vision processing which may be used to enhance videos or images by removing distortion or unwanted objects [18,19]. Inpainting encompassing two fields, image inpainting, and video inpainting, where image inpainting work in just spatially information (information in the same image around the inpainted region) as a basis for reconstruction pixels that used in inpainting process [1, 3]. In the other hand, video inpainting can be depended on Temporary information in addition to spatial information (information that calculated from all frames in the video around the inpainted region) that's make it closer to the truth environment [4]. The spatio-temporal dependency achieved by utilizing the similarity between video frames [6,8]. There are wide fields that utilize Inpainting in their application such as image enhancement, object removing, and films industry [3]. Many defaults that may face the inpainting process such as illumination change, shadow recreation, moving camera, etc. occlusion is one of the important challenges because of the difficulty of prediction of obscured objects that want to keep up in the scene. The method will implement in this paper on the single unwanted object and assumed the video is captured by a static camera and for a single scene for two objects as wanted an unwanted objects. And there is an occlusion between an unwanted moving object and another wanted object. The proposed system working on remove the unwanted object that covers the wanted object in a period in the video.

And filling the holes that produced from deleting operation either by background pixels depending on the background image that obtained from background modeling or object pixels. Background modeling is one of the important methodologies that used to obtain foreground objects. There are many classes of background modeling [20]. The method that implemented in this paper is the accumulation histogram method based on mean and variance criteria [22]. On other hand the detection of best object pixels that used for inpainting process will depend on Harris corner detector [21]. In addition to this suction, this paper contains the flowing sections. Section 2,explain the theoretical background of the implemented methods. Section 3, is a Reviewing of some related work. The proposed system of this paper in Section 4. Section 5,is an implementation of the proposed system in addition to discussion and evaluation to the final result. and conclusions of the paper in Section 6.

2. Theoretical Background

2.1 Mean Shift Tracker

Mean Shift Method one important methods in video tracking it uses the histogram feature to track the moving target object. The tracking in this method depends on comparing between the target models histograms based on the similarity.

Mean shift method uses either a rectangle or an oval area to represents the target [17]. This method also contains two models to compare, the target model and the candidate model. By depending on the histogram principle this method calculates the similarity

between the two models. in order to find the best similarity in each frame the Bhattacharya coefficient is calculated between the target model and the candidate model [16].

Then the method finds the probability of the color distribution in the histogram for the target in order to construct the target model by using equation (1). where q represents the probability of the color u, u=1,..m in target model where color distribution by discrete l of bin color histogram.

$$q_u = C \sum_{i=1}^n k(\|x_i\|^2)\delta(\text{b}(x_i) - u) \qquad (1)$$

The operator xi, i=1, ..n refer to the pixel locations of the target model. n is the number of pixels in an image. (xi) denote the color bin of the color at xi. C is the normalization constant.
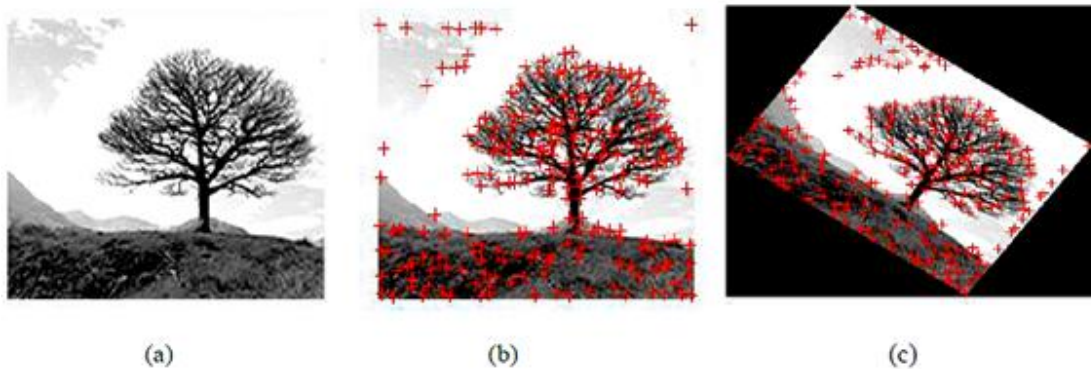
2.2 Harris Corner Detector

Harris corner detector is one of the famous algorithms that use for discovering interest points that which represent the angle head figure (1) illustrates the result of implementing this algorithm of an image.

Applying this algorithm depend on computing autocorrelation matrix technique which computed by using equation (1). This matrix is characterized as 2*2 symmetric [14].

$$M(x.y) = \sum_{u.v} w(u.v) * \begin{pmatrix} I_x^2(x.y)I_xI_y(x.y) \\ I_xI_y(x.y)I_y^2(x.y) \end{pmatrix} \qquad (1)$$

Fig. 1: Result Harris Corners (a) Original Image (b) Corners Detection (c) Rotate Image [14]

WhereIx and Iy are the local part derivatives in the x and y directions. the Ix,Iy are calculated by equations (2 , 3). W (u, v) represent the window weight over the area (u, v)..

$$I_x = G_\sigma^X * I \qquad (2)$$

$$I_y = G_\sigma^y * I \qquad (3)$$

Where $G\sigma x$, $G\sigma y$ are Gaussian, I is I(x, y) which consider the gray pixel density area (u, v) to be (x, y). The eigen value of the matrix M is used to find the interest points by computing this matrix for each pixel in the image. If both eigenvalues are large that means is interest points (corner) [9], see the Figure (2). To classify the image point by an autocorrelation matrix the corniness measure C (x, y) is used for each pixel by using equation (4):

$$C(x.y) = \det(M) - K(trace(M))^2 \qquad (4)$$

Where det $(M)= \lambda1*\lambda 2$ $and trace(M)=\lambda1+\lambda 2$
The K is a constant parameter and its range from (0.04) to (0.5). λ1 and λ2 are the eigenvalues of autocorrelation. Harris is a simple and steady way of detecting corners [14].
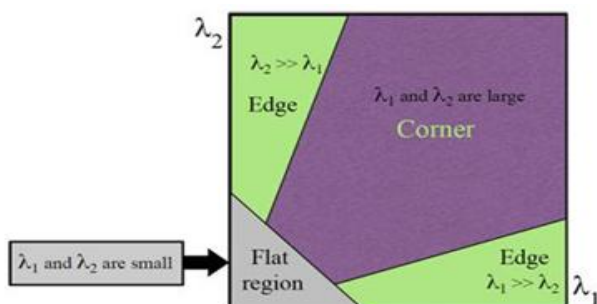

Fig. 2: Classification Eigen value of the Autocorrelation Matrix [15]

3. Literature Works

Kader A. Patwardhan et al. (2006) suggested a process for video inpainting his proposal dealt with videos that captured by a fixed camera and contain some moving object. His methodology performs inpainting for a target object and uses the best matches background pixels to fill its pixels by using a priority for each candidate filling pixels [2]. SamehZarif et al. (2013) presented a paper to remove a static object from a video scene by use local similarity in video completion algorithm. This algorithm uses the adjacent pixels to the target object to fill the pixels of inpainted object[8].Ali Mosleh et al. (2013). Worked on remove text that founded in the video content by two-stages "automated framework work proposed to detect and remove the text using unsupervised clustering and stroke width transform" [4]. Pierre-Luc et al. (2014) use local binary similarity patterns to enhance background subtraction he derives his proposed algorithm from ViBe method and low-cost to produce improved background subtraction method depending on spatio-temporal descriptor based on binary similarity [7]. MouniraEbdelli et al. 2015 proposed a new methodology for video inpainting with freehand camera in short-term windows [6].Sixue. Yang et al. (2016). Basing on the color edge detector worked on remove caption " Caption detection and removal from video images with complicated background using intelligent inpainting scheme " [10].LavinaDatey and Rajesh Parashar, (2016). Propose a system that works based on background separation and support vector machine to make a classifier on object's movement. The base of The basis for the work of this proposal is applying a subtraction process between the current frame and next frame. This operation will produce black frames with white pixels represent the motion between the two frames. They calculated a number of white pixels, then employed SUV algorithm to track an object and classify its movement [11].T.Saikumar, et al. (2016) produced Mean-Shift Tracking Algorithm for Salient Object Detection in videos[12].

IsraaHadi Ali and Roa`a M. Al-Airaji presented a method for video inpainting without occlusion [13].

4. Proposed System

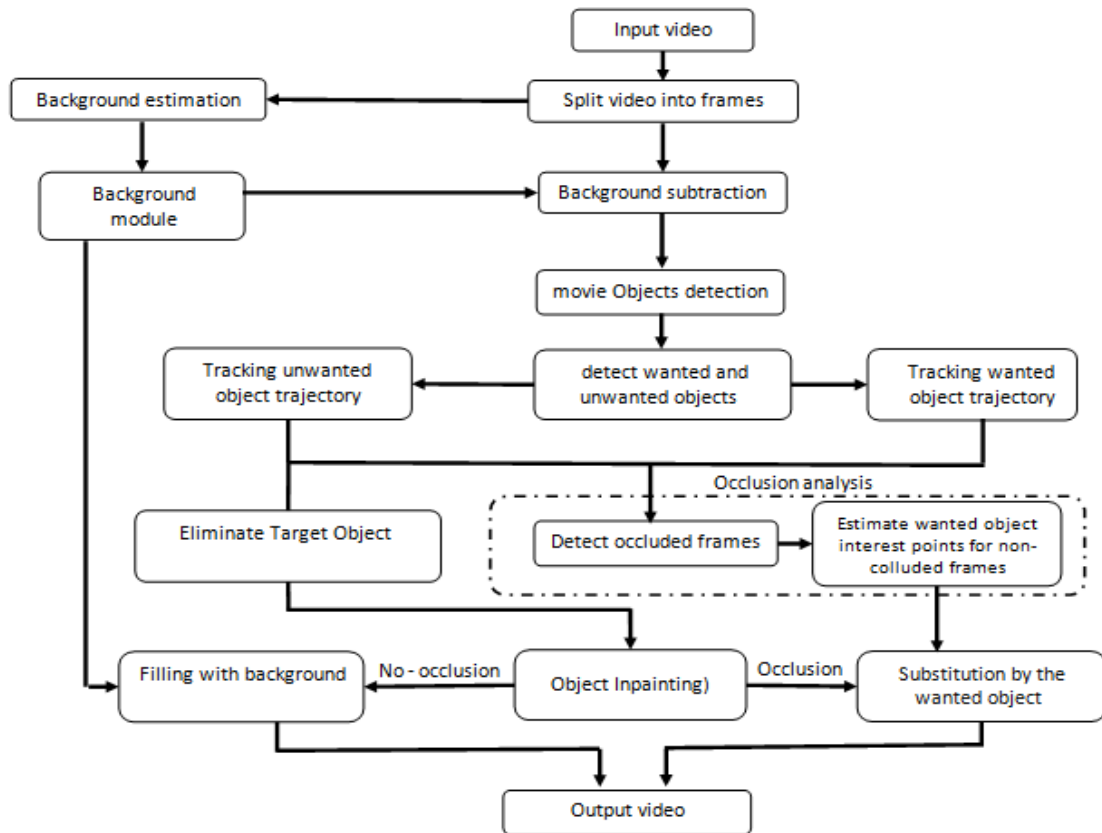Figure (3) shows the main stages of the proposed method

Fig. 3: Block diagram of the proposed method

4.1. Input video

In this stage, the video will be input to the proposed method in order to be split into frames those are will use in the next stage

4.2. Background estimation

The proposed method will find the ideal background model using accumulation histogram. This achieved by four steps. The first step taking each frame that produced from the previous stage and divides it into k blocks each block N*M. The second step takes each block in a frame and find the mean for it and count the frequency for each mean founded for the blocks that occupy the same place. Thirdly take the blocks that have maximum mean frequency and compute variance for each block, then count frequency for each variance. Finally, detect the block that belongs to the background by chose it depending two factors maximum mean frequency then maximum variance frequency. As shown in the figure (4) At the last of these three stages, we will get a compilation of the highest frequency blocks which represent the background model blocks.
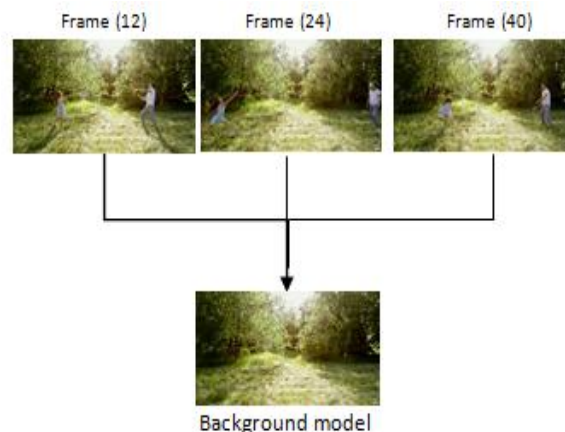

Fig. 4: Background estimation

The previous methods were based on either the mean or median value to detect the block that belongs to the background. Table (1) shows a comparison between three methods (mean, median, mean and variance) dependency in the term of time and precise by using PNSR (Peak signal-to-noise ratio) between the obtained backgrounds and original captured background that was taken by the camera. The comparison implemented for the same video and original captured background

Table 1: Background modeling methods comparisons

Methods	Time	PSNR
Mean	14 sec.	58.36
Median	15 sec.	60.15
Mean & variance	11 sec.	65.23

As it's shown in the above table the implemented method is faster and more accurate in obtaining the background model than other two methods because of the dependence on the two criteria and the large size of the block that used compared to the other two methods.

4.3. Background subtraction

This stage is important to get the foreground objects by subtracting each frame from the background model using that produced from stage process second stage.

Fig. 5: Background Subtraction

4.4. Object segmentation and labeling

The result from the above subtraction is the pixels of moving object. The difference between the background model and each frame represent the foreground object pixels which will be white pixels. The proposed system performs the region growing algorithm for labeling each foreground object in the video. When this stage finished the proposed system will provide with an array that contains information of foreground objects position and pixels in each frame.

4.5. Tracking for objects

After the user selection for the wanted and unwanted object by drawing a rectangle around the wanted object and another one around the unwanted one on the first frame of the original video. the proposed system will perform the mean-shift algorithm to tracking them and save the coordination of the rectangles that surrounded them in order to prepare the information about their position to the next stage.

4.6. Occlusion analysis

This stage includes two steps.
a. Flag the occluded frames: propose system will flag the frames that containing an occlusion between the two objects. This operation performing by comparing the coordination between there rectangles by the coordination of up-right and down-left corners as shown in the figure (6).

Fig. 6: Flag occluded frames

The system will create a flag array that corresponds to the video frames show the state of each frame. This array will be relied upon in the next step to detect the occluded frames.
b. Estimate interest points for the wanted object: the using of interest points will support the system to choose the closest part that used in the inpainting process. This stage will be performed by using the Harris corner algorithm. Harris corner algorithm is used to detect the interest points in the image by computing autocorrelation matrix technique.The system estimates the interest point for the wanted region for all non-occluded frames. In the inpainting process, the system will inpaint the missing part by filling its pixels from frames that have best interest points matching.

4.6. Eliminate unwanted object and pixels filling

Depending on object trajectory this process will detect the pixels of unwanted object in each frame and replace are with two chooses of pixels according to the occlusion state in each frame either background pixels by taking are from background model or object pixels by taking are from best matching of interest points between the appeared part of wanted object and non-occluded frames. The occlusion checking will be not just on whole

unwanted object but even on unwanted pixels. That means the filling object will be mixed between background and the wanted object in cases that containing partial occlusion.

5. Results and Discussion

By taking video file(the boy and his sister) with 320 frames as an example for applying in the proposed system. Which captured by a stationary camera. Containing two objects the first one semi-moving (the boy) as a wanted object and the second one which moves in front of the first one as the unwanted object (the girl).

Fig. 7: original video (The Boy and his Sister)

Fig. 8: Random frames from the video

The proposed system splits the video into frames and starts applying the first stage to obtain the background model as shown in the figure (8).

Fig. 9: Background model

By using the subtraction method between the video frames and the background model the proposed system will construct a binary array for frames that detect the moving foreground object then implement the third stage on it in order to segment and label the foreground object as shown in both figure (10)

Fig. 10: Background subtraction

After that, the user selects the wanted and unwanted objects by drawing a rectangle by the mouse of both objects as shown in the figure (11).

Fig. 11: Objects Selection

The proposed system will perform the mean shift tracker on two objects and detect the non-occluded frames then compute the interest points for the wanted object. The occlusion state determined by making a comparison between the two rectangles by the coordination of up-right and down-left corners. Table (2) shows the position of each rectangle in a sample of frames for this video.

Table 2: The coordination of objects positions

Frame Number	Unwanted object target		Wanted object target		Occlusion state
	Up-right	Down-left	Up-right	Down-left	
1	(63, 270)	(156, 302)	(55,144)	(152,181)	Non-occluded

35	(74, 207)	(167, 239)	(57,74)	(154,111)	Non-occluded
80	(63, 149)	(156, 181)	(59,132)	(156,169)	occluded
90	(63, 110)	(156, 142)	(59,137)	(156,174)	occluded
200	(69, 25)	(162, 57)	(59,195)	(156,232)	Non-occluded
245	(59, 128)	(152, 160)	(57,130)	(154,167)	occluded
257	(66, 199)	(159, 231)	(62,172)	(159,209)	occluded
300	(68, 238)	(161, 315)	(64,229)	(161,266)	Non-occluded

Table (3) show how the occlusion flag array will be for the same frames.

Table 3: The occlusion flag for video frames

Frame no	1	35	80	90	200	245	257	300
Flag	0	0	1	1	0	1	1	0

Then the system implements the inpainting process for occluded and non-occluded frames by filling holes by either background (blue area) or wanted object pixel (green area) or both According to the occlusion manner. As shown in the figure (12).
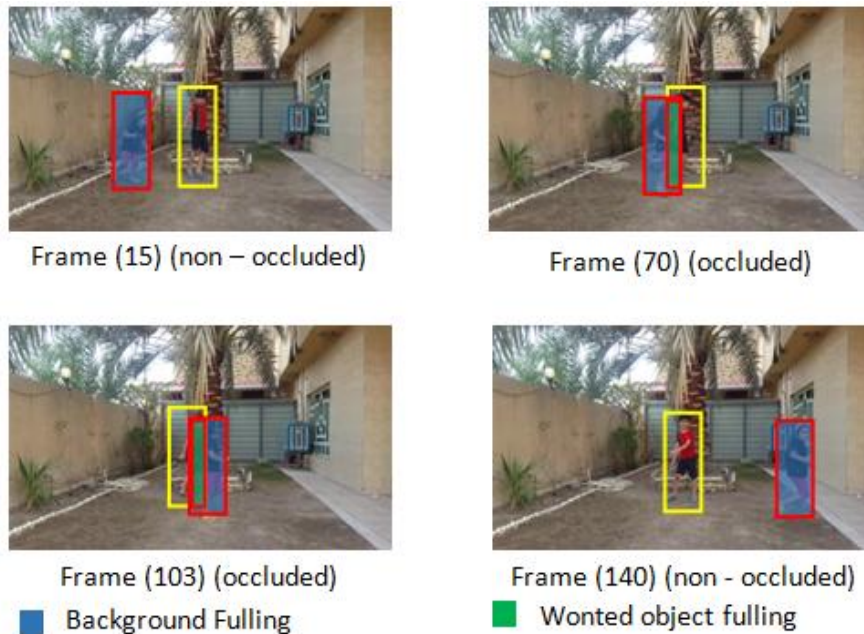

Frame (15) (non – occluded)

Frame (70) (occluded)

Frame (103) (occluded)

Frame (140) (non - occluded)

■ Background Fulling   ■ Wonted object fulling

Fig. 12: unwanted object inpainting

The result of the proposed system after inpainting process shown in figure (13).

Frame (15) (non – occluded)

Frame (70) (occluded)

Frame (103) (occluded)

Frame (140) (non - occluded)

Fig. 13: The inpainting system results

At the end of implementing the proposed system, the obtained results were very close to reality. Where there is no indication that an extra object has been deleted from the video environment. The resulting video contained only one object moving smoothly and naturally with no sudden movement movements indicating that the video was modified. The proposed system evaluated by make a comparison between the final result of this case study after deleting the unwanted object and a captured video that already don't contain the unwanted object. This evaluation implemented to additional two methods that performed on the same videos. the

first method The is "Intelligent video inpainting system for texture reconstruction" [6] and the second method is "Video inpainting with short-term windows". The criterion that used for the evaluation process is PNSR value. Below table (4) contains the comparison between the three final videos for a sample of frames.

Table 4: Evaluation of the final result

Frame Number	PSNR			Occlusion state
	Proposed system	First Method	second Method	
1	40.75	39.82	40.80	Non-occluded
35	39.38	39.3	38.8	Non-occluded
80	34.68	30.26	27.34	occluded
90	35.2	31.42	28.91	occluded
200	41.17	40.8	40.25	Non-occluded
245	32.65	28.35	26.78	occluded
257	38.21	34.92	32.67	occluded
300	40.5	39.26	38.48	Non-occluded

As shown in the table above the PNSR values were close in non-occluded frames. However, the value was the highest of the proposed system due to the accuracy of the background model extracted in the second stage. Which is the basis for the Inpainting process in non-occluded frames. On the other hand the frames that have an occlusion, the PNSR value of the proposed system exceeded other systems by a higher margin. That means the ability to choosing the right and real part of the hidden object that used in inpainting process.

6. Conclusion

This system uses a methodology for video inpainting overcoming the most important and difficult challenge which occlusion that may occur in any video. especially when the unwanted object takes the position between the camera and another wanted one. The difficulty appears by how to filling by another object and predict the right pixel of the hidden object in the right place over video and the right movement that kept the video consistency. The system takes the benefit of object movement repeating and filling with pixels for covered places from other uncovered ones that have the similar interest point (similar movement).

References

[1] Criminisi, P. Perez and K. Toyama, "Region filling and object removal by exemplar-based Image inpainting", Transaction on image processing IEEE, Vol. 13, No. 9, (2004).

[2] Kader A. Patwardhan, Guillermo Sapiro and Marcelo Bertalmio, "Video inpainting under constrained camera motion", IEEE transaction on image processing, Vol.16, No.2, (2006).

[3] Anu Rachel Abraham, A. KethsyPrabhavathy, PhD. J. Devi Shree, "A Survey on Video Inpainting", International Journal of Computer Applications, Vol.55 No.9, (2012).

[4] Ali Mosleh, Nizar Bouguila and A. Ben Hamza, "An Automatic inpainting scheme for video text detection and removal", Transaction on image processing IEEE, Vol.22, No.11 (2013).

[5] SamehZarif, IbrahimaFaya and DayangRohaya, "Static object removal from video scene using local similarity", International colloquium on signal processing and its applications IEEE, (2013).

[6] Narendra Bhatewara, Prashant Kumar and Anupam Agrawal, "Intelligent video inpainting system for texture reconstruction", 4th International Conference on Computer and Communication Technology (ICCCT), (2013).

[7] St-Charles, P. and G. Bilodeau. "Improving background subtraction using Local Binary Similarity Patterns". in IEEE Winter Conference on Applications of Computer Vision,(2014).

[8] MouniraEbdelli, Olivier Le Meur, Christine Guillemot, "Video inpainting with short-term windows: application to object removal and error concealment", IEEE Transaction on Image Processing, IEEE, Vol.24 No.10, pp.3034-47, (2015).

[9] Madbouly, A., M. Wafy, and M.-S.M. Mostafa, "Performance Assessment of Feature Detector-Descriptor Combination".International Journal of Computer Science Issues (IJCSI), Vol.12 No.50, p.87(2015).

[10] Sixue. Yang, Juntao. Xue and Yunri. Zong, "Caption detection and removal from videoimages with complicated background using intelligent inpainting scheme", world congress on intelligent control and automation IEEE, (2016).

[11] Datey, L., "Object Detection and Tracking From Video Using Support Vector Machine". International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol.5 No.12, (2016).

[12] T.Saikumar, MamathaNalavapani, "Mean-Shift Tracking Algorithm for Salient Object Detection in videos", International Journal of Advances in Computer Science and Technology, Vol.5 No.8, (2016).

[13] IsraaHadi Ali and Roa`a M. Al-Airaji , "Processing for Intelligent Movie Editor", Asian journal of Information technology, Vol.15 No.16, (2016).

[14] Peng, W., et al., "Harris Scale Invariant Corner Detection Algorithm Based on the Significant Region". International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.9 , No.3,p. 413-420,(2016).

[15] Hassaballah, M., A.A. Abdelmgeid, and H.A. Alshazly, "Image features detection, description and matching, in Image Feature Detectors and Descriptors", Springer. p. 11-45, (2016).

[16] Shweta K. Talmale, N.J.J., "Object Tracking In Images And Videos. International Journal Of Engineering And Computer Science", Vol.5 No.1: p. 5. (2016).

[17] DivyaM ,R.k.A.V., "Single Object Tracking System By Using Labview". International Journal on Recent and Innovation Trends in Computing and Communication,Vol.4 , No.5,p. 5.(2016).

[18] Jignasha H. Patel and Prof. Jignasa N. Patel , "A Survey : Different techniques of video inpainting" , International journal for scientific research & development vol.4 , No.12 (2017).

[19] A.GhanbariTalouki, M. Majdi and S. A. Edalatpanah, "Video Inpainting Using a Contour-based Method in Presence of More than One Moving Objects", International Journal of Advanced Engineering and Management, Vol. 2, No. 2, pp. 37-44, (2017).

[20] Tawfiq A. Al-Asadi and Fanar Ali Joda, "A Survey: Background Modelling and Object Detection Using Local Texture Features", Journal of Engineering and Applied Sciences, Vol.12 No.11, (2017).

[21] IsraaHadi Ali and Qasim Jaleel, "Determining the Best Method to Extract Interest Point for Video Hidden Object Recovery". Journal of Engineering and Applied Sciences, Vol.13 No.1, (2018).

[22] IsraaHadi Ali and M. Mustafa Ehsan, "Video Inpainting Based on Background Modelling". Journal of Engineering and Applied Sciences, Vol.13 No.6, (2018).