VIDEO INPAINTING USING BACKGROUND REGISTRATION

Sanket Sonawane[1], Sahil Sarangdhar[2], Siddhi Bhatt[3], Divya Sharma[4]

[1]*Department of Computer Engineering, Sandip Institute of Technology & Research Center, Nashik, Maharashtra, India.*
[2]*Department of Computer Engineering, Sandip Institute of Technology & Research Center, Nashik, Maharashtra, India*
[3]*Department of Computer Engineering, Sandip Institute of Technology & Research Center, Nashik, Maharashtra, India*
[4]*Department of Computer Engineering, Sandip Institute of Technology & Research Center, Nashik, Maharashtra, India*

Abstract

Video inpainting is a process of repairing the damaged areas of a video or removing any desired part of a video. Dealing with such problems requires a sturdy image inpainting algorithm along with a regeneration technique for filling the missing parts of a video sequence recorded from a static camera. Many automatic techniques for video inpainting are available but most of them are computationally intensive and fail to repair the damaged areas. Also, from the remaining video inpainting algorithms a masked video sequence has to be provided. To overcome this problem, inpainting process is carried out by using the background registration method which is proposed in this paper. The video is first converted into distinct image frames and the first frame is registered. Second, the edges of an object to be removed are detected by comparing the registered frame with each succeeding frame of the video. Next, a masked frame is generated for each time frame. Then the inpainting process is performed separately for each and every time frame of the images. Next, these processed image frames are displayed sequentially, so that it appears as a video.

Keywords: *PCA, KPCA, Criminisi algorithm*

---***---

1. INTRODUCTION

There has been a tremendous increase in the use of cameras worldwide over the last ten years. This increase has resulted in huge collection of data. Applications of cameras have also been increased in various domains. The rise in the number of application of the cameras gave rise to various new concepts in the domain of image processing. Video Inpainting is one such concept.

Video inpainting is the process of filling the missing or damaged parts of a video and reconstructing the video in such a way that the viewers cannot know if that video is automatically generated or not. As compared to image inpainting the search space and number of pixels to be worked and the in a video is colossal. Image inpainting is performed in spatial domain whereas video inpainting is performed in spatial temporal domain. [1]

Video inpainting plays an important role in the field of image processing. There are scores of goals and applications of this technique ranging from the reconstruction of damaged videos and images to the removal of the specific objects in the video. Video inpainting is used to remove objects or restore missing or tainted regions present in a video sequence by utilizing spatial and temporal information from neighboring scenes. [1] The main objective is to create a video sequence in which the inpainted area is merged smoothly in the video thus maintaining visual consistency i.e. no distortion is observed by the human eye when that video is played.

The main difference between the video and image inpainting methods using texture synthesis is in the size and characteristics of the region to be inpainted. For texture synthesis, the region can be much larger with the main focus being the filling in of two-dimensional repeating patterns that must some textures. In contrast, inpainting algorithms concentrate on the filling in of much smaller regions that are characterized by linear structures such as lines and object contours. Criminisi et al. [2] presented a single algorithm to work with both of these cases provided both textures and structures in images are present. [1][2]

Various studies and researches are carried out in the field of video inpainting due to its numerous applications with the main motive being developing an automatic algorithm for removing unwanted objects like any obscene gesture that is deemed inappropriate for the audience but for which re-shooting would be infeasible and expensive.

Restoring the video is another important application of video inpainting. Videos may get deteriorated in due time. These videos may have dust spots or scratches or may have some missing frames. So it is necessary to repair such videos.

These applications were initially performed manually by the professionals i.e. restoration experts which were meticulous, expensive and took a very long time. Hence automatic video restoration/inpainting technique lured both the private and commercial organizations.

Video inpainting though new, is a challenging concept in image processing. Video Inpainting algorithms are usually time-consuming due to large search spaces.

2. PREVIOUS WORK

Although this paper is about video inpainting, various techniques of video inpainting have been adopted. Initially, the basic idea was to use the concepts of image inpainting in video inpainting.

In image inpainting, several attempts have been made by using various techniques. Most of these techniques can be categorized in the following two categories: 1. Variational image inpainting and 2.Texture synthesis method. Also, there are some techniques that are a combination of both variational image inpainting and texture synthesis method. The method of variational image inpainting emphasizes on gradient propagation. These methods mainly focus on image structure continuity and propagation of the structure from the undamaged parts to the damaged or missing parts.[5] One important work was done by Bertalmio (Bertalmio, Sapiro, Caselles and Ballester, 2000) in which the employing method was used. On the other hand, the texture synthesis method tries to copy the structural information or the statistical data of the pixel from the undamaged part to the damaged part. This method was used by Efros in his paper, *Texture synthesis by non parameter sampling* (Efros and Leung, 1999). But the above two methods produced improper or undesirable results. Osher in his paper, *Simultaneous structure and texture image inpainting* (Osher, Sapiro, Vese and Bertalmio, 2003), has described a method combining the above two methods.

In recent years there has been another approach that has produced better results than the above discussed methods. This method uses global Markov random field (MRF) model and has been implemented by (Fidaner, 2008). In the MRF method, the method of image inpainting is considered as a problem of belief propagation. Komodakis has described this method in his paper, *Image completion using global optimization*, 2006. As compared to image inpainting, the total number of works in video inpainting is quite less. The authors (Jia, Tai, Wu, and Tang), in their paper, *Video repairing under variable illumination using cyclic motions*, proposed a complicated pipeline by combining some complex techniques. Along with this additional user interactions were required for distinguishing the different depth layers. In the paper, *Space time completion of video* (Wexler, Shechtman, and Irani, 2007), Wexler considered video inpainting as a problem of global optimization. This method had a slow running time because of its complexity. Lastly, there was a paper by Patwardhan (Patwardhan, Sapiro, and Bertalmio, 2007), in which a pipeline for video inpainting was proposed. This pipeline however, could not perform well in a majority of the cases. So, various attempts are being made for enhancing the running time and the quality of the output.

3. METHODOLOGY

Video inpainting can be considered as image inpainting done on video frames in their proper sequence. The entire process of video inpainting is carried out in the following steps:

1. Dilation Process.
2. Background Subtraction Algorithm.
3. Criminisis' Exemplar based Algorithm.
4. Flood Fill Algorithm.

3.1 Dilation Process

The video is first divided into time frames. These video frames are then smoothened using the dilation filter. Smoothening is done to remove any spots present in the frames. In dilation the operator enlarges the boundaries of the foreground pixels. Hence, the pixels of the foreground grow thus reducing the size of the spots. For this process two inputs are required, first the image to be dilated and second the set of coordinates called as kernel or structuring element. The set of coordinates taken should be small. The kernel decides the accuracy of dilation of the input image. For a binary image, the mathematical definition of this process can be given as:

Let the coordinates of the input image and the coordinate set of the kernel be X and S respectively. Let Sp be the translation of S in a way that its origin is at x. Then dilation of X by S is the set of point p so that the intersection of Sp with X is not empty.

In the dilation process of a binary image, for the kernel the pixels in the foreground are given the value '1' and those in the background are given the value '0'. Then we consider each and every background pixel of the input image and superimpose it with the kernel on top in a way that the position of the input pixel and the origin of the kernel coincide with each other. If there is a pixel in the kernel that coincides with the foreground pixel of the underlying input image, then that background pixel is set with the foreground value. However, if there is no such then the value of that background pixel is kept as is. So, if the dilation effect is set to foreground color then the background pixel having its neighbour a foreground pixel must be the edge of the spot. Thus the foreground region grows reducing the spot.

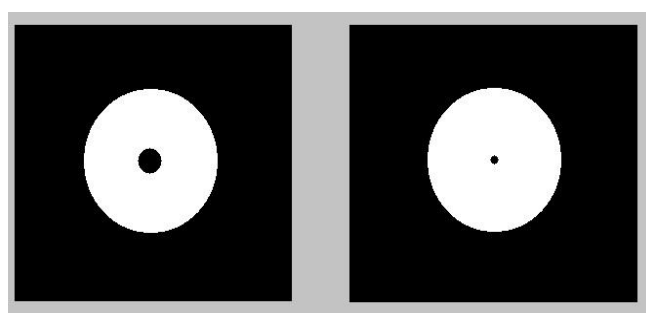

Fig 1: Dilation Process

3.2 Background Subtraction Algorithm

After the image frames have been dilated, the first frame is temporarily stored and used as a reference to compare with the current frames of the video in order to detect the significant difference between the frames. The purpose of this algorithm is to distinguish the moving objects. Thus the moving objects often referred to as foreground are separated from the static scene called as background. This is the basic functioning of the background subtraction algorithm.

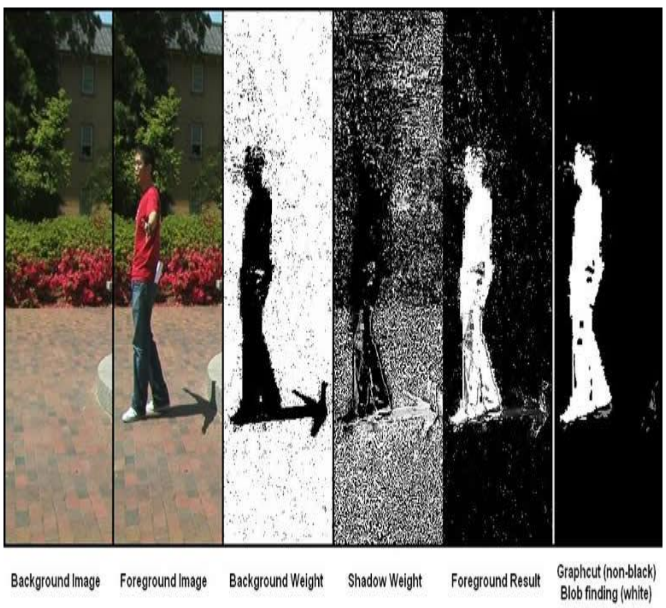

Fig 2: Background Subtraction Process

3.3 Criminisi Examplar based Inpainting

Criminisi use the sampling concept from Efros' approach. The improvement over Efros' is that the new approach takes isophote into consideration, and gave higher priority to those "interesting points" on the boundary of the gap. Those interesting points are parts of linear structures, and thus should be extended into the gap in order to obtain a naturally look. To identify those interesting points, Criminisi gives a priority value to all the pixels on the boundary of the gap. The "interesting points" will get higher priorities according to the algorithm, and thus the linear structures would be extended first.

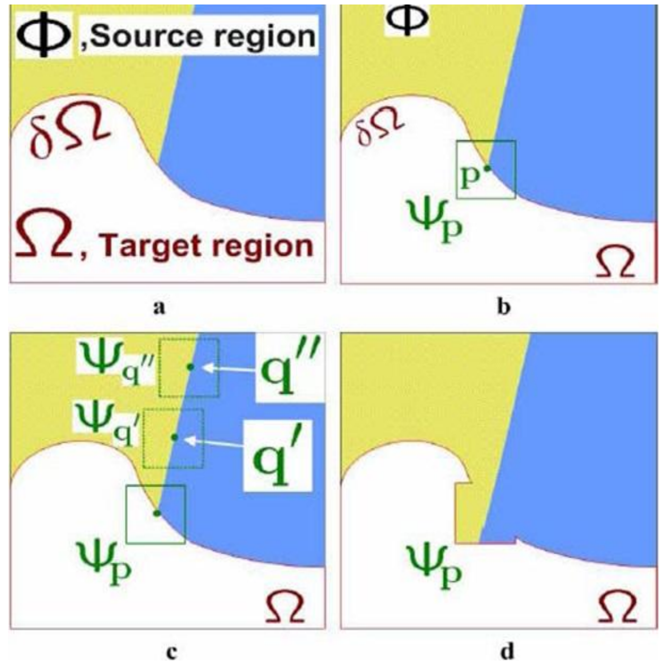

Fig 3: Structure propagation by exemplar-based inpainting

For each pixel on the boundary, a patch is constructed with that pixel at the center. The patch's priority is the product of two elements: a confidence term C(p), and a data term D(p). C(p) describes how many pixels are there in the patch. It is obvious that with more pixels in the patch, we would have a better confidence that a success target patch would be selected. D(p) describes how strong the isophote is hitting the boundary. This term boosts the priority of a patch that an isophote "flows" into. D(p) is especially important, since it encourages linear structures to be synthesized first, and thus propagated securely into the target region. The user will be asked to select a target region, Ω, manually. (a)The contour of the target region is denoted as δΩ. (b)For every point p on the contour δΩ, a patch Ψp is constructed, with p in the center of the patch. A priority is calculated based on how much reliable information around the pixel, as well as the isophote at this point. (c)The patch with the highest priority would be the target to fill. A global search is performed on the whole image to find a patch, Ψq that has most similarities with Ψp. (d) The last step would be copy the pixels from Ψq to fill Ψp. With a new contour, the next round of finding the patch with the highest priority continues, until all the gaps are filled.

Then, we performed a search to find a patch in the source area that is the most similar to the target patch. In our implementation, we calculate the color distance between the non-empty patch pixels at the same position. Sum of Square Difference (SSD) is used to calculate each color channel's difference, and then use SSD to sum up the overall color distance between the pixels. Once the best-fit patch has been found, we copy the color values from the source patch to the target patch. A target patch contains portion of source region and portion of target region. Only those pixels in the target region will be filled. After filling the patch, I update the contour list. Those contour points that fall into the boundary of the target region will be removed from the list. At the same time, the pixels on the boundary of the target patch will be added to the contour list, if those pixels hadn't been filled yet.

This is illustrated by the Fig 1(d). We keep select patches whose centre point is on the contour to be filled. After the filling, we would update the contour list. Eventually, the whole target region will be entirely filled, and the contour list would be empty. At this point, we may have the result picture.

Fig 4: Example of object removal from a frame

3.4 Flood Fill Algorithm

We use this algorithm to select the object having same pixel value. In this algorithm when we click on any point of the image then this point is passed as input to this algorithm as start point with color value of that pixel.

INPUT: Start point = Click point

FILL COLOR = color value of point.

When we apply this algorithm by clicking on any point, the algorithm starts to fill the region in a group of surrounding 8 pixel recursively.

4. CONCLUSIONS

In this paper, we have discussed a method of video inpainting by using the background registration technique. The discussed technique simply compares the background of the generated image frames for object removal. The Criminisi's examplar based algorithm is used for structure regeneration. Hence by using the above technique, we get the inpainted video.

REFERENCES

[1] B. Vidhya and S.Valarmathy, "Novel Video Inpainting Using Patch Sparsity", in IEEE International Conference on Recent Trends in Information Technology, ICRTIT 2011 MIT, Anna University, Chennai, June 3-5, 2011

[2] Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by Examplar-based image inpainting," IEEE Trans. Image Process., vol. 13, pp. 1200–1212, 2004.

[3] J. Jia and C. K. Tang, "Image repairing: Robust image synthesis by adaptive and tensor voting", in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, pp. 643–650, 2003.

[4] Olivier Barnich and Marc Van Droogenbroeck. "ViBe: A Universal Background Subtraction Algorithm for Video Sequences", in IEEE Transactions On Image Processing, Vol. 20, No. 6, June 2011.

[5] NGUYEN Quang Minh Tuan, "Video Inpainting", Department of Computer Science School of Computing, National University of Singapore 2008/09.

[6] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proc. SIGGRAPH, 2000, pp. 417–424.

[7] A. L. Bertozzi, M. Bertalmio and G. Sapiro, "Navier–Strokes, fluid dynamics, and image and video inpainting," in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, pp. 417–424, 2001.

[8] K.A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting of occluding and occluded objects", in Proc. ICIP 2005.Vol. II, pp. 69-72.

[9] Y. Zhang, J. Xiao, and M. Shah, "Motion layer based object removal in videos," 2005 Workshop on Applications of Computer Vision, 2005.