A Survey on Video Inpainting

Anu Rachel Abraham
M.Tech, Department of
Computer Science
Karunya university

A. Kethsy Prabhavathy
M.E, Department of
Computer Science
Karunya university

J. Devi Shree, PhD.
M.E PhD, Department of
Electrical Engineering
Anna University, Chennai

ABSTRACT

Video inpainting is a growing area of research. Many algorithms have been developed so far among which some are directly extended from the image inpainting algorithms. In this survey various patch based and object based video inpainting techniques are reviewed considering its merits and demerits.

Keywords

Video inpainting, object completion, dimensionality reduction, texture synthesis.

1. INTRODUCTION

Inpainting also known as retouching is the process by which we try to fill in the damaged or missing portions of an image in such a way that it is unable for the person seeing the image to find the fault within the image. Inpainting is not constrained to images alone. Rather it can be used for restoring video frames also. Video and image inpainting can be applied on a wide range of applications such as photograph restoration, removing occlusion, video and image modification, adding special effects and much more.

When compared with video inpainting there have been much research going on in the field of image inpainting. This is because when we consider video inpainting we need to cover considerably large number of pixels, and the search space to be considered is also very large. Moreover rather than considering the spatial consistencies within a frame, in video inpainting we must ensure that temporal consistencies between the frames are also maintained.

There are a wide range of image inpainting algorithms developed so far. But applying image inpainting algorithms directly into video inpainting without considering the temporal aspects may not be a good idea. It will cause inconsistencies and undesirable artifacts in the results. Hence video inpainting is often considered as a much more challenging problem when considered to image inpainting. Still there are many cases where image inpainting algorithms are successfully applied for video inpainting algorithms directly with minor extensions. For example belief propagation approaches along with Distance Transform and Principal Component Analysis (PCA) is used for solving the video inpainting problems.

The inpainting algorithms depending on how the images are restored can be categorized into texture-synthesis based methods, partial difference equation based and patches based techniques. In the texture synthesis method w a chose input texture is inserted into the damaged or missing area. In PDE based method information is propagated from the boundary towards the centre of the missing region. The patch based image inpainting was introduced because the above two methods could not maintain the structural information of the image and caused blurring artifacts. Since patch based image

inpainting techniques were successful researchers have extended this concept for video inpainting also.

Though video inpainting is very challenging recently many methods have been proposed in this area. Video inpainting methods can be generally classified into two types:

1) Patch based methods.
2) Object based methods.

2. PATCH BASED TECHNIQUES

Patch based video inpainting methods have been extended from the various patch based image inpainting methods. There are many patch based video inpainting algorithms proposed so far. Patch based algorithms work well when applied for image inpainting but have difficulty in handling spatial and temporal continuity problems when applied for video inpainting.

Among the patch based video inpainting algorithms the first one was the Navier Stokes video inpainting algorithm [1] which was the application of computational fluid dynamics. It was based on Navier-Stokes equation for fluid dynamics. Classical fluid dynamics concepts were used to propagate isophote lines from the exterior continuously. Here image intensity is used as a stream function for a two-dimensional incompressible flow. The information is automatically transported to the inpainting region once the user selects the region to be inpainted. Simultaneous filling in of multiple regions with completely different structures and surrounding backgrounds is allowed. It considered only the spatial aspects of the video and had very limited applications.

In Wexler's approach [2] the video completion is taken as a global optimization problem. It can handle both structured and unstructured dynamic objects automatically. Here frames which do not exist in the dataset are constructed from various space time patches selected from different parts of the video sequence. The global objective function removes local inconsistencies and heuristics of using large patches and it will rank the quality of completion. Fixed size cubes with three dimensions are used as the unit of the similarity measure function which is the Sum of Squared Differences. Though both spatial and temporal in formations are handled simultaneously the multi-scale nature of results will cause blurring and it is very slow.

In [3] a video completion using tracking and fragment merging was proposed by Jia et.al. It was based on texture synthesis where holes are completed fragment by fragment. Texture synthesis methods cannot be directly used in video inpainting algorithms, since amount of data in video sequences is very large and searching for a source fragment in the whole dataset will be slow. This is an iterative process which includes selecting the most apt target at the edge of the hole, whose space-time target fragment is known followed by choosing a source fragment similar to the target fragment from the search region using mean shift algorithm and finally

merging the source and target fragments, thereby reducing the size of the hole. The size of the search space for source fragments was reduced by introducing tracking. Temporal consistency is gained by ensuring consistency of the source fragments. This technique is only suitable when the object motion is periodic and when the images do not change in scale. The filling process may fail if the object tracking is lost, and it may cause artifacts in the hole region. When compared to the Wexler's approach it is more efficient and is faster since the search space is carefully selected.

Cheung et.al introduced video epitome [4] which contain the basic structure and motion characteristics of video which is useful for video inpainting. It is a probabilistic patch based model where patches are used for synthesizing images and videos. Patches from a part of the image is stitched together to synthesize new image with similar texture. It fills missing or occluded regions of the video data. Epitomes provide a representation which retains the natural flow of input data and it provides computational and statistical advantages over the patch libraries. Also epitomes can be trained directly on corrupted or degraded data as long as the data is repetitive. Here the results are of low resolution and can contain over smoothing artifacts.

In video completion for perspective camera under constrained motion [5] Shen et.al proposed an approach to reconstruct missing static background and moving foreground pixels where the camera motion can be either stationary or moving. Here the search space is reduced from 3D to 2D by creating manifolds of the space-time volume. It offers good temporal coherence thereby providing an effective periodic motion. Unlike the approaches described above this approach was based on the motion manifold, which is a good representation of temporal coherence. It is efficient for PTZ (Pan Tilt zoom) cameras for repairing video with projective distortion of moving foreground and global illumination.

Another work [6] done by Patwardhan et.al for inspecting missing parts of a video recorded with moving or stationary camera consisted of a simple pre-processing stages followed by two video inpainting steps. In the pre-processing stage the frames are segmented into foreground and background which helps in the creation of three image mosaics namely foreground mosaic, background mosaic and optical flow mosaic. They will help in reducing the search space of the algorithm. In foreground inpainting the holes caused by the moving objects in the foreground are filled by copying the information from moving foreground in other frames using a priority based approach. In background inpainting the frames are aligned and pixels are copied directly wherever possible, while the remaining pixels are filled by extending the spatial texture synthesis techniques. Though this technique provides good visual effect for each frame and works well in rich and cluttered backgrounds it is not suitable for arbitrary camera motions and do not address the case of complete occlusion. This technique does not maintain temporal continuity along the temporal axis and hence produces flickering artifacts.

In contour based video inpainting [7] it is assumed that the background is stationary and the foreground object has periodic motion. It involves separating the foreground and background using a threshold value. The rectangular window is place around the moving object, and by aligning the reference windows together an image mosaic is obtained. Finally background inpainting is done using exemplar based technique. This method overcomes the over smoothing of background and consumes less amount of time but

maintaining the quality of video keeping the run time low is a real challenge.

In another work [8] the basic technique to fill in a missing part of the video taken from a static camera is done. The main goals of this method include firstly filling in the static background keeping it temporal consistency and secondly to fill the moving foreground keeping the motion globally consistent. Here confidence values are assigned to each pixel in the frame. Background filling involves temporal filling in and spatial filling in. The moving object is completed frame by frame. It performs well when region to be inpainted is large, but the algorithm is suitable only for stationary cameras.

Another exemplar based video inpainting technique without ghost shadow artifacts was proposed [9].Here different motion segments with different temporal continuity takes different candidate patches and are used to inpaint holes after a selected video object is tracked and removed. This algorithm produces very few ghost shadows unlike most of the inpainting algorithms directly applied on videos. Here the earlier algorithms were also modified to handle complicated camera motions. Here multiple objects are removed one by one and hence the user will have to run the program again and again.

3. OBJECT BASED TECHNIQUES

Video inpainting is employed by object based approaches also. Using object based approaches both spatial and temporal aspects consistencies cam be covered simultaneously. Object based approaches produce high quality visual results.

Jia et al. in his paper video repairing under variable illumination using cyclic motions [10] proposed a user assisted video-layer segmentation which decomposes the video into color and illumination videos followed by a tensor voting technique which maintains consistency in both spatio-temporal and illumination domains. Here the available objects are synthesized to recover the occluded objects. Only textures are used in the background since the synthesized objects do not have a real trajectory. Here the repaired video preserves scene structure spatio-temporal consistency and variable illumination but it cannot repair shadows of a damaged object and there are chances of occurrence of incorrect lighting on the repaired movel.

In another work proposed by Zhao et al. [11] an efficient object based video inpainting technique for video recorded by stationary cameras. Background pixels most compatible with the current frame is used for background inpainting and all available objects are used for foreground inpainting. A fixed sliding window includes a set of all continuous objects and defines a similarity function which measures the similarity between two continuous object templates. This technique has the ability to cover large holes and cases where the occluded objects are completely missing from frames. Insufficient number of postures can produce unsatisfactory results .Also this scheme does not propose a systematic way to identify good filling objects.

Another object based video inpainting technique [12] which can handle spatial and temporal aspects simultaneously was proposed by Ling et.al. It could also handle over smoothness artifacts and insufficiency of available postures. Here at first virtual contour construction is done and a posture sequence matching is applied to retrieve the most similar sequence of postures from the available non-occluded postures. Then the key posture selection process is done and posture sequence estimation is made a substring search problem. Synthetic

postures are generated in case there are no good matches among the available postures in the database. It generates objects with good subject quality but cannot cases where the object movement is not periodic.

Lin et al. in his work virtual contour guided video object inpainting using posture mapping and retrieval [13] proposed a technique which mainly has three main steps-virtual contour construction, key posture section and mapping and synthetic posture generation. It is assumed that the trajectory of the occluded objects can be approximated by linear line segmentation during the period of occlusion. Mosaic based schemes and correspondence maps are used for background inpainting. This technique also avoids the problem caused due to the insufficient number of postures. But the synthetic posture generation technique used here is not suitable for generating complex postures and it do not deal with illumination change problems.

Tang et.al proposed another video repairing approach [14] to infer the missing static background and moving foreground due to severe damage or occlusion from a video. Extension of various image repairing methods were used for recovering the background pixels and layer segmentation and homography blending is used to preserve temporal coherence and flickering. Repairing the moving foreground involves the sampling and alignment phases. Here since reference video mosaic is used the system can work for large classes of camera motions. It is not applicable for more complex dynamics and a few spatial and temporal artifacts are observed at times.

A rank minimization approach for video inpainting [15] was proposed by Ding et al. This was a manifold based learning method and local linear embedding was used here to map the image to a low-dimensional space. The method included three steps. Firstly a set of descriptors with the necessary information to reconstruct the frame was found. Secondly, the optimum values for the descriptors were found and finally the frames were reconstructed based on the estimated values. Here Wexler's approach of finding the best matching patch from the adjacent frames was used for filling, difference being in the searching technique used. Descriptors for the missing area are found using Rank Minimization Interpolation (RMI) which uses dynamic models thereby reducing the computational complexity. This technique can be used for video sequences taken by moving cameras also. It is a non-iterative and computationally attractive algorithm and is not restricted to periodic motion. But this technique cannot handle scaling information and causes blurring and ghost image artifacts where the object's motion is not periodic due to the RBF inverse mapping.

Another posture sequence estimation method was proposed by Chih-hung et al. in [16]. This is a manifold based learning approach. Here human objects are considered and involves mainly three main steps-human posture synthesis, graphical model construction and posture sequence estimation. There are also two constraints introduced in this approach to maintain the motion continuity. First one limits the search range and the second constraint limits the search direction hence making the reconstructed motion continuous. In this approach since the postures from the same video are taken to complete or reconstruct the video, it avoids blurring and ghost artifacts. But as in all object based algorithms if the object segmentation is not done accurately it will also cause unpleasant results.

4. COMPARISON OF VARIOUS ALGORITHMS

The above sections show different video inpainting algorithms. Each algorithm has its own merits and demerits. Table 1 below describes the merits and demerits of the various algorithms which have been discussed above.

Table 1. Comparison of Various Video Inpainting Approaches.

APPROACH	YEAR	MERITS	DEMERITS
PATCH BASED TECHNIQUES			
Navier-stokes, fluid dynamics, and image and video inpainting.	2001	Produce good results. There are no color or motion artifacts	Only suitable for filling small non textured holes.
Space-time completion of video.	2004	Can complete frames or portions of frames that never existed in the dataset. Both spatial and temporal information's are handled simultaneously.	Multi-scale nature of results may cause blurring. Computationally expensive. Low resolution videos are only considered.
Video completion using tracking and fragment merging	2005	Fast since holes are completed frame-by-frame. Produce visually appealing results.	Suffers from artifacts at the hole region. Filling process may fail if object tracking is lost. Apt only when object motion is periodic and has no scale change.
Video epitomes	2005	Can synthesize data that do not have structure information.	Results are of low resolution and can contain over-smoothing artifacts.
Video inpainting of occluding and occluded objects	2005	Performs well when region to be inpainted is large. Static background filled-in is consistent	Works only for stationary cameras. Cannot handle completely occluded objects.
Video completion for perspective	2006	Works well for PTZ and stationary	Cannot handle spatial & temporal

camera under constrained motion		camera. Spatial and temporal coherence, as well as periodic motion patterns, are well maintained.	dimensions simultaneously. Computationally expensive, complex& time consuming
Video inpainting under constrained camera motion.	2007	Works well in rich and cluttered backgrounds.	Do not maintain temporal continuity and hence produces flickering artifacts.
Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity	2009	Produces very few ghost shadows unlike other inpainting algorithms. It can deal with different camera motions.	Block matching and selecting continuous blocks are issues which are to be handled.
Contour based video inpainting	2011	There is no over-smoothing on the background. Time spent to find the best patch is reduced.	Results obtained are not acceptable when there is non-stationary background and non-periodic motion. Main challenge is to reduce the running time while keeping the quality.
OBJECT BASED TECHNIQUES			
Video repairing: Inference of foreground and background under severe occlusion.	2004	Can work for a large class of camera motions.	Produce artifacts due to abrupt change of light & shadow. Not applicable for complex structures.
Video repairing under variable illumination using cyclic motions.	2006	Variable illumination& spatio-temporal consistency is preserved.	Works only for periodic motion. Cannot repair shadows of a damaged movel.
Efficient object-based video inpainting.	2006	Can address videos from both static and moving cameras. Can handle large holes including cases where the occluded object is completely missing from several frames.	Results are unsatisfactory if number of postures is insufficient. No systematic way to find a good filling position.
A rank minimization approach to video inpainting.	2007	Can be used for non-periodic motion. Non-iterative and computationally attractive algorithm.	It cannot handle scaling or deformations. It causes blurring and ghost image artifacts if the objects motion is not periodic.
Virtual contour-guided video object inpainting using posture mapping and retrieval.	2011	Avoids over-smoothing artifacts, and compensates insufficient number of postures.	If object moves nonlinearly during occlusion virtual contour constructed may not be accurate. The synthetic posture synthesis is not apt for complex postures.
Human object inpainting using manifold learning-based posture sequence estimation.	2011	Can be used for both periodic and not periodic motion. The reconstructed video has continuous motion.	Inaccurate segmentation will generate artifacts.

5. CONCLUSION

This survey shows the various patch based and object based video inpainting algorithms. Patch based video inpainting algorithms are usually extended from the existing image inpainting algorithms while object based video inpainting algorithms was introduced since most of the patch based algorithms were unable to perform both spatial and temporal aspects simultaneously. Each algorithm has its own merits and demerits. We can choose our algorithm depending on our own requirements.

6. REFERENCES

[1] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 355–362, 2001.

[2] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 463– 476, March 2007.

[3] Y.-T. Jia, S.-M. Hu, and R. R. Martin, "Video completion using tracking and fragment merging," *Proceedings of Pacific Graphics*, vol.21, no. 8–10, pp. 601–610, 2005.

[4] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," *of IEEE Conference on Computer Vision and Pattern Recognition*, pp.42–49, June 2005.

[5] Y. Shen, F. Lu, X. Cao, and H. Foroosh, "Video completion for perspective camera under constrained motion," *Proceedings of IEEE Conference on Pattern Recognition*, pp. 63–66, August 2006.

[6] K. A. Patwardhan, G. Sapiro, and M. Bertalmío, "Video inpainting under constrained camera motion," *IEEE Transactions on Image Processing*, vol.16, no. 2, pp. 545–553, February 2007.

[7] Amanna Ghanbari, Mohsen Soryani,"Contour based video inpainting", *Presented at the IEEE conference on Machine Vision and Image Processing*, 2011.

[8] K. A. Patwardhan, G. Sapiro, and M. Bertalmío, "Video inpainting of occluding and occluded objects," *presented at the IEEE International Conference on Image Processing,* 2005.

[9] T. K. Shih, N. C. Tang, and J.-N. Hwang, "Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity,"*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp.347–360, Mar. 2009.

[10] J. Jia, Y. Tai, T. Wu, and C. Tang, "Video repairing under variable illumination using cyclic motions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 832–883, May 2006.

[11] S.-C. S. Cheung, J. Zhao, and M. V. Venkatesh, "Efficient object-based video inpainting," *Proceedings of IEEE Conference on Image Processing*, pp. 705–708,Oct. 2006.

[12] J. Jia, Y.-W. Tai, T.-P. Wu, and C.-K. Tang, "Video inpainting using posture mapping," *IEEE Transactions on Pattern Analysis and. Machine Intelligence,* vol. 28, no. 5, pp. 832–839, May 2006.

[13] C.-H. Ling, C.-W. Lin, C.-W. Su, Y.-S. Chen, and H.-Y. M. Liao, "Virtual contour-guided video object inpainting using posture mapping and retrieval," *IEEE Transactions on Multimedia,* vol. 13, no. 2, pp. 292–302, Apr.2011.

[14] J. Jia, T. Wu, Y. Tai, and C. Tang, "Video repairing: Inference of foreground and background under severe occlusion," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,*vol. 1, 2004.

[15] Chih-Hung Ling,Yu-Ming Liang,Chia-Wen Lin,Yong-Sheng Chen,Hong-Yuan Mark Liao," Human object inpainting using manifold learning-based posture sequence estimation", *IEEE Transactions on Image Processing,* vol.20,no.11,pp. 3124 - 3135 , 2011.

[16] T. Ding, M. Sznaier, and O. I. Camps, "A rank minimization approach to video inpainting," *Proceedings of IEEE Conference on Computer Vision*, pp. 1–8, October 2007.