Video Completion for Perspective Camera Under Constrained Motion

Yuping Shen, Fei Lu, Xiaochun Cao, Hassan Foroosh
University of Central Florida
Computational Imaging Laboratory
{ypshen,feilu,xccao,foroosh}@cs.ucf.edu

Abstract

This paper presents a novel technique to fill in missing background and moving foreground of a video captured by a static or moving camera. Different from previous efforts which are typically based on processing in the 3D data volume, we slice the volume along the motion manifold of the moving object, and therefore reduce the search space from 3D to 2D, while still preserve the spatial and temporal coherence. In addition to the computational efficiency, based on geometric video analysis, the proposed approach is also able to handle real videos under perspective distortion, as well as common camera motions, such as panning, tilting, and zooming. The experimental results demonstrate that our algorithm performs comparably to 3D search based methods, and however extends the current state-of-the-art repairing techniques to videos with projective effects, as well as illumination changes.

1 Introduction

The problem of video completion is critical to many applications, such as video repairing, movie postproduction and video editing. The goal is to reconstruct the missing pixels in the holes created by damage to the video or removal of selected objects. The key issues in video completion are to keep the spatial-temporal coherence, and the faithful inference of pixels.

Many techniques have been proposed on the video completion problem. Bertalmo *et. al.* [1] proposed a frame-by-frame PDEs based approach, which extended image inpainting techniques to video sequences. Wexler *et. al.* [12] have proposed a space-time volume based method, in which they treated video completion as a global optimization problem, and enforced global spatio-temporal coherence by a well-defined objective function. An extension of this work was proposed in *et. al.* [6], in which tracking and fragment merging are used to reduce search space and enhance completion results. Patwardhan *et. al.* [9] extended the image inpainting techniques in [3] to video inpainting by copying the spatial patch in known area to holes in the background/foreground frame by frame. Jia *et. al.* [5] have proposed an approach to repair videos with periodically moving object under static/moving camera. A motion layers approach, which is working on moving rigid objects, has also been reported by Zhang *et. al.* [14]. Another interesting work on video stabilization and video completion using motion inpainting was presented in [8].

In this paper, we propose a novel approach to reconstruct missing static background and moving foreground pixels in video sequences, where the camera can be either stationary or moving. We slice the space-time volume of a video along the motion manifolds of objects. Then the missing pixels are recovered on the slices such that the spatial and temporal coherence are maintained. The repaired slices are finally combined to construct the repaired video.

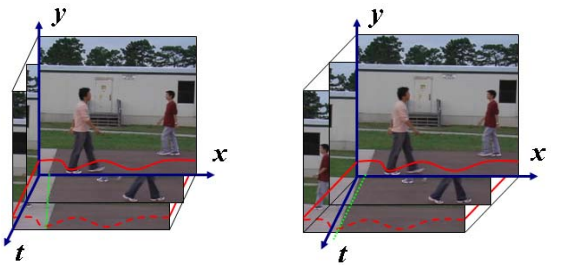
The main contributions of this paper are: (1) Different from previous frame-by-frame [1, 5, 14] and 3D-volume based [12] methods, our method is based on the motion manifold of the video volume, which offers a good representation of temporal coherence information and periodic characteristics of motion, thus providing a novel way to solve the temporal coherence problem and to give a faithful prediction of motion. Compared with the 3D-volume based approaches, our approach has 2D search space, while preserving spatial and temporal coherence of the video. (2) Our approach is able to handle more general video sequences with high resolution, such as non-rigid motion of objects under perspective distortion, nontrivial camera motion, and variable global illumination.

The rest of this paper is organized as follows. The details of our approach are described in Section 2. Experimental results are presented in Section 3, followed by summary and conclusion in Section 4.

2 Video completion Using Motion Manifold

Spatial and temporal coherence are critical in video completion problem. If we treat a video sequence as a space-

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE



(a) Original volume (c) Warped volume

(a) Original manifold (d) Warped manifold

Figure 1. Volume warping under moving camera. After warping, the backgrounds in all frames are aligned to the temporal axis.

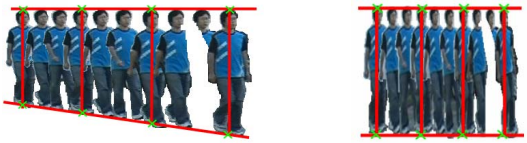
time 3D volume, the goal of keeping temporal coherence becomes equivalent to preserving the smoothness of space-time motion trajectories of pixels. In [12] a volume based approach is proposed to achieve this goal. However, the search in large space-time space is extremely computational expensive. Here we propose to construct manifolds (Figure 1 (c)) of the space-time volume, named as *motion manifolds*, on which the entire trajectory of a pixel is projected to a 2D curve, thus reducing the search space from 3D to 2D. By enforcing spatial and temporal coherence constraints in the propagation, we can restore a visually realistic video sequence with no or acceptable visual artifacts. The procedure of our method is described in the following.

2.1 Volume warping

This step is required only for moving camera. We only consider a subclass of camera motion, when the camera center is fixed (*e.g.*, purely rotating, and zooming camera). These special camera motions are particularly of interest due to their ubiquity in video shots of various sources such as movie and television industry. As is well-known under PTZ camera motion the correspondences between two video frames are defined by a homography that can be readily computed by a set of correspondences. Given the homography $\mathbf{H_i}$ from frame f_i to the first frame f_0, we warp each frame to f_0 and generate a warped spatial-temporal volume V', which is equivalent to the case of static camera(See Fig. 1). A multi-layer scheme proposed in [5] can also be adapted here to get better correspondence.

2.2 Separate foreground and background

We use a coarse-to-fine technique to segment the foreground from the background. First, a coarse boundary of the foreground in the first frame is offered by user, and then tracked by the Mean-Shift tracker [2]. Here the boundary need not to be very accurate but should cover all foreground pixels. After removing the foreground and possibly also some background pixels, we use the technique described in Section 2.4 to repair the background. For each pixel (x, y) in the field of view, we use an adaptive multi-modal background subtraction method [10] to model the pixel color as a mixture of Gaussians and detect the foreground objects.



(a) Original volume (a) Original volume

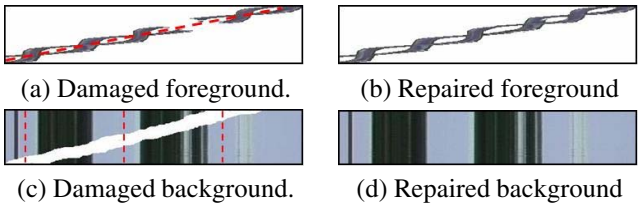
Figure 2. Rectification of foreground volume based on the detected vanishing point.

2.3 Repairing moving foreground

Extracting motion manifolds for object is difficult for arbitrary motions. Here we concentrate on the two common motions in reality, pure translation of rigid objects and periodic motion of rigid and non-rigid objects.

In this stage, the damaged moving foreground is repaired by a two-phase motion volume repairing scheme:

(1) Foreground volume rectification. The goal of this phase is to normalize the foreground volume such that the global trajectory of motion is parallel to the X-T plane. This is required when the trajectory is not parallel to X dimension, or when the scaling of the object also changes during the motion. We use an example in which a person is walking from far to near (see Fig. 2 (a)). In the mosaic generated from the segmented foreground, we pick up the locations $\{p_0, p_1, \ldots, p_n\}$ of the same point P on the object in frames $\{t, t+T, \cdots, t+n \cdot T\}$, where T is the period of the motion cycle and at time t the two legs of the person are folded together. Here we pick up two sets of such points $\mathcal{S}_H$: $\mathcal{S}_H = \{p_i^h\}_0^n$ on the head and $\mathcal{S}_F$: $\mathcal{S}_F = \{p_i^f\}_0^n$ on feet, such that the lines connected by p_i^f and p_i^h, $i = 0, \ldots, n$, are vertical to the ground plane in the world (see Fig. 2 (a) and (b)). To detect $\mathcal{S}_H$ and $\mathcal{S}_F$, we adapt the Eigen analysis scheme described in [7]. $\mathcal{S}_H$ and $\mathcal{S}_F$ are then used to compute the vanishing points along walking and vertical directions. Given these vanishing points, we can rectify the mosaic by applying affine and metric rectification [4] in order. The resulted transformation matrix $\mathbf{H_{rect}}$ is then applied to each frame to get a rectified volume. After the rectification, the X-T slices in the rectified volume is equivalent to the motion manifold of the original video.



(a) Damaged foreground. (b) Repaired foreground

(c) Damaged background. (d) Repaired background

Figure 3. Temporal texture propagation.

(2) Motion pattern propagation. We extend the structure propagation method [11] to the motion manifold by defining a new energy function for minimization, which enforces spatial-temporal coherence in the propagation.

Obtaining propagation paths. The curves parallel to the trajectory of global motion of the object are used as propagation path. On the motion manifold (X-T slice), the trajectory is obtained automatically by applying curve fitting to all foreground points. The red curves in Fig. 3 (a) and (c) are propagation paths.

Texture propagation as global optimization. Given the propagation path L, we first generate a set of M anchor points $\{p_i\}_{i=1}^{M}$ along L in the unknown region, and then sample a set of patches $P = \{P(i)\}_{i=1}^{N}$ along L in the known region such that the distance from the centroid of these patches to L is less than 5 pixels. Similar to structure propagation in [11], the temporal texture propagation is regarded as an optimal graph labelling problem, and dynamic programming technique is used to solve the optimization problem. For a labelling $Y = \{y_i\}_{i=1}^{M}$, *i.e.*, copying $P(y_i)$ to p_i, the energy function is defined as a spatial-temporal coherence error function $E(Y)$:

$$E(Y) = \sum_{i,j=1}^{M} E_1(y_i, y_j) + \sum_{i=1}^{M} (\alpha \cdot E_T(P(y_i)) + E_S(P(y_i))) \tag{1}$$

where $y_i, y_j \in P$, α is a constant and

$$E_1(y_i, y_j) = SSD(Overlap(P(y_i), P(y_j))), \tag{2}$$

$$E_T(P(y_i)) = SSD(Overlap(P(y_i), \overline{\Omega})), \tag{3}$$

The temporal coherence error $E_1(y_i, y_j)$ contributed by overlapping patches $P(y_i)$ and $P(y_j)$ is defined as the normalized SSD between overlapped regions of $P(y_i)$ and $P(y_j)$. The temporal coherence error $E_T(P(y_i))$ introduced by $P(y_i)$ is defined as the normalized SSD between overlapped regions of $P(y_i)$ and the known area $\overline{\Omega}$.

The spatial characteristics of pixels $p = (x, y, t)$ can be represented by $(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y})$ and $(\frac{\partial^2 P}{\partial x^2}, \frac{\partial^2 P}{\partial y^2})$, where P is a spatial-temporal patch. Therefore the spatial coherence error $E_S(P(y_i))$ caused by labelling y_i is defined as:

$$\begin{aligned} E_S(P(y_i)) = \beta_1 \cdot SSD(\tfrac{\partial P(y_i)}{\partial x}, \tfrac{\partial P'(y_i)}{\partial x}) \\ + \beta_2 \cdot SSD(\tfrac{\partial P(y_i)}{\partial y}, \tfrac{\partial P'(y_i)}{\partial y}) \\ + \beta_3 \cdot SSD(\tfrac{\partial^2 P(y_i)}{\partial x^2}, \tfrac{\partial^2 P'(y_i)}{\partial x^2}) \\ + \beta_4 \cdot SSD(\tfrac{\partial^2 P(y_i)}{\partial y^2}, \tfrac{\partial^2 P'(y_i)}{\partial y^2}), \end{aligned} \tag{4}$$

where $\beta_{1...4}$ are constants and $P'(y_i)$ is obtained by copying $P(y_i)$ to the anchor point $p(y_i)$.

The video completion is strongly guided by propagation paths, and the search space here is limited to a narrow area on the 2-D slice along L, therefore is greatly reduced compared with 3-D search space in [12].

After the holes in the rectified volume are filled-in, we need to post-warp it back to the original volume, which is an inverse process of the rectification.

2.4 Repairing background

The static background is a simplified case of moving object, in which the whole background is treated as a single object with no motion. Therefore, the same technique applied on foreground can also be used in static background, though the volume rectification is not required and the propagation paths are straight lines along temporal dimension(See Fig. 3 (c) and (d)).For severely damaged videos, some unfilled holes may exist in the same location of all frames after the previous processing. In this case, image inpainting technique in [3] is used to fill holes in the first frame and then the filled-in pixels are copied to all other frames.

The advantage of our background repairing method is: when the global illumination is changing periodically or the background has minor periodic motion, the layered mosaic based method [5] and motion layer based method [14] may fail, since they are unable to restore these changes of pixels by a copy-and-paste manner from one frame to all others. Our approach works very well in these cases since we also propagate the patterns of background changes to the unknown area. Fig. 4 shows an example of periodically changing illumination. The lighting condition inside the door is changing periodically, and the walking person is removed.

3 Results

The supplemental videos we submitted show the results of video completion in three cases. Fig. 5 shows the result on Walking-Across sequence, where the camera is panning and we want to remove the person closer to the camera. The video volume is warped first such that the motion of background is parallel to temporal dimension. Then the damaged video is separated into foreground and background volumes, which are then repaired individually and finally

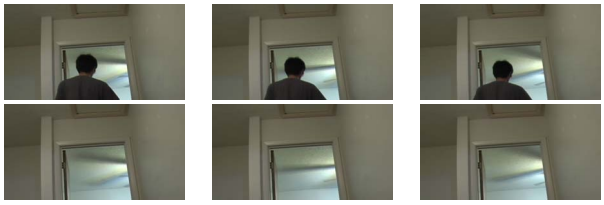


Figure 4. Repairing non-static background. First row: frames from input video, in which lighting condition is changing. Second row: result of removing walking person.

Figure 5. Video completion under camera motion. First row: frames from original video. Second row: repaired video with the closer person removed.

combined together. In the second column of Fig. 5, the person to repair is totally occluded, but well recovered in our result. Fig. 6 shows the result on Perspective-Distortion sequence, in which the occluded person is walking from far to near, which causes perspective distortion. The foreground volume is first extracted and then rectified. We then repair the holes in the rectified foreground volume and then warp it back, and finally combine it with the repaired background volume. Fig. 4 shows the result of Changing-Illumination sequence. The person walking through the door is removed, and the occluded ceiling, which is under changing global illumination caused by the fan, is repaired. These results show that our method works very well in the case of moving camera, projective effect and changing global illumination.

4 Summary and conclusion

In this paper, we propose a novel method for completing damaged videos. Our method works well for stationary and P-T-Z camera. Spatial and temporal coherence, as well as periodic motion patterns, are well maintained in the completed video. Our method can repair video with projective distortion of moving foreground, non-static background with period change of global illumination. In the future, we plan to extend our approach to some more general camera and foreground motion.

Figure 6. Video completion with perspective distortion. First row: frames from original video. Second row: repaired video with the closer person removed.

References

[1] M. Bertalmío, A. L. Bertozzi, and G. Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. In *Proc. IEEE CVPR*, pages 355–362, 2001.

[2] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE CVPR*, pages 142–151, 2000.

[3] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Processing*, 13(9):1200–1212, 2004.

[4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[5] J. Jia, T.-P. Wu, Y.-W. Tai, and C.-K. Tang. Video repairing: Inference of foreground and background under severe occlusion. In *CVPR*, pages 364–371, 2004.

[6] Y. Jia, S. Hu, and R.Martin. Video completion using tracking and fragment merging. *The Visual Computer*, 21(8):601–610, 2005.

[7] F. Lv, T. Zhao, and R. Nevatia. Self-calibration of a camera from video of a walking human. In *Proc. IEEE ICPR*, pages 562–567, 2002.

[8] Y. Matsushita, E. Ofek, X. Tang, and H.-Y. Shum. Full-frame video stabilization. In *Proc. IEEE CVPR*, pages 50–57, 2005.

[9] K. Patwardhan, G. Sapiro, and M. Bertalmío. Video inpainting of occluding and occluded objects. In *Proc. IEEE ICIP*, pages 69–72, 2005.

[10] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):747–757, 2000.

[11] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum. Image completion with structure propagation. *ACM Trans. Graph.*, 24(3):861–868, 2005.

[12] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. In *Proc. IEEE CVPR*, pages 120–127, 2004.

[13] Y. Wexler and D. Simakov. Space-time scene manifolds. In *Proc. IEEE ICCV*, volume 1, pages 858–863, 2005.

[14] Y. Zhang, J. Xiao, and M. Shah. Motion layer based object removal in videos. In *WACV/MOTION*, pages 516–521, 2005.