Panoramic Mosaics with VideoBrush™ *

Shmuel Peleg
Inst. of Computer Science
The Hebrew University
91904 Jerusalem, ISRAEL

Joshua Herman
David Sarnoff Research Center
CN 5300
Princeton, NJ 08540, USA

Abstract

As the field of view of a picture is much smaller than our own visual field of view, it is common to paste together several pictures to create a panoramic mosaic having a larger field of view. While scissors and glue are the tools used in film photography, more sophisticated methods were enabled with digital video.

Panoramic mosaics can be created by special devices which rotate around the camera's optical center (Quicktime VR, Surround Video), or by aligning, and pasting, frames in a video sequence to a single reference frame. Existing mosaicing methods have strong limitations on imaging conditions, and distortions are common.

Manifold projection enables the creation of panoramic mosaics from video sequences under very general conditions. The panoramic mosaic is a projection of the scene into a virtual manifold whose structure depends on the camera's motion. This manifold is more general than the customary projections onto a single image plane or onto a cylinder. VideoBrush, which is a real-time, software only, implementation on a PC, proves the superior quality and speed of this approach.

1 Introduction

The need to combine pictures into panoramic mosaics existed since the beginning of photography, as the camera's field of view is always smaller than the human field of view.

Three major issues are important in image mosaicing:

- Image alignment, which determines the transformation that aligns the images to be combined into a mosaic.
- Image cut and paste is necessary since most regions in the panoramic mosaic are overlapping, and are covered by more than one picture.
- Image blending is necessary to overcome the intensity difference between images, differences that are present even when images are perfectly aligned.

The simplest mosaics are created from a set of images whose mutual displacements are pure image-plane translations. This is approximately the case with some satellite images. Other simple mosaics are created by rotating the camera around its optical center using a special device, and creating a panoramic image which represents the projection of the scene onto a cylinder [7, 15, 14, 13]. Since it is not simple to ensure a pure rotation around the optical center, such mosaics are used only in limited cases.

In more general camera motions, that may include both camera translations and camera rotations, more general transformation for image alignment are used [5, 8, 12, 16, 9]. In all cases images are aligned pairwise, using a parametric transformation like an affine transformation or planar-projective transformation. A reference frame is selected, and all images are aligned with this reference frame and combined to create the panoramic mosaic.

Aligning all frames to a single reference frame is reasonable when the camera is far away and its motion is mainly a translation and a rotation around the optical axis. Significant distortions are created when camera motions includes other rotations.

Manifold Projection overcomes many of the difficulties in photo-mosaicing:

- The projection is defined for almost any arbitrary camera motion and any scene structure. This is enabled by narrowing the goal of image alignment from perfect alignment of all overlapping image regions to alignment only along the seam between the images.
- There are no distortions caused by the alignment to a reference frame. Object size in the panoramic

*This research was mostly done at David Sarnoff Research Center, Princeton, NJ, USA. VideoBrush is a trademark of Sarnoff. Contact E-Mail: peleg@cs.huji.ac.il

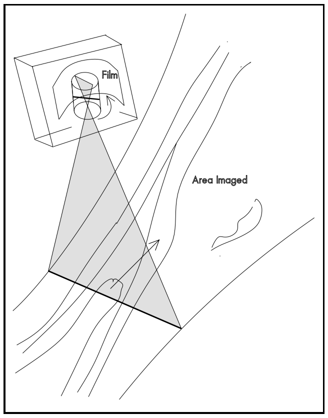

Figure 1: Aerial photography with a 1D scan system.

mosaic is the same as in the original images, and therefore the resolution in the mosaic is the same as the image resolution.

- Computation is simplified as the only image warping used are image-plane translations and rotations.

VideoBrush is an initial implementation of Manifold Projection which performs real-time mosaicing from a video sequence on a PC without any hardware acceleration.

2 Manifold Projection

Manifold Projection simulates the sweeping of the scene with a plane using a one-dimensional sensor array (Figure 1). Such a 1-D sensor can scan the scene by arbitrary combinations of rotations and translations, and in all cases the scanning will result in a sensible panoramic image if we could figure out how to align the incoming 1D image strips. Some satellite images are created by scanning the earth with a 1-D sensor array using a rotating mirror. Since in this case the alignment of the sensors can be done using the location of the satellite and the position of the mirror, panoramic 2D images are easily obtained. Figure 1 is an example of such a 1D scan system.

In more general cases the motion of the sweeping plane may not be known. It seems impossible to align the 1-D image strips coming from an arbitrary plane sweep, but the problem becomes easier when the input is a video sequence. A 2D frame in a video sequence can be regarded as having a 1-D strip somewhere in the center of the image ("center strip"), embedded in the 2D image to facilitate alignment. The motion of the sweeping plane can then be computed from the entire image, and applied on the center-strip for alignment and mosaicing.

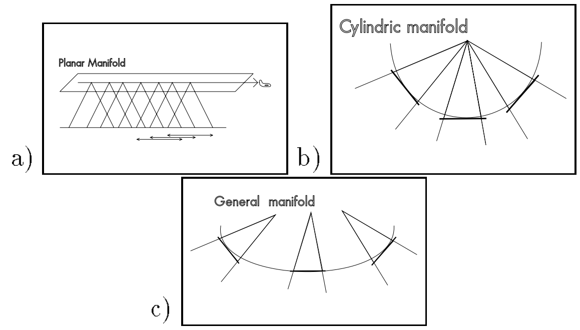

Figure 2: Different cases in Manifold Projection, where the projection is onto a smooth manifold passing through the centers of the image planes used for mosaicing. The camera is located at the tip of the "field-of-view" cone, and the image plane is marked by a bold segment.
a. Pure camera translation: parallel projection onto a plane.
b. Pure camera rotation: projection onto a cylindrical manifold.
c. Combined translation and rotation: the manifold is not simple any more.

The image transformations of the 1D strips generated by the sweeping plane are only rigid transformations: image plane translations and rotations. Therefore, rigid transformations should also be the transformation used in Manifold Projection. It should be noted that general camera motions induce, in general, non-rigid image-plane transformations. However, to simulate the plane sweep only rigid transformations should be used for the center-strip.

The panoramic mosaics generated by combining the aligned 1D center-strips form a new scene-to-image projection, called the Manifold Projection. This is a projection of the scene into a general manifold which is a smooth manifold passing through centers of all image planes constructing the mosaic. In the case of pure camera translations (Figure 2.a), Manifold projections turns out to be a parallel projection onto a plane. In the case of pure camera rotations (Figure 2.b), it is a projection onto a cylinder. But when both camera translations and rotations are involved, as in Figure 2.c, the manifold is not a simple manifold any more. The ability to handle such arbitrary combinations of camera rotations and translations is the major distinction between Manifold Projection and all previous mosaicing approaches.

The type of camera motion has a very significant impact on the type of projection and on the appearance of the panoramic mosaic. In camera panning, where the camera motion is a pure rotation around the Y-axis, the resulting projection is onto a cylinder. This generates a mosaic which is, locally, very similar to every input image.

In a pure camera translation, where the camera moves parallel to the image plane, manifold projection is a semi-parallel projection onto a plane. Semi-parallel means that each center-strip is parallel to the other center-strips, but within the center-strips the projection is still perspective. Parallel projection is very different from a perspective projection in the sense that far-away objects do not appear smaller than close-by objects.

3 Image Alignment

Simulation of scene sweeping by a plane from a given video sequence can be done once the full 3D motion of the camera ("ego-motion") is known [11]. However, the implementation of the manifold projection described in this paper uses only 2D alignment, rather than using full ego-motion analysis. Nevertheless, results are impressive in most cases. It has most of the desired features of the theoretical manifold projection, e.g. that each object in the mosaic appears in the same shape and size as it appears in the video frames, avoiding any scaling, and therefore avoiding the possible associated distortions and loss of resolution. The 2D alignment used therefore compensates only for image translations and rotations. Another assumption in this implementation is that scale changes are minimal: there is no change of focal length, and the effects of forward motion are significantly smaller than the effects of other motions.

To assure that the motion computation will always result in the image motion of a single object, methods similar to [10, 6] were used.

4 Cut and Paste

Combination of the sequence of aligned image frames into a single panoramic mosaic can be done in several ways. In those cases where image alignment is close to perfect, pixel values in the panoramic mosaic can be computed by averaging the corresponding values in all overlapping pixels of the aligned original frames.

When the alignment between images is not perfect, averaging may result in blurring and in deterioration of image quality. In this case it is preferred to select only one of the input images to represent a region in the mosaic. Such a selection should be done to minimize effects of misalignment. The most logical selection is to select from each image that part closest to its center. There are two reasons for that selection:

- Alignment is usually better at the center than at the edges of the pictures.
- Image distortion is minimal at the center of the images.

This selection corresponds to the Voronoi tessellation [3]. Using the Voronoi tessellation for image cut-and-paste also served to minimize visible misalignment due to lens distortions. Voronoi tessellation causes every seam to be at the same distance from the two corresponding image centers. As lens distortions is a radial effect, features that are perpendicular to the seam will be distorted equally on the seam, and therefore will remain aligned regardless of lens distortion.

5 Color Merging in Seams

Changes in image brightness, usually caused by the mechanism of automatic gain control (AGC), cause visible brightness seams in the mosaic between regions covered by different images. These seams should be eliminated in order to get a seamless panorama.

The process of blending the different images into a seamless panorama must smooth all these illumination discontinuities, while preserving image sharpness. A method that fulfills this requirement is described in [4]. In this approach, the images are decomposed into band-pass pyramid levels, and then combined at each band-pass pyramid level. Final reconstruction of the images from the combined band-pass levels give the desired panorama.

6 Examples

Figure 3 shows some panoramic mosaic images created with VideoBrush. More examples can be viewed in "http://www.sarnoff.com/VideoBrush".

7 Concluding Remarks

Manifold Projection enables the fast creation of low-distortion panoramic mosaics under very general camera motions. Implementation under the assumptions of limited change of scale and limited parallax gives unparalleled speed and quality of mosaicing. Future extensions will address the issues of motion parallax, as well as forward motion and zoom which are not addresses in the current scheme.

Bibliography

[1] *ARPA Image Understanding Workshop*, Monterey, California, November 1994. Morgan Kaufmann.

[2] *Fifth International Conference on Computer Vision*, Cambridge, MA, June 1995. IEEE-CS.

Figure 3: Examples of panoramic images using manifold projection. The curved boundary is created by the unstabilized motion of the hand-held camera.

[3] F. Aurenhammer. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, September 1991.

[4] P.J. Burt and E.H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. on Graphics*, 2(4):217–236, October 1983.

[5] P.J. Burt and P. Anandan. Image stabilization by registration to a reference mosaic. In *ARPA Image Understanding Workshop* [1], pages 457–465.

[6] P.J. Burt, R. Hingorani, and R.J. Kolczynski. Mechanisms for isolating component patterns in the sequential analysis of multiple motion. In *Proc. IEEE Workshop on Visual Motion*, pages 187–193, Princeton, NJ, October 1991. IEEE-CS.

[7] Tom R. Halfhill. See you around. *Byte Magazine*, pages 85–90, May 1995.

[8] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P.J. Burt. Real-time scene stabilization and mosaic construction. In *ARPA Image Understanding Workshop* [1], pages 457–465.

[9] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Fifth International Conference on Computer Vision* [2], pages 605–611.

[10] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In G. Sandini, editor, *Second European Conference on Computer Vision*, pages 282–287, Santa Margherita, Italy, May 1992. Springer.

[11] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, WA, June 1994.

[12] P. Jaillon and A. Montanvert. Image mosaicking applied to three-dimensional surfaces. In *12th International Conference on Pattern Recognition*, pages 253–257, Jerusalem, Israel, October 1994. IEEE-CS.

[13] Arun Krishnan and Narendra Ahuja. Panoramic image acquisition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 379–384, San Fransisco, California, June 1996.

[14] Steve Mann and Rosalind Picard. Virtual bellows: Constructing high quality stills from video. In *First IEEE International Conference on Image Processing*, Austin, Texas, November 1994.

[15] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *SIGGRAPH*, Los Angeles, California, August 1995. ACM.

[16] H.S. Sawhney, S. Ayer, and M. Gorkani. Model-based 2D & 3D dominant motion estimation for mosaicing and video representation. In *Fifth International Conference on Computer Vision* [2], pages 583–590.