

Panoramic Mosaics by Manifold Projection

Shmuel Peleg*
Institute 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. Images with a wider field of view can be generated by using fish-eye lens, or 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 more general conditions, and in particular the unrestricted motion of a hand-held camera. 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.

In addition to being more general than traditional mosaics, manifold projection is also computationally efficient, as the only image deformations used are image-plane translations and rotations. Real-time, software only, implementation on a Pentium-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. Also, many times large objects could not be captured in a

single picture. Using a wide field of view (fish-eye) lens can be a solution, but the images obtained have substantial distortions, and mapping an entire scene into the limited resolution of a video camera compromises image quality. A more common solution is photo-mosaicing: aligning, and pasting, frames in a video sequence, which enables a more complete view. Digital photography created new applications for mosaicing [14, 15, 16, 3], which were first implemented for aerial and satellite images.

Three major issues are important in image mosaicing:

- **Image alignment:** Determines the transformation that aligns the images to be combined into a mosaic. Paper photo-mosaicing uses rigid transformations for alignment: picture translations and rotations. Digital processing enables more general transformations, like affine or planar-projective.
- **Image cut and paste:** Most regions in the panoramic mosaic are overlapping, and are covered by more than one picture. The cut and paste process involves either a selection of a single image for each overlapping region, or some kind of a combination of all overlapping images.
- **Image blending:** Necessary to overcome the intensity difference between images, differences that are present even when images are perfectly aligned. Such differences are created by a dynamically changing camera gain.

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. Such translations can either be computed by manually pointing to corresponding points, or by image correlation methods. 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 [5, 13, 12, 11]. Since it is not simple to ensure a pure rotation around the optical center, such mosaics are used only in limited cases.

*This research was mostly done when S. Peleg was on sabbatical at David Sarnoff Research Center, Princeton, NJ, USA. Contact E-Mail: peleg@cs.huji.ac.il.

†J. Herman passed away on October 1, 1996, a few weeks after completing the first version of the mosaicing program VideoBrush.

In more general camera motions, that may include both camera translations and camera rotations, more general transformation for image alignment are used [4, 6, 10, 18, 8]. 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 include 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 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.

Manifold Projection has been initially implemented in a software package which performs real-time mosaicing from a video sequence on a Pentium-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 some cases mosaics generated in this manner can be considered at linear pushbroom cameras [7].

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

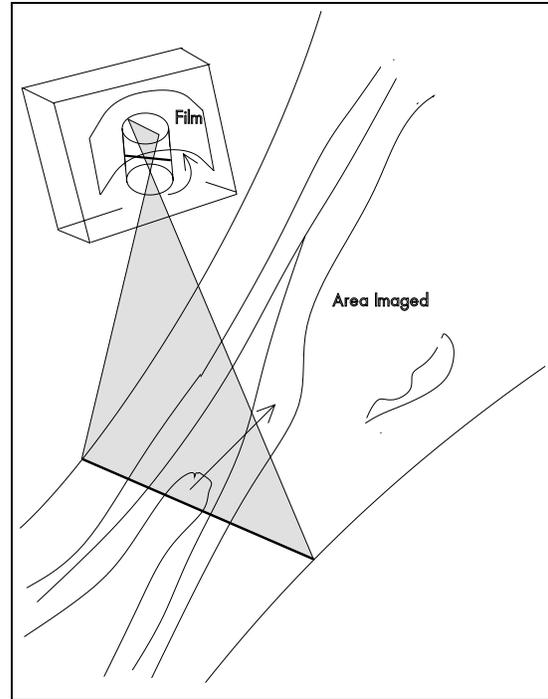


Figure 1. Aerial photography with a 1D scan system (pushbroom camera).

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.

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 Projection 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.

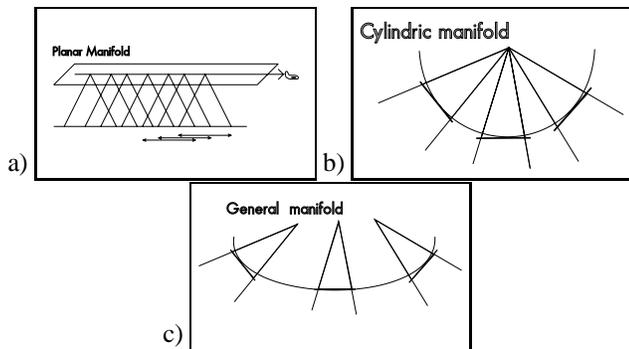


Figure 2. Different cases in Manifold Projection, where the projection is onto a smooth manifold passing through the centers of the image planes used for mosaicing. The camera is located at the tip of the “field-of-view” cone, and the image plane is marked by a bold segment.

- a. In pure camera translation the projection is a parallel projection onto a plane.**
- b. In pure camera rotation the projection is onto a cylindrical manifold.**
- c. When both camera translation and rotation are involved, the manifold is not simple 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 [9]. Given the 3D ego-motion, and given the range of objects whose aspect ratio should be best, it can be determined for each image where

is the 1-D strip to be aligned and pasted into the mosaic as depicted in Figure 1.

However, in certain simple cases the manifold projection can be constructed without ego-motion computation. These are the cases having little motion parallax and little change in scale: pure camera rotations with arbitrary scenes; arbitrary ego-motion with a planar scene (or a very far-away scene); etc. Even in more challenging cases manifold projection can be computed without ego-motion computation as long as a good 2D image alignment method is being used. 2D image alignment will fail to produce good panoramic mosaics only in cases of motion parallax between foreground and background, and when the 2D alignment does not consistently align either the foreground or the background.

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. The following steps were taken:

1. Real-time translation-only alignment of incoming images using correlation. Correlation is performed in a multi-resolution manner, where the results of the correlation in lower resolutions are used to guide the correlation in higher resolutions. This multiresolution correlation enables to detect large displacements, while keeping the computational complexity low.

It is important that this part will be done in real-time, displaying to the user the evolving panorama as the camera is being moved.

2. In order to increase accuracy, additional alignment is done off-line, this time computing both 2D image translation and 2D image rotation. This is done in a multi-resolution fashion by solving 3 equations with 3 unknowns (the motion parameters for image translation and rotation), as described in [2].

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, it

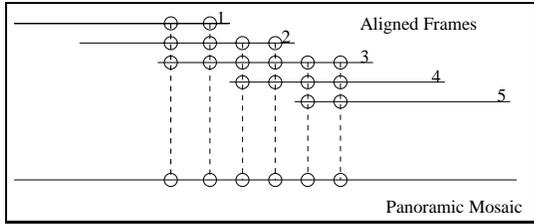


Figure 3. Pixel values in the panoramic mosaic can be computed by averaging the corresponding values in all overlapping frames. In addition to a simple mean, the median is often used to compute the pixel value in the mosaic from all overlapping regions.

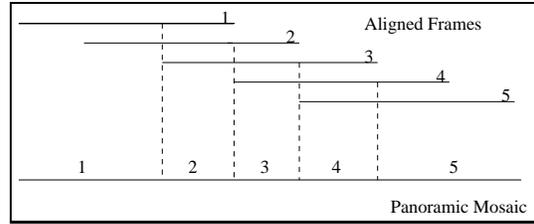


Figure 4. Pixel values in the panoramic mosaic are taken from a single image whose center, after alignment, is closest to the corresponding pixel. All pixels in Region 2 of the mosaic are therefore taken from the strip at the center of Image 2, etc. This construction corresponds in 2D to the Voronoi tessellation.

is desirable to use all overlapping images to produce the mosaic as described in Figure 3. In this case, pixel values in the panoramic mosaic can be computed by averaging the corresponding values in all overlapping pixels of the aligned original frames. In addition to a simple mean, the median is often used to compute the pixel value in the mosaic from all overlapping regions.

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 [1], and is shown in Figure 4. 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 [3]. 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 5 and Figure 6 show some panoramic mosaic images created with an implementation of the manifold projection on the PC. Note that the processing is all done in full color. More pictures, in color, can be viewed in “<http://www.sarnoff.com/VideoBrush>” and in “<http://www.cs.huji.ac.il/~peleg/brush>”.

7 Concluding Remarks

Manifold Projection enables the fast creation of low-distortion panoramic mosaics under very general camera motions. The basic principle is the alignment of the strips which contribute to the mosaic, rather than the alignment of the entire overlap between frames. Each strip undergoes the minimal distortion to be aligned with the neighboring strip, eliminating global distortions.

An implementation under the assumptions of insignificant change of scale and limited parallax gives unparalleled speed and quality of mosaicing. Extensions should address the issues of motion parallax, as well as forward motion and zoom which are not addresses in the current scheme. Initial experiments in this direction are reported in [17].



Figure 5. Examples of panoramic images using manifold projection. The curved boundary is created by the unstabilized motion of the hand-held camera.



Figure 6. Manifold projection with vertical scanning. The curved boundary is created by the unstabilized motion of the hand-held camera.

8 Acknowledgment

Many thanks to colleagues at Sarnoff for providing the right environment for this work. D. Dixon initiated this research by insisting that mosaicing should work real-time on a PC. Encouragement and support was given by P. Anandan, J. Bergen, P. Burt, C. Carlson, G. Gendel, T. Kumar, H. Sawhney, and N. Winarsky.

References

- [1] F. Aurenhammer. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, September 1991.
- [2] J. Bergen, P. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14:886–895, September 1992.
- [3] P. Burt and E. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. on Graphics*, 2(4):217–236, October 1983.
- [4] P. Burt and P. Anandan. Image stabilization by registration to a reference mosaic. In *ARPA Image Understanding Workshop*, pages 457–465, Monterey, California, November 1994. Morgan Kaufmann.
- [5] T. R. Halfhill. See you around. *Byte Magazine*, pages 85–90, May 1995.
- [6] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P. Burt. Real-time scene stabilization and mosaic construction. In *ARPA Image Understanding Workshop*, pages 457–465, Monterey, California, November 1994. Morgan Kaufmann.
- [7] R. Hartley and R. Gupta. Linear pushbroom cameras. In J. Eklundh, editor, *Third European Conference on Computer Vision*, pages 555–566, Stockholm, Sweden, May 1994. Springer.
- [8] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Fifth International Conference on Computer Vision*, pages 605–611, Cambridge, MA, June 1995. IEEE-CS.
- [9] 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.
- [10] 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.
- [11] A. Krishnan and N. Ahuja. Panoramic image acquisition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 379–384, San Francisco, California, June 1996.
- [12] S. Mann and R. Picard. Virtual bellows: Constructing high quality stills from video. In *First IEEE International Conference on Image Processing*, Austin, Texas, November 1994.
- [13] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *SIGGRAPH*, Los Angeles, California, August 1995. ACM.
- [14] D. Milgram. Computer methods for creating photomosaics. *IEEE Trans. on Computers*, C-24:1113–1119, 1975.
- [15] D. Milgram. Adaptive techniques for photomosaicking. *IEEE Trans. on Computers*, C-26:1175–1180, 1977.
- [16] S. Peleg. Elimination of seams from photomosaics. *Computer Graphics and Image Processing*, 16:90–94, May 1981.
- [17] B. Rousso, S. Peleg, and I. Finci. Generalized panoramic mosaics. In *DARPA Image Understanding Workshop*, New Orleans, Louisiana, May 1997. Morgan Kaufmann.
- [18] H. 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*, pages 583–590, Cambridge, MA, June 1995. IEEE-CS.