

Mosaicing on Adaptive Manifolds

Shmuel Peleg, *Member, IEEE*, Benny Rousso, Alex Rav-Acha, and Assaf Zomet

Abstract—Image mosaicing is commonly used to increase the visual field of view by pasting together many images or video frames. Existing mosaicing methods are based on projecting all images onto a predetermined single manifold: A plane is commonly used for a camera translating sideways, a cylinder is used for a panning camera, and a sphere is used for a camera which is both panning and tilting. While different mosaicing methods should therefore be used for different types of camera motion, more general types of camera motion, such as forward motion, are practically impossible for traditional mosaicing. A new methodology to allow image mosaicing in more general cases of camera motion is presented. Mosaicing is performed by projecting thin strips from the images onto manifolds which are adapted to the camera motion. While the limitations of existing mosaicing techniques are a result of using predetermined manifolds, the use of more general manifolds overcomes these limitations.

Index Terms—Mosaicing, motion analysis, image alignment.

1 INTRODUCTION

CREATING pictures having larger field of view by combining many smaller images is common since the beginning of photography, as the camera's field of view is smaller than the human field of view. In addition, some large objects cannot be captured in a single picture as is the case in aerial photography. Using omnidirectional cameras [19] can provide a partial solution, but capturing a wide field of view with the limited resolution of a video camera compromises image resolution. A common solution is photo-mosaicing: aligning and pasting pictures, or frames in a video sequence, to create a wider view. Digital photography enabled new implementations for mosaicing [17], [18], [20], [3], [10], [28], which were first applied to aerial and satellite images and later used for scene and object representation.

The simplest mosaics are created by panning the camera around its optical center, in which case the panoramic image can be created on a cylindrical or a spherical manifold [15], [5], [16], [29], [12], [28]. The original images, which are formed by a perspective projection onto a plane, are warped to be perspectively projected into an appropriate cylinder, where they can be combined to a full 360 degrees panorama, as in Fig. 1. While the limitations to pure sideways rotation enable easy mosaicing without the problems of motion parallax, this approach cannot be used with other camera motions.

Simple mosaicing is also possible from a set of images whose mutual displacements are pure image-plane translations. This is the case for a translating camera orthogonally viewing a planar scene. For somewhat more general camera motions, more general transformation for image

alignment can be used, like a global affine transformation or a planar-projective transformation [4], [7], [11], [26], [10]. In most cases, images are aligned pairwise using the global parametric transformation, a reference frame is selected, and all images are aligned to this reference frame and combined to create the panoramic mosaic. Such methods imply the perspective projection of all the images onto the planar manifold corresponding to the image plane of the reference frame. Using a planar manifold and aligning all frames to a single reference frame is reasonable only when there are no considerable depth differences in the scene and the camera motion is mainly a sideways translation and rotation around the optical axis. Significant distortions are created, for example, in more general camera motion which includes sideways rotation or when there are scale changes in the image due to camera translation, as shown in Fig. 2.

Most restrictions on the motion of the camera used for mosaicing can be eliminated by using a manifold whose shape is determined adaptively during the mosaicing process. To enable undistorted mosaicing the selected manifold should have the property that, after projecting the images onto the manifold, the *optical flow*¹ vectors become approximately uniform: parallel to each other and of equal magnitude. Typical cases for this optical flow are sideways image translation, where the manifold is a plane and a panning camera, where the manifold is a vertical cylinder and the optical flow in a center of the image is approximately uniform. Different manifolds relating to more general camera motions will be presented in this paper. It will be shown, for example, that the general case of a translating camera can be handled using a cylindrical manifold whose axis is the direction of motion.

When a moving camera captures a general static scene, the optical flow depends on the scene depth, making mosaicing difficult. A technique for mosaicing such scenes is the "slit camera," or the "pushbroom camera," used in

• S. Peleg, A. Rav-Acha, and A. Zomet are with the School of Computer Science and Engineering, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel. E-mail: {peleg, alexis, zomet}@cs.huji.ac.il.

• B. Rousso is with Impulse Dynamic, Israel. E-mail: Benny@Impulse.co.il.

Manuscript received 4 Dec. 1998; revised 19 Apr. 2000; accepted 23 July 2000.

Recommended for acceptance by R. Szeliski.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108405.

1. Image motion is represented by the optical flow: the displacement vectors associated with each image point which specify the location of the image point in the next frame relative to its location in the current frame.

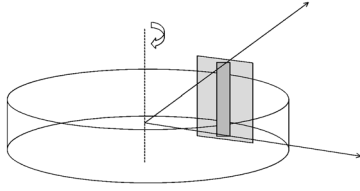


Fig. 1. A panoramic image can be generated from a panning camera by combining the images on the surface of a cylinder.

aerial photography [8]. This camera can be modeled as a 1D sensor array which collects strips by “sweeping” the scene, as described in Fig. 3.

The imaging process of the pushbroom camera can be modeled by a multiperspective projection: For each strip, the projection is perspective, while different strips may be acquired from different viewpoints. Thus, in the direction of the strips, the projection is perspective, while, in the scanning direction, the projection is parallel. Since under parallel projection there is no parallax, the strips in the resulting mosaic are aligned at the seams.

All mosaicing techniques described in this paper process video sequences acquired by perspective cameras moving on a smooth route. They approximate the mosaic image which would have been acquired by a pushbroom camera moving on the same route. This is done by reprojecting thin strips from the images onto a manifold such that the optical flow becomes approximately uniform: parallel and of equal magnitude. Usually, both the manifold and the reprojection transformation are computed implicitly.

The implementations of manifold mosaicing presented here are based on the motion computed between the images. Unlike other methods for multiperspective mosaics [22], [30], the mosaics are usually constructed without knowing or recovering the structure of the scene and without knowing explicitly the full motion and calibration of the camera.

The warping of the strips to induce parallel and uniform optical flow can sometimes be achieved in two steps: First, the images are rectified [9], [14] to get a parallel and horizontal optical flow. After rectification, the images are resampled by interpolating the coordinates linearly along the flow. This method has several drawbacks: 1) Image rectification assumes camera translation. Mosaicing must handle pure rotations. 2) Image rectification rotates the image planes to be parallel to the camera translation vector. In order to rectify more than two images, it should be assumed that the camera translates in a constant direction.

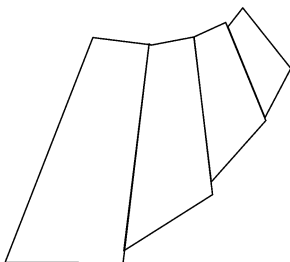


Fig. 2. Mosaicing images of a planar scene under a tilted view by warping the input images to the coordinate system of the reference image. The resulting mosaic will be curled.

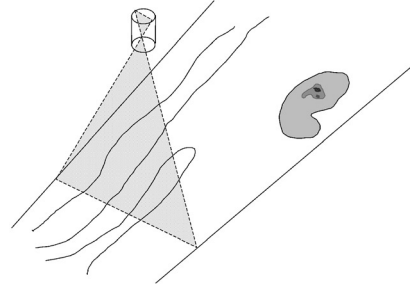


Fig. 3. An aerial pushbroom camera.

The projection of the strips onto the manifold can be viewed as a “Rectification” onto a nonplanar manifold.

Each region in the mosaic is taken from that image where it is captured at highest resolution. While this could have been neglected in the traditional mosaicing, which does not allow any scale changes, it is critical for general camera motions where, for example, a region is seen at higher resolution when closer.

Examples of manifold mosaicing using strips will be given for cases of almost uniform image translations caused by a panning camera [21], for a forward moving camera [24], [23], and for 2D planar projective transformation caused by a tilted panning camera or a tilted camera translating in a planar scene [33]. Mosaics generated in this manner can be considered as similar to the vertical “slits” [31] or the “linear push-broom cameras” [8]. However, unlike the straight “slit” or “broom,” the broom in manifold mosaicing may change its shape from a straight line to a circular arc to become mostly perpendicular to the optical flow. This demonstrates the flexibility in mosaicing with strips, and its adaptation to changes in camera motion.

2 MOSAICING WITH STRIPS

Most existing mosaicing systems align and combine full images or video frames [28], [10], [26], [13]. The combination of full frames into mosaics introduces some difficulties:

- It is almost impossible to accurately align complete frames due to lens distortion, motion parallax, moving objects, etc. This results in “ghosting” or blurring when the mosaic is constructed. Similar

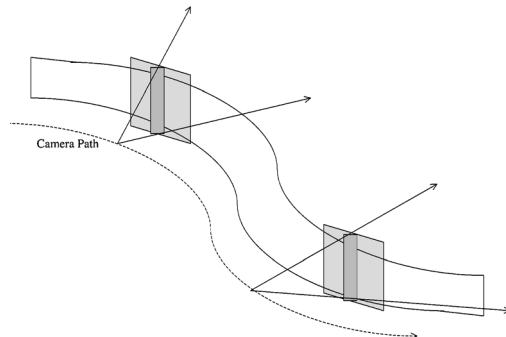


Fig. 4. A panoramic image generated from a vertical “slit” moving on a smooth path on a horizontal plane.

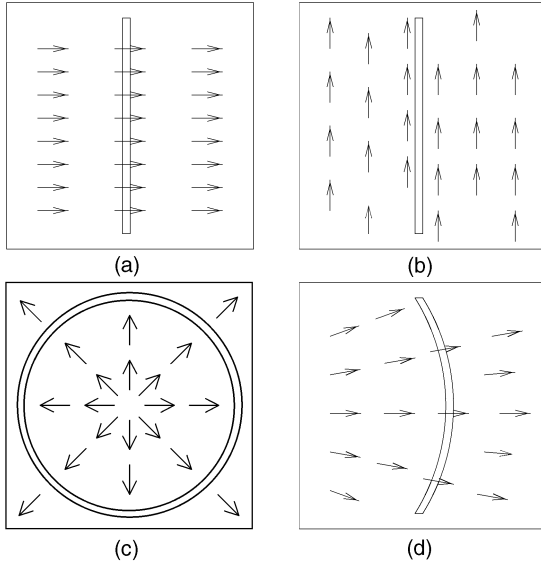


Fig. 5. The mosaicing process and the direction of the optical flow. (a) A vertical slit is optimal when the optical flow is horizontal. (b) A vertical slit is useless when the optical flow is vertical. (c) A circular slit is optimal when the optical flow is radial. (d) For general motion, optimal slits should be perpendicular to the optical flow and bent accordingly.

artifacts may appear due to accumulative error in the motion between nonsuccessive frames.

- It is difficult to determine the mosaicing manifold, e.g., if all images are aligned to one reference image, different reference images will give different mosaics. In the case of a projection onto a cylinder, it is important that the camera motion is a pure sideways rotation.

In order to overcome the above difficulties, we propose using mosaicing with strips. Thin strips are taken from the input images and placed, after warping, onto the mosaic. We present implementations of manifold mosaicing for several types of scenarios. All the implementations are satisfying the following principles:

- The width of the strips should be proportional to the motion.
- The borders of the strips should match. For example, when moving to the right, the right border of the

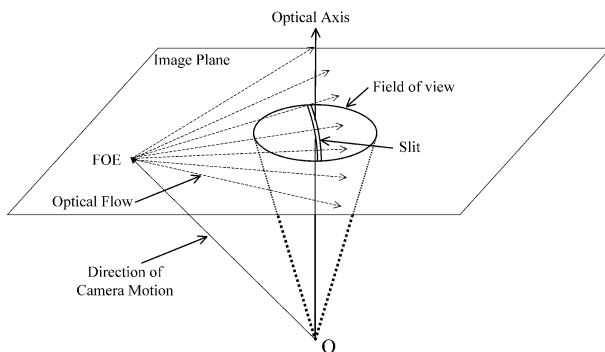


Fig. 6. Determining the shape of the slit with camera translation. In this case, the optimal slit will be the longest circular section having its center at the FOE and passing through the field of view. This is the longest curve in the FOV that is perpendicular to the optical flow.

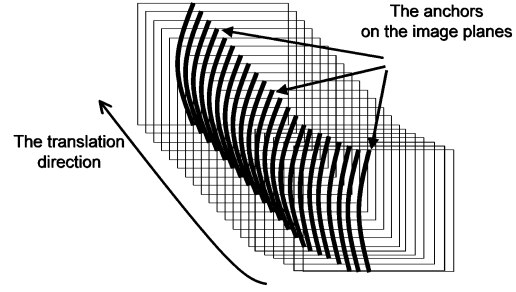


Fig. 7. The mosaicing manifold is formed by the motion of the camera and the shape of the anchors in the images.

strips taken from image I_n should correspond to the left border of the strip taken from image I_{n+1} . This is necessary in order to get a continuous mosaic.

- The collected strips should be warped and pasted into the mosaic image such that, after warping, their optical flow becomes parallel to the direction in which the panoramic image is constructed and of equal magnitude.
- In order to avoid global resizing, each image strip includes a feature (the *anchor*) which does not change under the warping.
- It is recommended to have the *anchor* perpendicular to the optical flow. This maximizes the information collected by the virtual 1D sensor array.
- It is recommended to take the strips from the center of the image to reduce effects of lens distortion.

The warping of the strips into the mosaic is equivalent to reprojection onto a manifold without the explicit computation of the manifold. No accumulative distortions are encountered as each strip contains an anchor and is warped to match just its neighboring strips. The anchors are placed parallel along the mosaic, completing a parallel projection in the direction of the motion of the camera. Thus, the anchors are the realization of the “slit” or “broom” of the virtual camera. Assuming the motion between successive

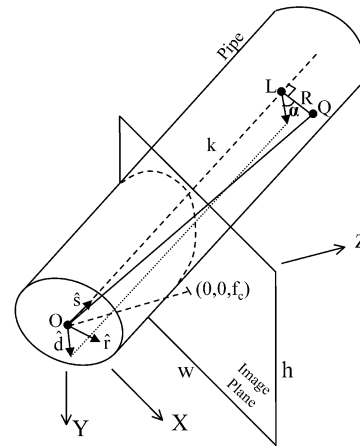


Fig. 8. The pipe projection geometry. The axes of the pipe \hat{s} pass through the optical center $O = (0, 0, 0)$ and through the FOE S . \hat{d} and \hat{r} are unit vectors chosen to form a Cartesian coordinate system together with \hat{s} . The image point $P = (x, y, f_c)$ is projected onto its corresponding point Q on the pipe. Q is represented by k , the position along the axis \hat{s} , and α , the angle from \hat{d} .

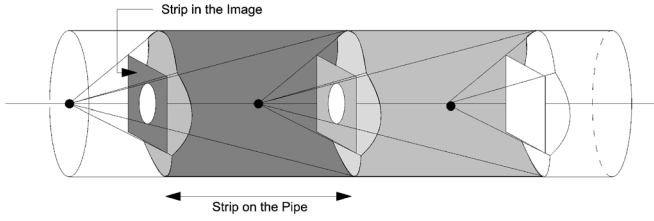


Fig. 9. The relation between pipe projection and mosaicing with strips. The strip on the pipe corresponds to a strip on the image that is warped to achieve parallel optical flow. Strips are taken from the images in which best resolution is obtained.

frames is small, canceling the parallax by linear interpolation of the coordinates along the flow is a satisfying approximation for the narrow gaps between the anchors. When the strips are wide, it is possible to reduce the parallax and simulate the parallel projection by generating intermediate views [27], [6]. The introduction of intermediate views simulates a denser image sequence, where the strips are narrower, with smaller discontinuities due to motion parallax.

In order to maximize the information collection rate of the 1D sensor, it is recommended to have the slit perpendicular to the optical flow, as illustrated in Fig. 5. An example for the determination of the shape of the slit is given for image motion generated by pure translation of the camera, as shown in Fig. 6. In this case, the image motion can be described by a radial optical flow emanating from the focus of expansion (FOE) and the field of view of the camera (FOV) can be described as a circle on the image plane. The optimal slit is the longest circular section having its center at the FOE and passing through the FOV. This is the longest curve in the FOV that is perpendicular to the optical flow.

The definition of the scanning slit as perpendicular to the optical flow is very simple for some cases.

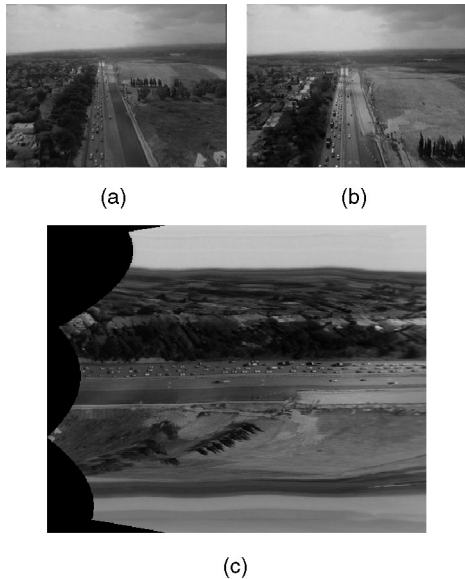


Fig. 10. Pipe projection with a forward moving camera and a planar scene. The focus of expansion is inside the image. (a) and (b) are individual frames from a video sequence. (c) The pipe mosaic of the sequence. The road is always at full resolution.

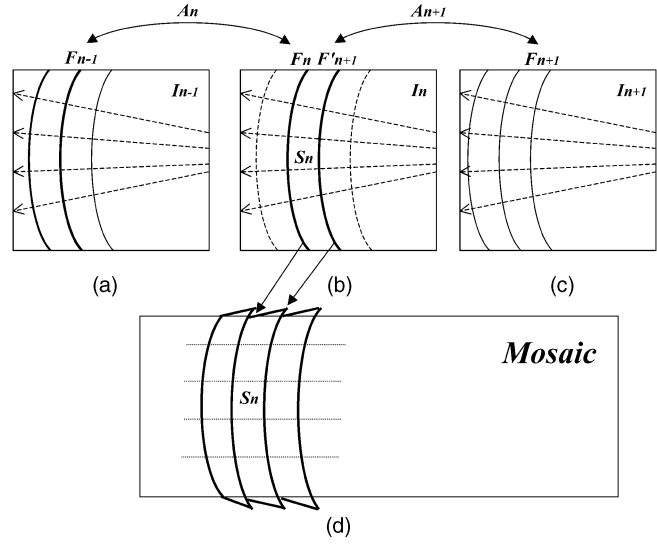


Fig. 11. Cutting and pasting strips. (a), (b), and (c) Strips are perpendicular to the optical flow. (d) Strips are warped and pasted so that their back side is fixed and their front side is warped to match the back side of the next strip.

- In sideways image motion, the optimal slit is vertical (Fig. 5a).
- In image scaling (zoom) and in forward motion, the optimal slit is a circle (Fig. 5c).
- In image motion generated by camera translation, the optimal slit is a circular arc (Fig. 5d).

Image motion is usually more general than these simple special cases. However, in most cases, slits that are straight lines, circular curves, or elliptic curves are sufficient for mosaicing.

The shape of the slit determines the shape of the manifold on which the mosaic is created. The circular slit, for example (Fig. 5c), forms a cylindrical manifold. In Fig. 7, it is demonstrated how the anchors form the manifold.

When mosaicing with strips, changes in image brightness, usually caused by automatic gain control, cause visible brightness seams between strips. These illumination discontinuities can be eliminated by blending the different images, for example, using the Laplacian pyramid [3].

3 IMPLEMENTATION EXAMPLES

In this section, implementations of manifold mosaicing are described. First, the simple case of a camera moving on a

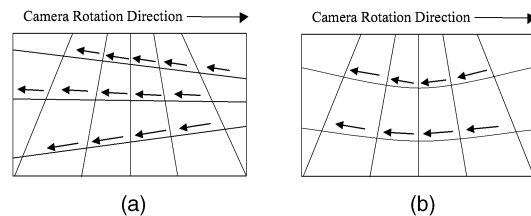


Fig. 12. The image of vertical and horizontal lines: (a) Tilted camera moving horizontally and viewing a planar scene. (b) Tilted camera panning horizontally. The lines are on a cylinder centered on the rotation axis.

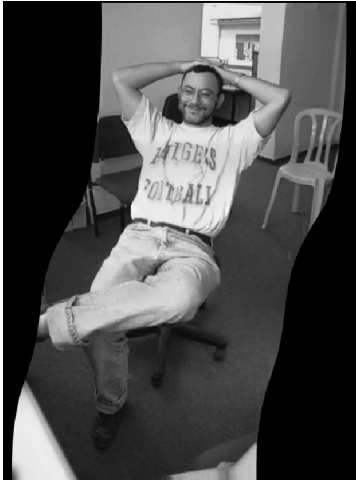


Fig. 13. Manifold mosaicing with vertical scanning. The curved boundary is created by the unstabilized motion of the hand-held camera.

horizontal path [31], [21], where the optical flow can be approximated at the center as horizontal and uniform.

Then, we describe two algorithms for mosaicing from a forward moving camera. The first algorithm constructs the mosaic by explicitly reprojecting the images onto the pipe manifold [24]. The second algorithm performs the reprojection implicitly by pasting strips onto the mosaic image [23].

Finally, an algorithm is described for cases in which the image motion can be described by a 2D homography. This algorithm handles the case of a tilted panning camera and the case of a tilted camera translating in a planar scene.

3.1 Horizontal Motion

A mosaicing approach which creates, for the first time, multiperspective panoramic views on general manifolds has been described in [31]. It was assumed that the camera motion and calibration are known from an external device and that its motion is a combination of translation sideways on a plane and panning. When the camera pans, the motion in the center of the image is approximately horizontal and uniform. Similar image motion occurs when the camera translates sideways, assuming there are no considerable depth differences. Thus, a mosaic can be constructed by copying thin vertical strips from the input images and pasting them side by side onto the mosaic image. This is

equivalent to projecting the strips onto a manifold which is a combination of cylindrical patches (when the camera pans, as in Fig. 1) and planar patches (when the camera only translates). For a more general motion combining panning and horizontal translation, the manifold follows the center strip of the images as shown in Fig. 1 and Fig. 4 and the process can be viewed as a 1D vertical slit camera which is scanning the scene.

Manifold mosaicing can be implemented without knowing the camera motion and internal parameters from an external device [21]. First, the motion between the images is computed pairwise, using 2D rotation and translation model. Then, the rotations about the Z-axis are canceled and narrow vertical strips from the centers of the images are pasted onto the mosaic image.

The strips are taken from the center of the image to minimize effects of misalignment. There are three reasons for that selection:

- The approximation of the motion generated by a panning camera as uniform is optimal at the center of the image.
- Lens distortion is minimal at the center of the images
- Alignment is usually better at the center than at the edges of the pictures.

This selection corresponds to the Voronoi tessellation [1]. Using the Voronoi tessellation for image cut-and-paste reduces 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 are radial, features that are perpendicular to the seam will be distorted equally on the seam and, therefore, will remain aligned regardless of lens distortion. The construction of the mosaic is very fast and has been demonstrated live on a PC [21]. Results are impressive in most cases and have the desired feature of manifold mosaicing: Each object in the mosaic appears in the same size as it appears in the video frames, avoiding any scaling and, therefore, avoiding distortions and loss of resolution. Mosaicing is done without the explicit assumption of pure rotation and without the need to project the images onto a cylinder before mosaicing. Fig. 14 and Fig. 13 show panoramic mosaic images created with an implementation of the manifold mosaicing on the PC [21].



Fig. 14. An example of panoramic imaging using manifold mosaicing with straight strips. The curved boundary is created by the unstabilized motion of the hand-held camera.

3.2 Forward Motion Using Pipe Projection

Forward camera motion used to be the classical case where traditional mosaicing fails. A theoretical camera model which handles the case of forward motion is a slit camera which scans the scene through a circular slit. This slit is symmetrical around the FOE and gives a wide FOV [32].

Manifold mosaicing can approximately simulate the “Circular slit” camera using a cylindrical manifold along the trajectory of the camera. We call the 3D projection on such a manifold “Pipe Projection” and its definition is described in this section.

The translation of the camera (and also zoom) induces radial optical flow which emerges from the FOE, except for the singular case of sideways translation in which the optical flow vectors are parallel. Cases of radial optical flow are much more complicated for mosaicing as the optical flow is not parallel and may depend on the structure of the scene. The pipe projection described in this section simplifies the mosaicing in these cases.

The internal parameters of the camera are assumed to be known throughout this presentation. It should be mentioned though that an error in the focal length results only in a scaling of the mosaic.

Given a sequence of images taken by a translating camera, we would like to transform the images such that the radial optical flow will turn into a parallel optical flow in the transformed representation. In order to do that, we project the 2D planar image onto a 3D cylindrical manifold, which we call a *pipe* (see Fig. 8). The axis of the pipe \hat{s} is chosen to pass through the optical center $O = (0, 0, 0)$ and through the FOE $S = (s_x, s_y, f_c)$, where f_c is the focal length, and, thus, $\hat{s} = S/|S|$. Each image point $P = (x, y, f_c)$ is projected onto its corresponding point Q on the pipe. The point Q is collinear with O and P and its distance from the pipe’s axis \hat{s} is R , where R is the radius of the pipe. The pipe projection is similar to the pipe-rectification proposed in [25] for stereo matching, but the projection from image to cylinder is different.

In the pipe representation of the image, the optical flow of each pixel Q is now parallel to the pipe’s axis \hat{s} . This does not solve the problem of motion parallax, but limits the distortion caused by parallax to be only along the trajectory of the camera. When the frame-rate is high and the depth differences are not significant, the pipe projection gives a good approximation of the mosaic generated from the “Circular slit” model. View interpolation can be used to reduce the parallax effects when the frame rate is not high enough.

The position in the pipe of a point Q is represented by k , the position along the axis \hat{s} , and α , the angle from \hat{d} . \hat{d} and \hat{r} are unit vectors chosen to form a Cartesian coordinate system together with \hat{s} . The 3D position of a point (k, α) on the pipe is $Q = (Q_x, Q_y, Q_z) = k\hat{s} + R\cos(\alpha)\hat{d} + R\sin(\alpha)\hat{r}$ and the corresponding pixel in the image plane for the point Q is $P = (x, y, f_c) = (f_c Q_x/Q_z, f_c Q_y/Q_z, f_c)$.

Pixels in the image whose original distance from the axis \hat{s} is less than R become magnified on the pipe, but, when projected back to the image, they restore their resolution. However, pixels with distance greater than R shrink on the

pipe, thus losing their original resolution. Note that selecting

$$R = \sqrt{f_c^2 + \left(\frac{w}{2}\right)^2 + \left(\frac{h}{2}\right)^2},$$

where w and h are the width and height of the image, ensures that no pixel will reduce its resolution at the projection, as the intersection of the pipe with the image plane never occurs within the image boundaries.

Most regions on the pipe are covered by projections from several images. For every point in the pipe, the projected values are taken from the image having the best resolution among all projected images. The best preserved resolution is around the intersection of the pipe with the image plane ($Q_z = f$) and the resolution decreases as $|Q_z - f_c|$ increases. This definition forms a strip which will be taken from that image having best resolution. An example is in Fig. 9.

The pipe representation is a natural generalization of the planar manifold and the cylindrical manifold: The planar manifold is analogous to the side of the pipe and the cylindrical manifold is analogous to a pipe with zero trajectory.

Cases like oblique view, forward motion, and zoom can be handled well using the pipe projection and give good results, while traditional mosaicing methods may fail in these cases. The pipe projection fits best, cases with close-to-linear translation. When the motion is more complicated, it is recommended to use the strip mechanism described in the next section. An example for the use of pipe projection is shown in Fig. 10.

3.3 Curved Strips for Forward Motion

Manifold mosaicing with curved strips can be used for the case of varying forward motion by implicitly simulating projection of strips onto a mosaicing manifold. The manifold is determined by the motion and can vary from one image to another, as demonstrated in Fig. 7. For simplicity, we show here the implementation of the scheme when the motion can be described by a parametric model.

3.3.1 Image Alignment

The image alignment used in our implementation for forward camera motion was a transformation describing a perspective projection of a plane (homography). While this is a transformation having only eight parameters, it gave reasonable overall results in our examples, as in Fig. 15. This transformation is described by the following equations:

$$\begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix} = \begin{pmatrix} \frac{a+bx_n+cy_n}{1+gx_n+hy_n} \\ \frac{d+ex_n+fy_n}{1+gx_n+hy_n} \end{pmatrix}. \quad (1)$$

Image alignment is performed between every two consecutive images using one of many methods (e.g., [2]). Since rotation about the optical axis does not introduce new information, we derotate the images by an approximation to such a rotation:² $\omega_z \approx \frac{e-c}{2}$. After the derotation, the homography is recomputed. The process of computing the

2. In a small rotation ω_z about the Z -axis, $x' = x - \omega_z y$ and $y' = y + \omega_z x$. Therefore, in (1), $e \approx \omega_z$ and $c \approx -\omega_z$.

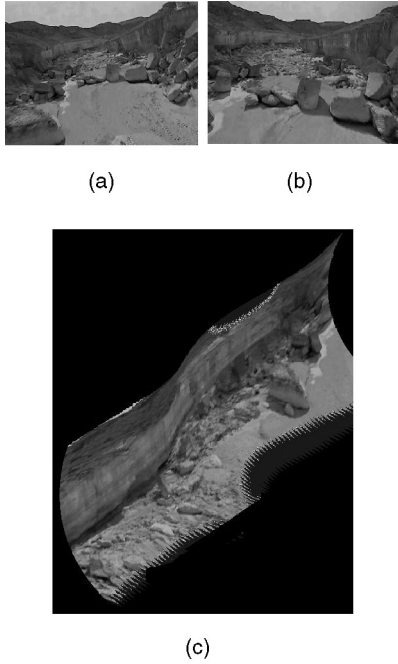


Fig. 15. Nonstable forward motion in a small canyon. (a) and (b) Two original frames. (c) Mosaic generated from curved strips.

homography followed by derotation is iterated several times until convergence.

3.3.2 Determining the Anchor

In each image, a curve which is maximally perpendicular to the optical flow is chosen as an anchor. Finding a curve that is exactly perpendicular to the optical flow is difficult in the case of a homography and sometimes an exact curve does not exist. Elliptic curves which are almost perpendicular to the image motion were used as an approximation. The ellipse is defined by using only six of the eight parameters (the affine parameters a, b, c, d, e, f) of the transformation. The elliptic curve is centered at the settle point of the transformation (the point where $(x_n, y_n) \approx (x_{n-1}, y_{n-1})$ or the point where this difference is the smallest). To find the settle point of the homography, an initial guess is used from the settle point of an affine transformation. From the equation:

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} a + b\tilde{x} + c\tilde{y} \\ d + e\tilde{x} + f\tilde{y} \end{pmatrix}, \quad (2)$$

the settle point can be derived:

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \frac{-dc + (f-1)a}{(b-1)(f-1) - ce} \\ \frac{ae - (b-1)d}{(b-1)(f-1) - ce} \end{pmatrix}. \quad (3)$$

This initial guess is used to start a minimization process to find the settle point (or the best approximation to a settle point) of the planar projective transformation. Note that the settle point can be outside the image itself, as in the case of sideways motion, where it is in infinity.

The radius of the ellipse depends on the self-characteristics of the sequence. Using a large radius will create a mosaic with higher resolution since the strip will be taken near the borders

of the image, where objects are closest to the camera. On the other hand, it is recommended not to use the most external borders where lens distortion is maximal.

3.3.3 Strip Selection

In order to determine the strip to be taken from Image I_n , the anchor of the the succeeding frame, I_{n+1} , should be considered.

Let \mathcal{A}_{n+1} be the transformation relating points $p_{n+1} = (x_{n+1}, y_{n+1})$ in Image I_{n+1} to the corresponding points $p_n = (x_n, y_n)$ in Image I_n . And, let the lines $\mathcal{F}_n(x_n, y_n) = 0$ and $\mathcal{F}_{n+1}(x_{n+1}, y_{n+1}) = 0$ be the anchors of the images I_n and I_{n+1} , respectively.

The strip that is taken from the image I_n is determined to be the region bounded between the two lines $\mathcal{F}_n(x_n, y_n) = 0$ and $\mathcal{F}'_{n+1}(x_n, y_n) = 0$ in I_n , where $\mathcal{F}'_{n+1}(x_n, y_n) = 0$ is the curve corresponding to the anchor $\mathcal{F}_{n+1}(x_{n+1}, y_{n+1}) = 0$ in I_n using the transformation \mathcal{A}_{n+1} (see Figs. 11a, 11b, and 11c).

This selection of the boundaries of the strip ensures that no information is missed or duplicated along the strip collection as the orthogonality to the optical flow is kept.

Mosaic construction is performed in a direction determined by the location of the settle point of the transformation and the center of the image. To create continuous mosaic images while avoiding accumulated distortions, the warping of the strips should depend only on an adjacent original frame, independent of the history of previous warpings. In our scheme, the back side of each strip is chosen to be an anchor and, thus, it is never changed. The front side of the strip is warped to match the back side of the next strip defined by \mathcal{F}'_{n+1} , as shown in Fig. 11d.

In the case of constant translation, the anchors for all images are identical and, thus, the optical flow becomes parallel to the direction in which the panoramic mosaic is constructed. Under the assumption of slowly varying motion, the original optical flow becomes close-to-parallel after the warping, avoiding significant scaling of the mosaic.

3.4 Rectified Mosaicing: A Tilted Camera

The mosaicing algorithm described in Section 3.1 flawlessly handles the following two cases:

- A panning camera, when the optical axis is perpendicular to the rotation axis. For example, when a camera is panning from left to right with a vertical rotation axis, its optical axis must be horizontal.
- A translating camera scanning a plane in the scene with the viewing direction perpendicular to the plane and the motion direction parallel to the plane.

When the camera motion and the viewing directions are different, e.g., when the camera is tilted, this mosaicing algorithm constructs a curled or otherwise distorted mosaic. In this section, an algorithm to handle this case is presented. The camera is assumed to be either panning and tilted or translating in a planar scene or moving horizontally, as in Section 3.1, but with a tilt. The image motion model used is a homography, which is assumed to be computed by one of many methods (e.g., [2]).

For simplicity of explanation, we assume that the optical flow is close to parallel and close to horizontal and,

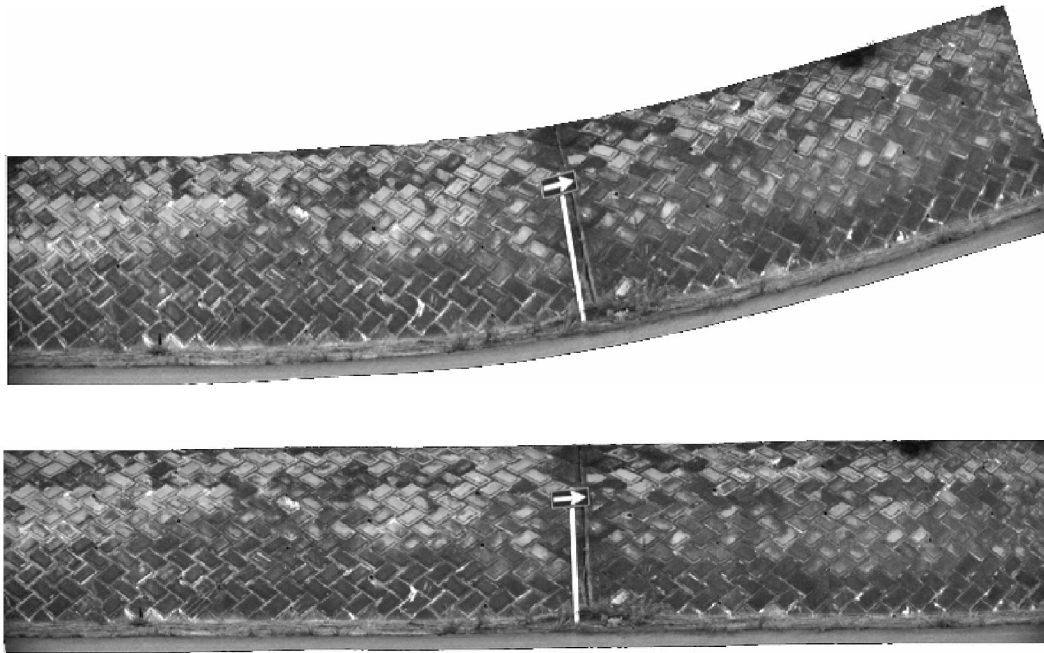


Fig. 16. A translating camera mosaicing a slanted wall. Regular mosaicing results in a curled image, while rectified mosaicing results in a straight mosaic.

therefore, we simulate a pushbroom camera having a vertical straight slit. Thus, the “anchor,” the feature in the strip that does not change with the warping of the strip, will also be a vertical straight line. The method can be easily adapted to nonhorizontal motion.

Consider the case of a camera translating to the right in front of a planar scene, with the viewing direction tilted up and to the right from a perpendicular view. When the plane is covered with equidistant horizontal and vertical lines, each image of this plane is similar to Fig. 12a. Parallel lines on the plane are not parallel in the image and a square in the plane is projected onto a general quadrangle in the image. The motion between consecutive images is similar to the one in Fig. 12a. A similar phenomenon occurs when the camera rotates about an arbitrary axis. Parallel vertical lines on the cylinder are not parallel in the image, as shown in Fig. 12b.

In order to use these images in the manifold mosaicing scheme, narrow strips should be taken from the images and

warped such that the optical flow becomes approximately parallel and of equal magnitude. Our method approximates a uniform horizontal optical flow by placing the anchors in parallel on the mosaic, with vertical offset which corresponds to the vertical motion of the camera. The narrow gaps between the anchors are filled using linear horizontal interpolation of the coordinates, assuring alignment along the seams. When the camera translates in a planar scene, the anchors are projections of parallel lines on the plane and, thus, parallel lines in the world are projected onto parallel lines in the mosaic image. When the camera pans, the anchors are projections of parallel lines on a cylinder centered in the camera rotation axis and, thus, parallel lines on the cylinder are projected on parallel lines in the mosaic.

The implementation details of rectified mosaicing are described in the Appendix. Comparisons of the results in the cases of a translating camera and a panning camera are shown in Figs. 16 and 17.

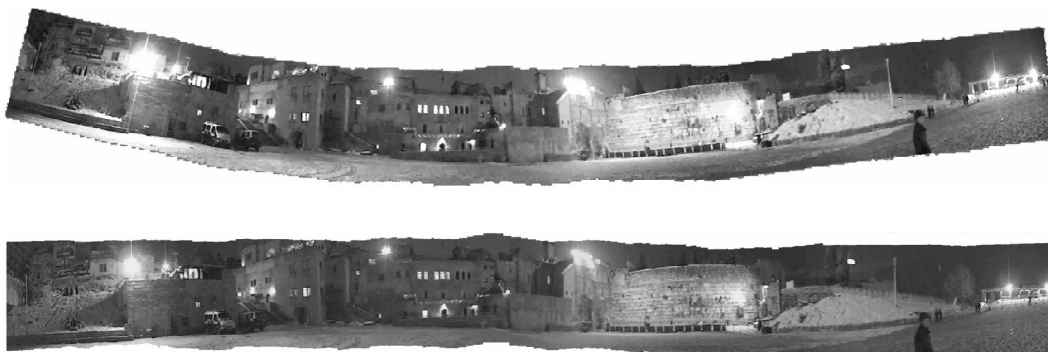


Fig. 17. Mosaicing from a panning camera which is slightly tilted upward. Regular mosaicing results in a curled image, while rectified mosaicing results in a straight mosaic.

$$\| \tilde{Q}_k - Q'_k \|_2 * \frac{h}{\| Q'_k - P'_k \|_2}$$

from the current strip, where h is the image height.

A.2 Symmetrical Strips

We assume similar imaging conditions as in Section 1.

Let C_k be the vertical line at the center of the image and let P_k, Q_k be its intersections with the top and bottom image borders.

We would like to choose a region which is approximately symmetrical around C_k , to reduce lens distortion. (This is the reason for choosing C_k as the anchor. In general, any other line can be used.) This region is illustrated in Fig. 19.

Given the homography H_{k-1} between Image I_k and Image I_{k-1} , let O_{k-1} be the center of image I_{k-1} and let d be the vertical offset between O_{k-1} and $H_{k-1}(O_{k-1})$. Let P'_k be a point shifted from P_k by d , and let Q'_k be a point vertically shifted from Q_k by d . A similar procedure is applied between images I_k and I_{k+1} , based on the homography H_k between the images.

Now, the homographies are used to map points P'_{k+1} and Q'_{k+1} from Image I_{k+1} and points P_{k-1} and Q_{k-1} from Image I_{k-1} , to Image I_k . Then, middle points are found: Let F_L be the homography mapping an arbitrary rectangle UVWX to the points $(H_{k-1}(P_{k-1}), P'_k, Q'_k, H_{k-1}(Q_{k-1}))$, respectively, and let F_R be the homography mapping UVWX to the points $(P_k, H_k^{-1}(P'_{k+1}), H_k^{-1}(Q'_{k+1}), Q_k)$, respectively. The region borders are defined by:

$$A_{11} = F_L\left(\frac{U+V}{2}\right), A_{21} = F_L\left(\frac{W+X}{2}\right),$$

$$A_{12} = F_R\left(\frac{U+V}{2}\right), A_{22} = F_R\left(\frac{W+X}{2}\right).$$

The polygonal region in the image is comprised of two quadrangles: the left quadrangle, with the corners at P'_k, Q'_k, A_{11} , and A_{21} , and the right quadrangle, with the corners at P_k, Q_k, A_{12} , and A_{22} . Each of these quadrangles is mapped to a rectangle in the mosaic. The left quadrangle is warped to a rectangle in the mosaic by some smooth (e.g., bilinear) interpolation of the coordinates of the corners like in the asymmetric case. A similar process is applied to the right part of the strip (rectangle) and the right part of the region.

The left part of the strip is placed at the same vertical offset as the right part of the previous strip and the right side of the strip is placed with vertical offset of d from the left part.

ACKNOWLEDGMENTS

This research was partially funded by DARPA through ARL Contract DAAL01-97-0101 and by the European ACTS project AC074 "Vanguard."

REFERENCES

- [1] F. Aurenhammer, "Voronoi Diagrams: A Survey of a Fundamental Geometric Data Structure," *ACM Computing Surveys*, vol. 23, no. 3, pp. 345-405, Sept. 1991.
- [2] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation," *Proc. European Conf. Computer Vision*, pp. 237-252, 1992.
- [3] P.J. Burt and E.H. Adelson, "A Multiresolution Spline with Application to Image Mosaics," *ACM Trans. Graphics*, vol. 2, no. 4, pp. 217-236, Oct. 1983.
- [4] P.J. Burt and P. Anandan, "Image Stabilization by Registration to a Reference Mosaic," *Proc. DARPA Image Understanding Workshop '94*, pp. 424-434, Nov. 1994.
- [5] S.E. Chen, "Quicktime VR—An Image-Based Approach to Virtual Environment Navigation," *Proc. ACM SIGGRAPH '95*, pp. 29-38, Aug. 1995.
- [6] S.E. Chen and L. Williams, "View Interpolation for Image Synthesis," *Proc. ACM SIGGRAPH '93*, pp. 279-288, Aug. 1993.
- [7] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P.J. Burt, "Real-Time Scene Stabilization and Mosaic Construction," *Proc. DARPA Image Understanding Workshop '94*, pp. 457-465, Nov. 1994.
- [8] R. Hartley and R. Gupta, "Linear Pushbroom Cameras," *Proc. Third European Conf. Computer Vision*, J.O. Eklundh, ed., pp. 555-566, May 1994.
- [9] R.I. Hartley, "Theory and Practice of Projective Rectification," *Int'l J. Computer Vision*, vol. 35, no. 2, pp. 1-16, Nov. 1999.
- [10] M. Irani, P. Anandan, and S. Hsu, "Mosaic Based Representations of Video Sequences and Their Applications," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 605-611, June 1995.
- [11] P. Jaillon and A. Montanvert, "Image Mosaicing Applied to Three-Dimensional Surfaces," *Proc. 12th Int'l Conf. Pattern Recognition*, pp. 253-257, Oct. 1994.
- [12] A. Krishnan and N. Ahuja, "Panoramic Image Acquisition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 379-384, June 1996.
- [13] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna, "Representation of Scenes from Collections of Images," *Proc. IEEE Workshop Representation of Visual Scenes*, pp. 10-17, June 1995.
- [14] C. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," *Proc. Computer Vision and Pattern Recognition '99*, vol. I, pp. 125-131, 1999.
- [15] S. Mann and R. Picard, "Virtual Bellows: Constructing High Quality Stills from Video," *Proc. First IEEE Int'l Conf. Image Processing*, vol. I, pp. 363-367, Nov. 1994.
- [16] L. McMillan and G. Bishop, "Plenoptic Modeling: An Image-Based Rendering System," *Proc. ACM SIGGRAPH '95*, pp. 39-46, Aug. 1995.
- [17] D.L. Milgram, "Computer Methods for Creating Photomosaics," *IEEE Trans. Computers*, vol. 24, pp. 1,113-1,119, 1975.
- [18] D.L. Milgram, "Adaptive Techniques for Photomosaicing," *IEEE Trans. Computers*, vol. 26, pp. 1,175-1,180, 1977.
- [19] S.K. Nayar, "Catadioptric Omnidirectional Cameras," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 482-488, June 1997.
- [20] S. Peleg, "Elimination of Seams from Photomosaics," *Computer Graphics and Image Processing*, vol. 16, pp. 90-94, May 1981.
- [21] S. Peleg and J. Herman, "Panoramic Mosaics by Manifold Projection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 338-343, June 1997.
- [22] P. Rademacher and Gary Bishop, "Multiple-Center-of-Projection Images," *Proc. ACM SIGGRAPH '98*, pp. 199-206, July 1998.
- [23] B. Rousso, S. Peleg, and I. Finci, "Generalized Panoramic Mosaics," *Proc. DARPA Image Understanding Workshop '97*, pp. 255-260, May 1997.
- [24] B. Rousso, S. Peleg, I. Finci, and A. Rav-Acha, "Universal Mosaicing Using Pipe Projection," *Proc. Sixth Int'l Conf. Computer Vision*, pp. 945-952, Jan. 1998.
- [25] S. Roy, J. Meunier, and I. Cox, "Cylindrical Rectification to Minimize Epipolar Distortion," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 393-399, June 1997.
- [26] H.S. Sawhney, S. Ayer, and M. Gorkani, "Model-Based 2D and 3D Dominant Motion Estimation for Mosaicing and Video Representation," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 583-590, June 1995.
- [27] S. Seitz and C. Dyer, "Physically Valid View Synthesis by Image Interpolation," *Proc. IEEE Workshop Representation of Visual Scenes*, pp. 18-95, June 1995.
- [28] H.Y. Shum and R. Szeliski, "Construction of Panoramic Image Mosaics with Global and Local Alignment," *Int'l J. Computer Vision*, vol. 36, no. 2, pp. 101-130, Feb. 2000.
- [29] R. Szeliski, "Video Mosaics for Virtual Environments," *IEEE Computer Graphics and Applications*, pp. 22-30, Mar. 1996.
- [30] D.N. Wood, A. Finkelstein, J.F. Hughes, C.E. Thayer, and D.H. Salesin, "Multiperspective Panoramas for Cel Animation," *Proc. ACM SIGGRAPH '97*, pp. 243-250, Aug. 1997.
- [31] J.Y. Zheng and S. Tsuji, "Panoramic Representation for Route Recognition by a Mobile Robot," *Int'l J. Computer Vision*, vol. 9, pp. 55-76, 1992.

- [32] J.Y. Zheng and S. Tsuji, "Generating Dynamic Projection Images for Scene Representation and Understanding," *Computer Vision and Image Understanding*, vol. 72, pp. 237-256, Dec. 1998.
- [33] A. Zomet, S. Peleg, and C. Arora, "Rectified Mosaicing: Mosaics without the Curl," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 459-465, June 2000.



Shmuel Peleg received the BSc degree in mathematics from The Hebrew University of Jerusalem, Israel, in 1976, and the MSc, and PhD degrees in computer science from the University of Maryland, College Park, in 1978 and 1979, respectively. He has been a member of the faculty at The Hebrew University of Jerusalem since 1980 and he has held visiting positions at the University of Maryland, New York University, and the Sarnoff Corporation. He

is a member of the IEEE and the IEEE Computer Society.

Benny Rousso received the PhD degree in computer science from The Hebrew University of Jerusalem, Israel, in 1998. He is now with Impulse Dynamics in Israel.



Alex Rav-Acha received the BSc degree in mathematics and computer science in 1997 from The Hebrew University of Jerusalem, Israel. Currently, he is an MSc student at The Hebrew University and his research area is computer vision and image processing.



Assaf Zomet received the BSc and MSc degrees in computer science from The University of Jerusalem, Israel, in 1997 and 1999, respectively. Currently, he is a PhD candidate at The Hebrew University. His research interests are in computer vision, with emphasis on motion analysis. For the past year, he has been a consultant to industry on computer vision problems such as mosaicing, super resolution, and motion analysis.