Int J Comput Vis (2009) 85: 307–315
DOI 10.1007/s11263-009-0264-7

Rectified Surface Mosaics

Robert E. Carroll · Steven M. Seitz

Received: 1 February 2008 / Accepted: 18 June 2009 / Published online: 30 June 2009

Abstract We approach mosaicing as a camera tracking problem within a known parameterized surface. From a video of a camera moving within a surface, we compute a mosaic representing the texture of that surface, flattened onto a planar image. Our approach works by defining a warp between images as a function of surface geometry and camera pose. Globally optimizing this warp to maximize alignment across all frames determines the camera trajectory, and the corresponding flattened mosaic image. In contrast to previous mosaicing methods which assume planar or distant scenes, or controlled camera motion, our approach enables mosaicing in cases where the camera moves unpredictably through proximal surfaces, such as in medical endoscopy applications.

Keywords Image registration and mosaicing · Pose estimation · Endoscopy

1 Introduction

Mosaics enable capturing the appearance of an entire scene in a single image. Many techniques have been proposed in the literature, based on a range of projection models including both perspective (Chen 1995; Szeliski and Shum 1997) and multi-perspective varieties (Wood et al. 1997; Agarwala et al. 2006; Seitz and Kim 2002). A common feature of almost all known projection models is that they introduce distortions, as certain scene characteristics (e.g., linearity, parallelism, length, angles, etc.) are not preserved in the mosaic. While some distortions may be tolerable or even desirable for casual visualization tasks, they can pose problems when precision or measurement is required.

One approach that avoids distortions is to define the projection based on a known reference plane in the scene. In this case, the images can be *rectified* to align with the physical coordinates on the reference plane, and stitched together. The result is a mosaic that removes the effects of perspective distortion and preserves lengths and angles for points on the reference plane (Szeliski 1996).

In this paper, based on our original work presented in Carroll and Seitz (2007), we generalize the approach of plane rectification to allow distortion-free rectified mosaics for any *developable surface* (e.g., boxes, cylinders, cones, generalized cylinders, etc.). More generally, our framework supports any kind of parametric surface, in which case the resulting mosaic image has the same parameterization as the surface itself. This generalization has a number of possible uses, but is particularly motivated by endoscopy applications, i.e., building mosaics of internal organs for medical diagnostics.

Our approach takes as input a video from a camera moving through a tube or other surface, and produces as output a planar image that represents the surface texture unwrapped onto a plane. The form of the surface must be known in advance, but the camera path may be unknown and unconstrained. Towards this end, we seek to align the images using a global objective function that is parameterized by surface geometry and camera pose in every image. We minimize this objective function using a global space-time version of Lucas-Kanade registration (Lucas and Kanade 1981), and

R.E. Carroll (✉)
University of California, Berkeley, CA, USA
e-mail: carroll@cs.berkeley.edu

S.M. Seitz
University of Washington, Seattle, WA, USA

composite strips of the aligned images into a mosaic. As a side effect, the approach solves for camera pose.

1.1 Related Work

Rousso et al. introduced "pipe projection" as a method that allows the mosaicing of video containing forward motion (Rousso et al. 1998). While this approach is very related to our goals, the "pipe" in their case is not meant to correspond to a physical surface in the scene, but is rather used as a manifold that allows radial optical flow to be transformed into parallel optical flow on the pipe's surface. Their approach is most effective when the scene is distant, or in the special case when the camera is moving along the center axis of a perfect cylinder. This pipe projection method is therefore not well suited for cameras moving freely through proximal physical pipes or other surfaces, as in the case of endoscopy video, and produces significant distortions in these cases.

In computing camera pose through a global optimization, our approach bears relation to work in computer vision and photogrammetry on structure from motion (SfM) and bundle adjustment (Hartley and Zisserman 2004). Indeed, SfM could in principle be used to compute the camera motion, as an alternative to the approach proposed in this paper. However, SfM methods require tracked features as input, which is particularly problematic for endoscopy video of anatomical surfaces, as such video tends to have few distinct features and large parallax between frames. Instead, we adopt an alignment approach that simultaneously solves for correspondence and camera pose in order to build the mosaic.

The problem of estimating camera pose in endoscopy imagery has been explored in the medical imaging community. Notably, several researchers (Bricault et al. 1998; Mori et al. 2000; Helferty et al. 2004) have demonstrated the ability to track a camera moving through the lung, computing six degree of freedom pose. This line of work builds on the use of *virtual endoscopy* to generate synthetic renderings of a CT model of the lung acquired off-line, which can be compared to the images captured online for the purpose of registration. This use of virtual images has the advantage of being able to correct for drift that would otherwise occur in frame-to-frame image registration. However, the accuracy of the alignment depends not just on the accuracy of the CT scan, but also the models for surface reflectance, shading, and lighting, and the ability to create renderings that approximate the true appearance of the scene. In contrast, our approach avoids the need to render a virtual model, and instead compensates for drift problems using a global multi-frame simultaneous alignment. A disadvantage of our approach is that the global optimization is not well-suited for real-time applications.

Vercauteren et al. (2006) present an endoscopic mosaicing technique where frame-to-frame motion can be approximated by rigid transformations. This approach is well suited for situations where the camera directly faces the surface being imaged and the surface is relatively flat. A number of other researchers (Reeff et al. 2006; Konen et al. 2007; Miranda-Luna et al. 2008) have presented similar techniques for endoscopic mosaicing, but they all assume each frame of the video sequence views a portion of the surface which is locally flat, or that the center of projection of the camera is mostly stationary. These assumptions allow the use of simple motion models (rigid, affine, homography, etc.) and a planar mosaicing surface, but they are not appropriate for applications we show, like that of a camera moving down a tubular structure. Seshamani et al. (2006) addressed the problem of mosaicing of tubular structures, but assume the camera motion is completely axial, allowing the warping and alignment procedures to be done independently. In contrast, our approach is aimed at applications where the camera moves arbitrarily, and we provide a more general formulation that works on other static surfaces.

Some authors have explored the related problems of unfolding 3D geometry (Carrascosa et al. 2006; Truong et al. 2006), computing triangle texture maps from images registered with CT scans (Rai and Higgins 2006), and unwrapping a single image using a circular parameterization (Warmath et al. 2005).

2 Overview

The input to our algorithm is a sequence of perspective views from a camera moving within a known type of surface. From this sequence we wish to output a mosaic which represents the texture of the surface. To do this we estimate the 6 DOF camera pose for each frame. With known pose, we can do an inverse projection of each frame onto the surface. The 2D mosaic is formed by flattening the surface onto a plane, which can be done without creating any distortions if the surface is developable.

Pose estimation is done by defining a warping function between neighboring video frames. This warp is based on an inverse projection of one image onto the mosaicing surface followed by a projection onto another image plane. The warp is a function of the pose parameters of both images and the surface parameters. We use this warp to define an intensity minimization between frames, using the framework of Lucas-Kanade alignment (Lucas and Kanade 1981; Baker and Matthews 2002; Szeliski and Shum 1997). We would like to compare each image to at least its two neighbors in the sequence, but this results in two (likely inconsistent) pose estimations for each frame. The series of duplicate pose estimations are not readily combined into a single camera path, so we instead generalize the registration algorithm into a global minimization across all frames.

To compute a consistent pose for each image, based on warps to multiple other images, we simultaneously solve the pose parameters of every frame. We use a generalization of Lucas-Kanade that handles multiple warps with interdependent parameters, which in our case are the camera pose parameters. Additionally, we show that the iterative update for the global optimization can be constructed from elements of the standard pairwise Lucas-Kanade method. Our global formulation is similar to the space-time tracking framework proposed by Agarwala et al. (2004b), who also used global optimization for tracking, but in the context of tracking 2D contours in a sequence of images.

Furthermore, we show that our framework can be extended to handle correspondences between input images and the projection surface. If the surface contains one or more known features that can be identified in the video, the pose estimations can be constrained to maintain the correspondence. This is useful in medical applications when the vanishing point of a cylindrical structure can be found or when known fiducials are present.

3 Surface Projection Warp

We model the video sequence as a pinhole camera moving freely within a static surface (Fig. 1). Under this model it is possible to warp an image taken at one location to an image taken somewhere else. We will use this warp to define an objective function parameterized by camera pose. The image warp we wish to solve for is modeled as the combination of an inverse perspective projection from one camera location onto the mosaicing surface, followed by the projection to another camera location. An arbitrary surface **S** in 3-space can be parameterized by two variables, (s,t), and the image plane is parameterized by (u,v).

The mapping from surface coordinates to image coordinates and vice-versa are given by

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{P}\left(\begin{bmatrix} s \\ t \end{bmatrix}, \mathbf{X}\right) \quad \text{and} \quad \begin{bmatrix} s \\ t \end{bmatrix} = \mathbf{P}^{-1}\left(\begin{bmatrix} u \\ v \end{bmatrix}, \mathbf{X}\right), \tag{1}$$

where $\mathbf{X} = (x, y, z, \alpha, \beta, \gamma)$ contains the six-degree-of-freedom camera pose.

The relationship between the 3D surface point $\mathbf{S}(s,t)$ and its projection onto the image plane at $\mathbf{u} = (u, v, f)$ is described by

$$\mathbf{S} = \mathbf{x} + \mathbf{Ru}c \tag{2}$$

where $\mathbf{x} = (x, y, z)$ is the position of the camera and $\mathbf{R} = \mathbf{R}_x(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_z(\gamma)$ is the rotation matrix representing the camera's orientation. The quantity **Ru** is the direction from the optical center to the 3D pixel location, adjusted to the coordinate system of the surface, and c is the scale factor corresponding to the intersection of this surface along **Ru**.

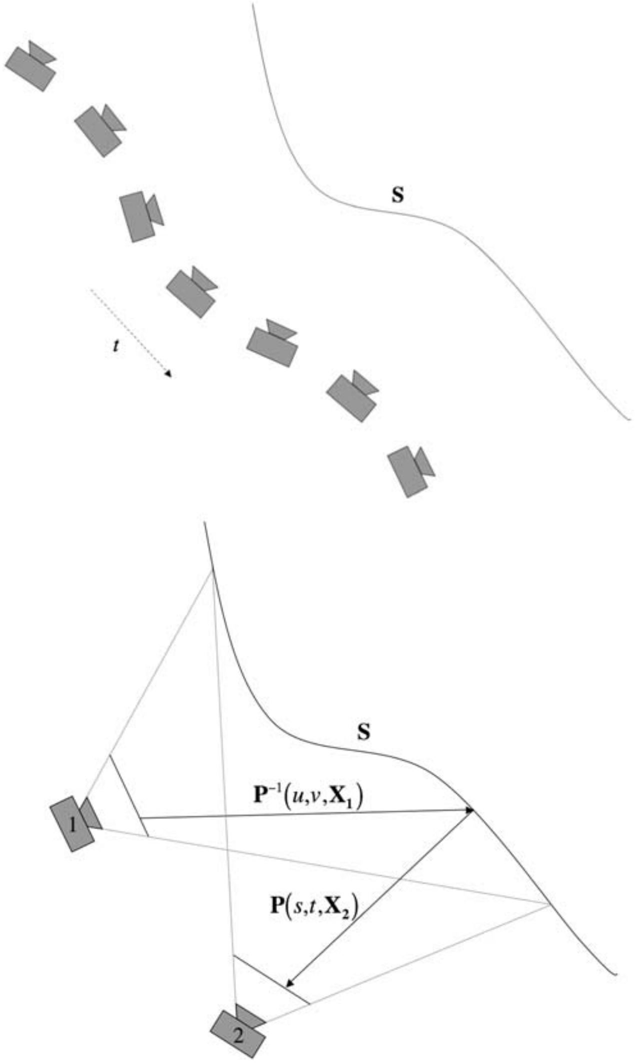
S
t
S
$\mathbf{P}^{-1}(u,v,\mathbf{X}_1)$
$\mathbf{P}(s,t,\mathbf{X}_2)$
1
2

Fig. 1 (*Top*) The motion model is that of a static surface and a freely moving cameral. (*Bottom*) The surface projection warp is a combination of inverse projection onto the surface from camera 1, and projection back to the image plane of camera 2

Solving (2) for **u** yields the following relation:

$$\mathbf{u} = \begin{bmatrix} u \\ v \\ f \end{bmatrix} = \mathbf{R}^{-1}(\mathbf{S} - \mathbf{x})/c. \tag{3}$$

We can easily find c, since the focal length f is known, and thus we have the forward projection.

The inverse projection is defined by intersecting a ray from the camera with the surface. We define the projection such that the intersection with the smallest positive c is used if there are multiple ray-surface intersections. See the Appendix for an example with a cylindrical surface.

Composing $\mathbf{P}^{-1}$ and $\mathbf{P}$ projects the first image through the surface into the coordinate system of the second image, giving the warp

$$\mathbf{W}(\mathbf{u}, \mathbf{X}_1, \mathbf{X}_2) = \mathbf{P}\left(\mathbf{P}^{-1}(\mathbf{u}, \mathbf{X}_1), \mathbf{X}_2\right). \tag{4}$$

4 Pairwise Pose Estimation

We wish to solve for a warp that minimizes the sum squared difference between two frames. That is, we wish to find $\mathbf{X}_1$ and $\mathbf{X}_2$ that minimizes the function

$$\sum_{\mathbf{u}} [I_1(\mathbf{u}) - I_2(\mathbf{W}(\mathbf{u}, \mathbf{X}_1, \mathbf{X}_2))]^2. \tag{5}$$

We use a Lucas-Kanade style forwards additive[1] approach as described by Baker and Matthews (2002). Given a current estimate of $\mathbf{X}_1$ and $\mathbf{X}_2$, we wish to find iterative updates $\Delta\mathbf{X}_1$ and $\Delta\mathbf{X}_2$ that reduces the error function

$$\sum_{\mathbf{u}} [I_1(\mathbf{u}) - I_2(\mathbf{W}(\mathbf{u}, \mathbf{X}_1 + \Delta\mathbf{X}_1, \mathbf{X}_2 + \Delta\mathbf{X}_2))]^2. \tag{6}$$

The closed form additive update for this equation is

$$\begin{bmatrix} \Delta\mathbf{X}_1 \\ \Delta\mathbf{X}_2 \end{bmatrix} = \mathbf{H}^{-1}\mathbf{b} \tag{7}$$

where $\mathbf{H}$ is the Hessian

$$\mathbf{H} = \sum_{\mathbf{u}} \left[\nabla I_2 \frac{\partial \mathbf{W}}{\partial(\mathbf{X}_1, \mathbf{X}_2)}\right]^T \left[\nabla I_2 \frac{\partial \mathbf{W}}{\partial(\mathbf{X}_1, \mathbf{X}_2)}\right], \tag{8}$$

and $\mathbf{b}$ is the residual (Szeliski and Shum 1997)

$$\mathbf{b} = -\sum_{\mathbf{u}} \left[\nabla I_2 \frac{\partial \mathbf{W}}{\partial(\mathbf{X}_1, \mathbf{X}_2)}\right]^T [I_1(\mathbf{u}) - I_2(\mathbf{W}(\mathbf{u}, \mathbf{X}_1, \mathbf{X}_2))]. \tag{9}$$

Our warping function is the combination of two projections, so the Jacobian of the warp can be expressed in terms of the Jacobians of the projections

$$\frac{\partial \mathbf{W}}{\partial(\mathbf{X}_1, \mathbf{X}_2)} = \left[\frac{\partial \mathbf{P}}{\partial \mathbf{s}} \frac{\partial \mathbf{P}^{-1}}{\partial \mathbf{X}_1} \quad \frac{\partial \mathbf{P}}{\partial \mathbf{X}_2}\right]. \tag{10}$$

In order to compute the Jacobian it is necessary that the surface be differentiable.

To improve convergence, we initialize the pose parameters based on the surface and the expected type of motion relative to that surface. For example, for a pipe we initialize the camera to be oriented axially, facing directly down the pipe. For a planar surface we initialize the camera to be facing the plane. All frames are given the same initial pose. For convenience, the world coordinates are chosen so that the initial pose is the zero vector. Depending on the surface there can be ambiguities in the warp, such as for a circular cylinder which is radially and axially symmetric. In these cases it may be desirable to partially constrain one of the frames, so as to prevent the minimization from searching among equivalent results. In the case of a cylinder, we constrain the first camera position to be at a particular depth and angle around the axis. The iterative update is run on a coarse to fine basis to improve convergence and performance.

5 Global Pose Estimation

The algorithm outlined in Sect. 4 estimates two camera poses given two input images. To register a full video, we could apply the pairwise registration to each pair of consecutive images. However, this would result in *two* pose estimates at each frame, and they would likely be inconsistent. To obtain a single consistent pose at each frame, we reformulate the pairwise optimizations into one global optimization that minimizes the error between successive frames simultaneously.

The error function we wish to minimize is a sum of the pairwise errors,

$$\sum_{i=1}^{n-1} \sum_{\mathbf{u}} (I_i(\mathbf{u}) - I_{i+1}(\mathbf{W}(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_{i+1})))^2. \tag{11}$$

As in Sect. 4, the Jacobian for the warp between frame i and $i+1$ is given by

$$\mathbf{J}_i = \frac{\partial \mathbf{W}(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_{i+1})}{\partial(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}, \tag{12}$$

the Hessian by

$$\mathbf{H} = \sum_{i=1}^{n-1} \sum_{\mathbf{u}} \left[\nabla I_{i+1}\mathbf{J}_i\right]^T \left[\nabla I_{i+1}\mathbf{J}_i\right], \tag{13}$$

and the residual by

$$\mathbf{b} = -\sum_{i=1}^{n-1} \sum_{\mathbf{u}} \left[\nabla I_{i+1}\mathbf{J}_i\right]^T \left[I_i(\mathbf{u}) - I_{i+1}(\mathbf{W}(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_{i+1}))\right]. \tag{14}$$

[1] The forwards additive algorithm is computationally more expensive than the alternatives described in Baker and Matthews (2002), but since the set of warps does not generally form a semi-group or group, the compositional algorithms are not applicable. Furthermore, the requirements for inverse additive approach are not satisfied.

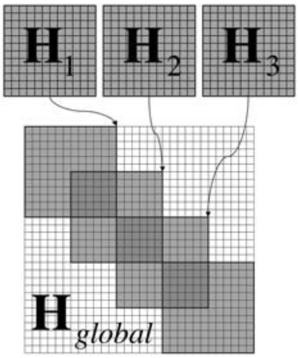

Fig. 2 The banded Hessian matrix used in the global optimization is constructed from Hessians of the pairwise registration. The overlapping regions are summed

The iterative update becomes

$$\begin{bmatrix} \Delta\mathbf{X}_1 \\ \vdots \\ \Delta\mathbf{X}_n \end{bmatrix} = \mathbf{H}^{-1}\mathbf{b}. \tag{15}$$

Note that the Jacobian is sparse, as the parameters for each frame depend only on previous and next frame in the sequence. The Hessian for the global optimization is a $6n \times 6n$ square matrix. However, since only consecutive frames are compared the Hessian is sparse and banded, allowing (15) to be solved efficiently. The global Hessian and residual can be constructed from their pairwise counterparts, as is illustrated in Fig. 2.

We have derived a global optimization where each frame is compared to the immediately previous and next frames. However, this method is easily extended to compare each frame to any number of neighbors, at the cost of additional computational complexity. If we alter the global error to be

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{i+k} \sum_{\mathbf{u}} \left(I_i(\mathbf{u}) - I_j\left(\mathbf{W}\left(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_j\right)\right)\right)^2 \tag{16}$$

the pose for each image will be determined by comparisons to k frames on each side. The value of k determines the width of the band in the global Hessian. The advantage of larger values of k is more reliable pose estimation that uses large and small motions for better local and global alignment. In the limit we compare each image to every other image and the objective function becomes

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{\mathbf{u}} \left(I_i(\mathbf{u}) - I_j\left(\mathbf{W}\left(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_j\right)\right)\right)^2, \tag{17}$$

in which case the Hessian is completely filled in.

6 Surface Feature Constraints

It may be possible to detect certain features in the input images that are known to correspond to a particular location on the surface. For example, for the case of endoscopy inside of cylindrical surfaces, it is straightforward to detect the cylinder's vanishing point as it appears as a dark spot in the video. We wish to constrain our optimization so that the location of the detected feature in each image coincides with the projection of its known surface location into that image. We can include these constraints in our optimization by adding the weighted squared distance of the two into the error function. The new global error is

$$\sum_{i=1}^{n-1} \sum_{\mathbf{u}} \left(I_i(\mathbf{u}) - I_{i+1}\left(\mathbf{W}\left(\mathbf{u}, \mathbf{X}_i, \mathbf{X}_{i+1}\right)\right)\right)^2 + \sum_{i=1}^{n} \sum_{f} w \left|\mathbf{P}(\mathbf{s}_f, \mathbf{X}_i) - \mathbf{u}_f\right|^2 \tag{18}$$

where $\mathbf{s}_f$ is the known location of a feature on the surface, $\mathbf{u}_f$ is the location of the feature in image i, and w is a weighting term. Since the constraints are independent of other images, we can treat them individually. The constraint for features in an image is given by

$$\sum_{f} \left|\mathbf{P}(\mathbf{s}_f, \mathbf{X}) - \mathbf{u}_f\right|^2 \tag{19}$$

We can proceed as in Sect. 4 to get the Hessian for the surface feature constraints

$$\mathbf{H} = \sum_{f} \frac{\partial\mathbf{P}}{\partial\mathbf{X}}^T \frac{\partial\mathbf{P}}{\partial\mathbf{X}}, \tag{20}$$

and the residual

$$\mathbf{b} = -\sum_{f} \frac{\partial\mathbf{P}}{\partial\mathbf{X}}^T \left[\mathbf{P}(\mathbf{s}_f, \mathbf{X}) - \mathbf{u}_f\right]. \tag{21}$$

The weighted Hessian and residual for each feature constraint can be added to their global counterparts in the same way as the pairwise pose estimation.

7 Mosaic Construction

With known camera pose for each input frame we are left with the problem of selecting patches of pixels from the appropriate input frames, and stitching them together seamlessly. We render the output mosaic using an inverse mapping, i.e., for each pixel in the output frame we sample from a corresponding point in the video sequence. However, each pixel in the mosaic image may be captured by multiple video frames. A simple approach is to traverse the video sequence, projecting onto previously unseen regions of the surface as they come into view. This is highly scene dependent, but

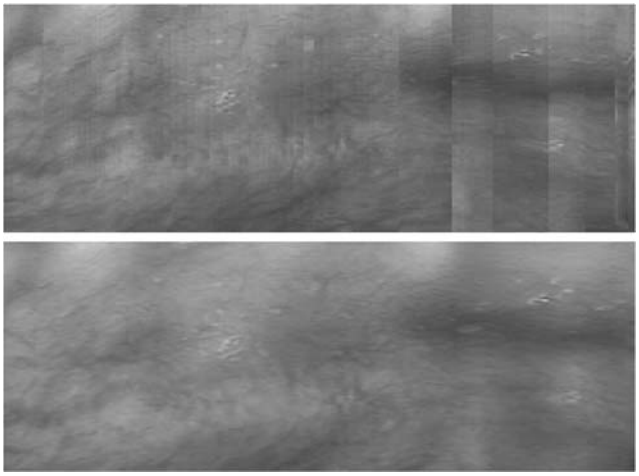

Fig. 3 Section of a mosaic created with and without gradient domain blending

we found that it works well for the case of a camera moving backwards out of a cylindrical surface (or a forward sequence traversed in reverse). We also disallow sampling from peripheries of the image frame since these regions tend to have more distortions and misregistrations.

Simply stitching together pixel intensities (or colors) will produce noticeable artifacts along seams due to exposure differences and any misregistations. To address this issue we use gradient domain blending (Agarwala et al. 2004a; Pérez et al. 2003; Fattal et al. 2002). In this technique, gradient fields from the different patches are stitched together instead of their intensity values. The output mosaic is produced by finding the image that most closely corresponds to this accumulated gradient field. Since there are no strong gradients on either side of the seems, the integrated image will have no visible seem (see Fig. 3).

8 Experimental Results

In this section we will demonstrate our algorithm on various data-sets to produce scene visualizations that could not have been captured with a single physical camera. Runtime for each of these experiments was 20–30 minutes on a 3.2 GHz Pentium IV.

8.1 World Map

This experiment involved moving a camera down a rigid 10 inch diameter, 5 ft. long tube lined with a rolled up world map. The camera was inserted into the tube on a plastic tray. The video was taken with a consumer camcorder and the scene was unevenly lit, as can be seen in the input images. Along with a limited depth of field, these issues make the registration challenging. The resulting mosaic (Fig. 4) is constructed from strips taken from 400 images. The bottom of the map is cut off since it is not visible when rolled up. The mixed exposure in the mosaic is caused by uneven lighting conditions. Despite the low quality input video, the recovered mosaic closely matches the reference image, demonstrating the algorithm's capability for metric accuracy.

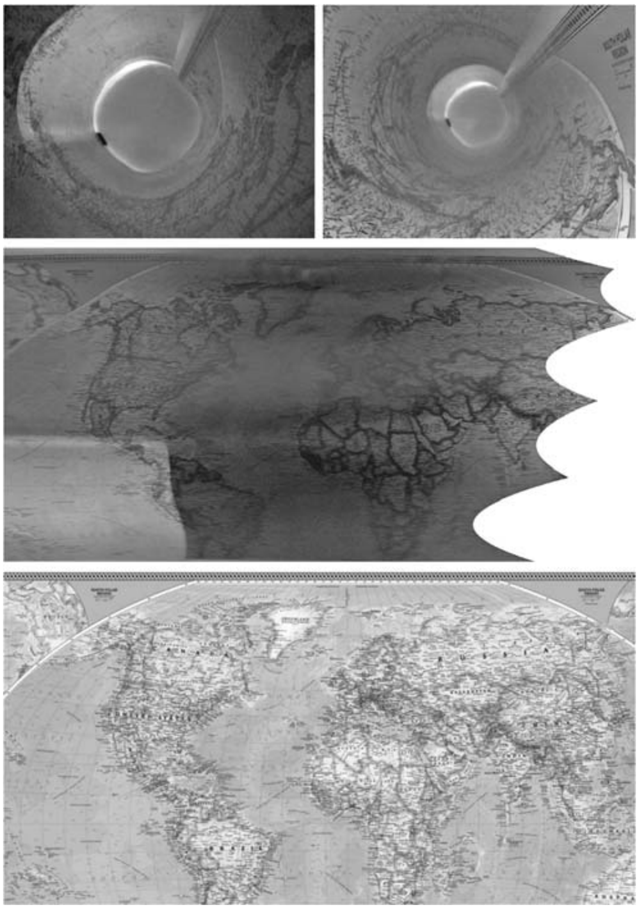

Fig. 4 Input frames from video of a rolled up world map, the resulting mosaic, and the ground truth map

8.2 Esophageal Endoscopy

This example (Fig. 5) was made from 220 frames of an esophageal endoscopy procedure on a sedated patient. This is one of the motivations for our project, to create a diagnostic tool for pre-cancer screenings. For this sequence a dark spot corresponding to cylinder's axial vanishing point was detected, and this constraint was used in the registration.

8.3 Model Esophagus

This mosaic (Fig. 6) was created as a proof of concept for the algorithm's use with an ultrathin Scanning Fiber Endoscope being developed at the University of Washington Human Photonics Lab (Seibel et al. 2006). The camera is unique in that it captures pixels sequentially from a spiraling optical fiber. The fiber scope was inserted into a model

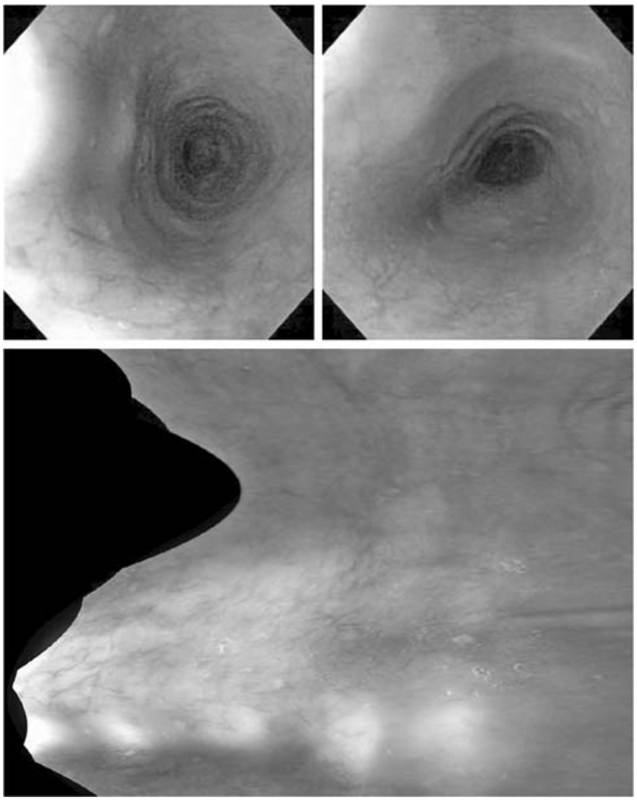

Fig. 5 Mosaic from esophageal endoscopy video using a cylindrical surface. The *dark spot* is constrained to correspond to the cylinder's vanishing point

esophagus to capture a 250 frame sequence. Given the low quality, grainy input video our registration algorithm produces an impressive result.

9 Conclusion

We have presented a technique for creating mosaics from a camera moving through a proximal surface with known geometry and demonstrated its applicability to endoscopic video of tubular structures. To our knowledge, in relation to other endoscopic mosaicing techniques, ours is the first to handle scenes that are not locally planar and where the camera moves arbitrarily. Our rectified mosaics use the known surface geometry as the mosaicing surface to reduce distortions and facilitate accurate measurement. We demonstrate our algorithm's accuracy by testing it on a cylindrical object that allows evaluation with respect to ground truth, and further demonstrate its robustness on real-world endoscopy video. The approach is shown to produce good results even with low quality input video.

There are a number of directions for future work in this area. Our current implementation is unoptimized and there is much potential for improving runtime by developing faster optimization techniques. Registration currently assumes brightness constancy, but this is a simplification in many cases since both the camera and light position can be moving. Specular highlights deviate from this assumption particularly and cause pixels to become saturated. Furthermore, our formulation can also be extended to include shape parameters to enable mosaicing of shapes where the geometry is not fully known. If allowed to vary at each time-step, the shape parameters could be used to model dynamic surfaces.

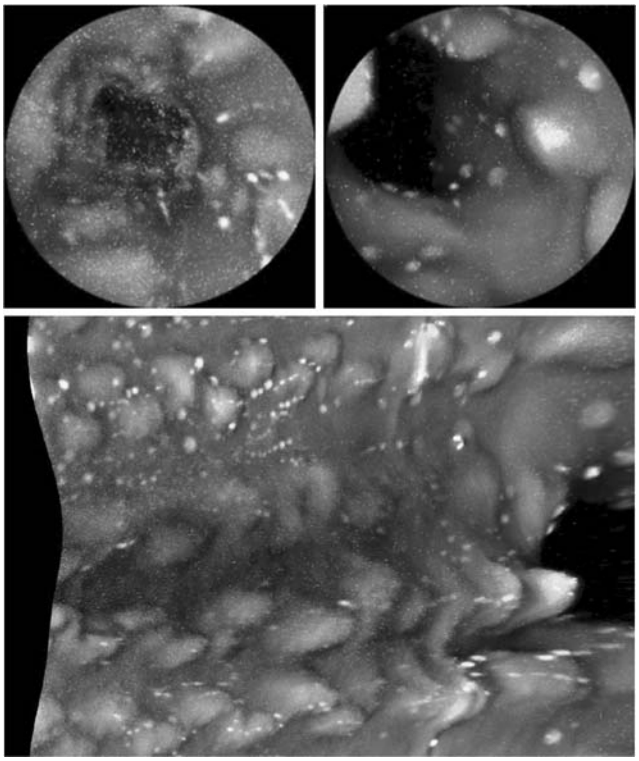

Fig. 6 Mosaic of a model esophagus imaged using a Scanning Fiber Endoscope (Seibel et al. 2006)

Acknowledgements The support of Pentax and NSF grant IIS-0413198 are gratefully acknowledged. We wish to thank Eric Seibel for suggesting this problem and for helping to guide our research.

Appendix: Cylindrical Pipe

The case of a cylindrical surface (Fig. 7) is derived by replacing the surface parameters (s, t) from Sect. 3 with the cylinder's parameters (k, θ), where k is depth along the

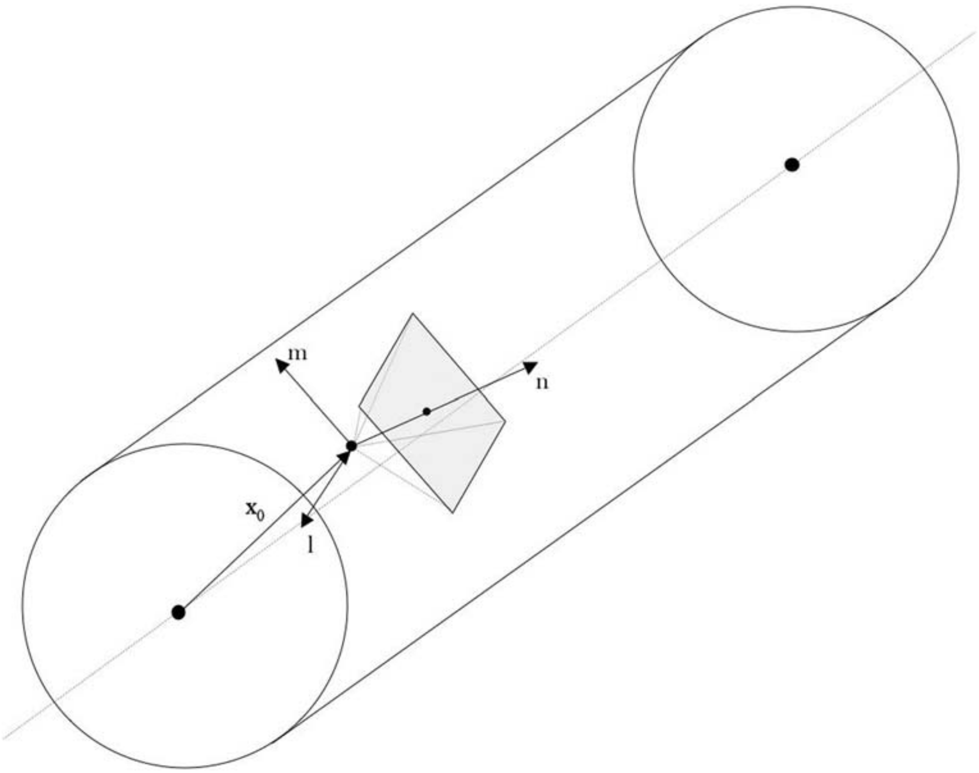

Fig. 7 The pipe projection: **l**, **m**, and **n** represent the image's basis vectors in scene coordinates. [**l m n**] = **R** represents the camera's orientation, and **x** is its position

pipe's axis and θ is the angle around the axis. Then

$$\mathbf{S}(k,\theta) = \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \\ k \end{bmatrix}$$

and relation between surface coordinate and image coordinate becomes

$$\begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \\ k \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{R} \begin{bmatrix} u \\ v \\ f \end{bmatrix} c.$$

From this equation the mapping from (k,θ) to (u,v) is straightforward, but going from (u,v) to (k,θ) is a bit more tricky. If u and v are known we can solve for c by noting

$$\begin{aligned} r^2 &= (r\cos(\theta))^2 + (r\sin(\theta))^2 \\ &= (c(\mathbf{Ru})_x + x)^2 + (c(\mathbf{Ru})_y + y)^2. \end{aligned}$$

This gives us a quadratic in c. With known c it follows that $\alpha = \arctan((y + (\mathbf{Ru})_x c)/(x + (\mathbf{Ru})_x c))$ and $k = z + (\mathbf{Ru})_z c$. Here it is assumed r is known, but this can be chosen arbitrarily as it is an ambiguity in the scene. Changing r will result in a stretching of the mosaic.

References

Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., & Cohen, M. (2004a). Interactive digital photomontage. *ACM Transactions on Graphics*, 294–302.

Agarwala, A., Hertzmann, A., Salesin, D. H., & Seitz, S. M. (2004b). Keyframe-based tracking for rotoscoping and animation. *ACM Transactions on Graphics*, *23*(3), 584–591.

Agarwala, A., Agrawala, M., Cohen, M., Salesin, D., & Szeliski, R. (2006). Photographing long scenes with multi-viewpoint panoramas. *ACM Transactions on Graphics*, *25*(3), 853–861.

Baker, S., & Matthews, I. (2002). *Lucas-Kanade 20 years on: A unifying framework: Part 1* (Tech. rep.). Robotics Institute, Carnegie Mellon University, Pittsburg, PA.

Bricault, I., Ferretti, G., & Cinquin, P. (1998). Registration of real and ct-derived virtual bronchoscopic images to assist transbronchial biopsy. *IEEE Transactions on Medical Imaging*, *17*(5), 703–714.

Carrascosa, P., Capunay, C., Lopez, E. M., Ulla, M., Castiglioni, R., & Carrascosa, J. (2006). Multidetector ct colonoscopy: evaluation of the perspective-filet view virtual colon dissection technique for the detection of elevated lesions. *Abdominal Imaging*.

Carroll, R. E., & Seitz, S. M. (2007). Rectified surface mosaics. In *Computer vision, 2007 ICCV 2007 IEEE 11th international conference* (pp. 1–8).

Chen, S. E. (1995). Quicktime vr: an image-based approach to virtual environment navigation. In *SIGGRAPH '95* (pp. 29–38).

Fattal, R., Lischinski, D., & Werman, M. (2002). Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, *21*(3), 249–256.

Hartley, R. I., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge: Cambridge University Press.

Helferty, J. P., Hoffman, E. A., McLennan, G., & Higgins, W. E. (2004). Ct-video registration accuracy for virtual guidance of bronchoscopy. In *SPIE* (Vol. 5369, pp. 150–164).

Konen, W., Breiderhoff, B., & Scholz, M. (2007). Real-time image mosaic for endoscopic video sequences. In *Bildverarbeitung für die Medizin* 2007 (pp. 298–302).

Lucas, B., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on artificial intelligence* (pp. 674–679).

Miranda-Luna, R., Daul, C., Blondel, W., Hernandez-Mier, Y., Wolf, D., & Guillemin, F. (2008). Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm. *IEEE Transactions on Biomedical Engineering*, *55*(2), 541–553.

Mori, K., Suenaga, Y., Toriwaki, J., Hasegawa, J., Kataa, K., Takabatake, H., & Natori, H. (2000). A method for tracking camera motion of real endoscope by using virtual endoscopy system. In *SPIE* (Vol. 3978, pp. 122–133).

Pérez, P., Gangnet, M., & Blake, A. (2003). Poisson image editing. *ACM Transactions on Graphics*, *22*(3), 313–318.

Rai, L., & Higgins, W. E. (2006). Image-based rendering method for mapping endoscopic video onto ct-based endoluminal views. In *SPIE* (Vol. 6141, pp. 1–12).

Reeff, M., Gerhard, F., Cattin, P., & Szkely, G. (2006). Mosaicing of endoscopic placenta images. In C. Hochberger & R. Liskowsky (Eds.), *Lecture notes in infomatics: Vol. P-93. Informatik 2006. Informatik für menschen* (pp. 467–474).

Rousso, B., Peleg, S., Finci, I., & Rav-Acha, A. (1998). Universal mosaicing using pipe projection. In *Intl. conf. on comp. vision* (pp. 945–952).

Seibel, E. J., Johnston, R. S., & Mellville, C. D. (2006). A full-color scanning fiber endoscope. In *Proc. SPIE, opt. fibers sens. for med. diagnost. treatment appl. VI* (Vol. 6083, pp. 9–16).

Seitz, S. M., & Kim, J. (2002). The space of all stereo images. *International Journal of Computer Vision*, *48*(1), 21–38.

Seshamani, S., Lau, W., & Hager, G. D. (2006). Real-time endoscopic mosaicking. In *MICCAI* (pp. 355–363).

Szeliski, R. (1996). Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, *16*(2), 22–30.

Szeliski, R., & Shum, H. Y. (1997). Creating full view panoramic image mosaics and environment maps. In: *SIGGRAPH '97* (pp. 251–258).

Truong, T., Kitasaka, T., Mori, K., & Suenaga, Y. (2006). Fast and accurate tract unfolding based on stable volumetric image deformation. In *SPIE* (Vol. 6143, pp. 412–423).

Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., & Ayache, N. (2006). Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis*, *10*(5), 673–692.

Warmath, J. R., Cao, Z., Bao, P., Herline, A. J., & Galloway, R. L. Jr. (2005). Semi-automatic staging system for rectal cancer using spatially oriented unwrapped endorectal ultrasound. In *SPIE* (Vol. 5744, pp. 425–434).

Wood, D. N., Finkelstein, A., Hughes, J. F., Thayer, C. E., & Salesin, D. H. (1997). Multiperspective panoramas for cel animation. In *SIGGRAPH '97* (pp. 243–250).