Real-time Endoscopic Mosaicking

Sharmishtaa Seshamani[1], William Lau[2], and Gregory Hager[1]

[1] Department of Computer Science, Johns Hopkins University, Baltimore, MD
{sharmi,hager}@cs.jhu.edu
[2] National Institutes of Health, Bethesda, MD **
william.lau@nih.gov

Abstract. With the advancement of minimally invasive techniques for surgical and diagnostic procedures, there is a growing need for the development of methods for improved visualization of internal body structures. Video mosaicking is one method for doing this. This approach provides a broader field of view of the scene by stitching together images in a video sequence. Of particular importance is the need for online processing to provide real-time feedback and visualization for image-guided surgery and diagnosis. We propose a method for online video mosaicking applied to endoscopic imagery, with examples in microscopic retinal imaging and catadioptric endometrial imaging.

1 Introduction

Endoscopy is an invaluable tool for surgical and diagnostic applications in pulmonary medicine, urology, orthopedic surgery and gynecology. It permits minimally invasive procedures, involving little or no injury to healthy organs and tissues. Current endoscopic technologies include fiberscopy, videoscopy, laparoscopy and wireless capsule endoscopy.

A drawback in these methods is the narrow field of view due to the size of most endoscopic imaging systems. As a result, individual images are often not very intuitive for evaluation. Automated mosaicking [1, 2] offers the opportunity to create an integrated picture or an environment map of a scene from a video sequence of endoscopic images. An essential first step in the process is the estimation of a registration estimate between captured images. One method of obtaining this estimate is the use of external optical tracking [3]. This however requires additional tracking equipment and a constant line of sight. A purely image based registration method is therefore an attractive alternative.

Current image registration methods generally apply to images related by planar homographies. Some examples are views of a plane from arbitrary camera positions and views of a general scene taken by a camera free only to pan, tilt and zoom [1]. Endoscopic images, however, are typically not related by planar homographies, due to the complexity of internal body scene structure and the impracticability of restricting camera motion. For example, bronchoscopy involves linear axial motion while imaging a tubular environment. It is therefore necessary to develop methods that, based on the imaging model and scene geometry, transform endoscopic images into a representation that is suitable for

** Disclaimer: The views and opinions of authors expressed herein do not necessarily state or reflect those of the NIH, DHHS, or the United States Government.

mosaicking. In particular, paracatadioptric imaging is potentially useful for endoscopy. A paracatadioptric system typically comprises of a parabolic mirror which reflects light onto a camera and thus provides a wider field of view which makes it a useful tool for endoscopy. However, image transformations imposed by the motion of a paracatadioptric imager are not linear, and further do not satisfy the requirement (for mosaicking) of forming a group, thus complicating the problem [4].

Methods for computing a registration estimate fall into two broad categories: direct methods which compute a transformation that optimizes some measure of photometric consistency over the entire image [5, 6], and feature based methods [7] which use a sparse set of corresponding image features to estimate the image-to-image mapping. Once a registration is computed, the construction of a mosaic entails resampling the images to a common coordinate system (given by the registration) so that they can be combined into a single image.

In most implemented systems, the entire mosaicking process is carried out offline, allowing the registration problem to be solved as a joint global optimization. However, this means that the quality of the mosaic and area it covers is difficult to determine until after the fact. A more intuitive approach is to develop an initial mosaic "online" as images are acquired. This provides the physician immediate and direct visual feedback as to the coverage and quality of the resulting mosaic.

In this paper, we describe methods for online image registration and mosaicking, and provide experimental results for retinal and endometrial imaging applications. In the next section, we first describe methods for performing traditional planar mosaicking and illustrate its application to retinal imaging. We then describe the modifications necessary to deal with paracatadioptric imaging of tubular structures and present results from that system.

2 Registration for Endoscopic Mosaicking

In the case of an endoscope viewing a planar or locally planar surface, the appropriate registration transformation is a homography [1]. However, in many cases the mapping between images of planar scenes can be described by mappings with fewer degrees of freedom which, consequently, can often be more reliably and rapidly estimated. In particular, affine mappings account for translation, rotation and scaling effects and are subgroups of a planar homography [1]. Although the affine transformation is, in general, necessary for mosaicking a large scene, it is often possible to make due with even simpler models, allowing for simpler computation and a more stable result. Our mosaicking system begins

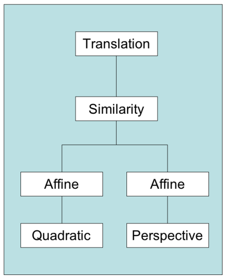

Fig. 1. Warping Model Hierarchy

with a simple translational model and moves through a hierarchy of models based on the scene structure (Fig. 1). The move to a more complex model is triggered when the registration error for the simpler model exceeds a fixed threshold. In the cases of locally planar (retinal) and cylindrical (endometrial) imaging which are presented in this paper, affine motion models generated small enough registration errors that did not exceed the threshold. Therefore quadratic and perspective models were not considered. The registration methodology we employ using this model is a direct technique. We denote the registration transformation as $D(p)$. In the case of an affine motion model, the transformation is linear which relates image coordinates as follows:

$$(u_1, u_2) = D(p)(x_1, x_2, 1)^T = f(x, p), \quad D(p) = \begin{pmatrix} 1 + p_1 & p_3 & p_5 \\ p_2 & 1 + p_4 & p_6 \end{pmatrix} \tag{1}$$

where $x = (x_1, x_2)$ is the pixel coordinate of a physical point on the first image, $u = (u_1, u_2)$ is the projection of the point in the second image, $p = (p_1, p_2, p_3, p_4, p_5, p_6)^T$ is the unknown parameter vector relating the two images and $f(x, p)$ is the transformation which is a function of x and p. [8, 5].

For online video mosaicking, motion between images is generally small, and the dominant motion is typically translation. As a result, it is common to compute an initial estimate of 2D translation by performing a brute-force search to maximize normalized cross-correlation between images. After this, a local continuous optimization method can be applied to compute the registration upto sub-pixel accuracy. Assuming brightness constancy, the goal of the registration algorithm is to minimize the sum of squared error between two images, I_0 and the image I_1 warped back onto the coordinate frame of I_0 ,with respect to the warping parameters p. For a general motion model with transformation function $f(x, p)$, this quantity is:

$$\sum_x [I(f(x, p + \Delta p)) - I_0(x)]^2 \tag{2}$$

This expression is linearized by a first order Taylor expansion on $I(f(x, p + \Delta p))$:

$$\sum_x [I(f(x, p)) + \nabla I \frac{\delta f}{\delta p} - I_0(x)]^2 \tag{3}$$

where ∇I is the image gradient vector and $\frac{\delta f}{\delta p}$ is the Jacobian of the transformation in (1). A linear closed-form solution can be obtained for the registration parameters p [6, 5]. In the interest of reducing computation time, a portion of the available image pixels can be chosen for optimization. In particular, since low magnitude image gradients have negligible effects on the solution, they can be eliminated to form an equivalent, smaller Jacobian matrix [8]. Once registration is computed to a subpixel level, the final step is to stitch warped images. The registration transformation between each pair of images j and k is defined as $D(p)_{j,k}$. Define K_i as the transformation of an image to the mosaic. For each new image, the transformation $K_{i+1} = K_i.D(p)^{-1}_{i,i+1}$. The new image is then weighted

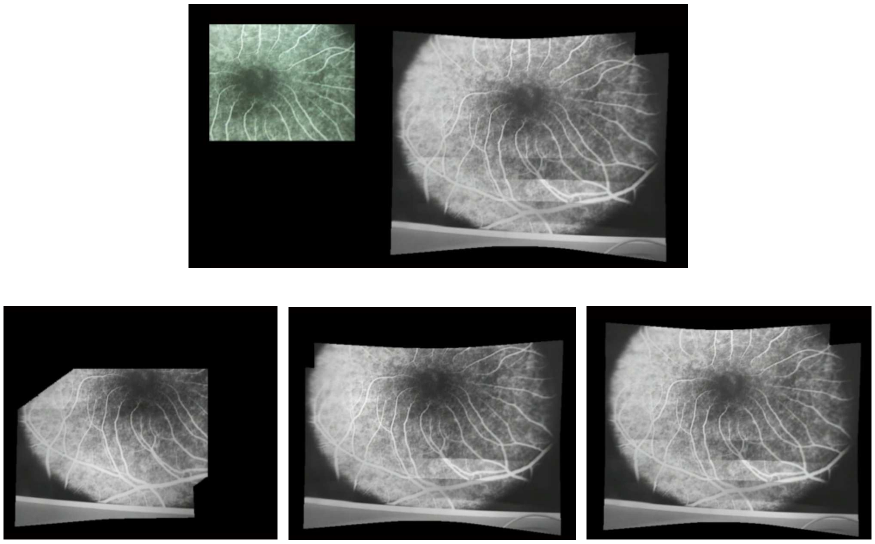

Fig. 2. An image mosaicking result for a retinal image sequence. The figure above shows an input image on the left and the resulting mosaic on the right; The 3 figures below show the mosaic in progressive stages. Note the change of field of view as more images are included

and projected onto the current mosaic with this estimated transformation. The method has been tested on simulated retinal sequences, and on endoscopic retinal images acquired with a GRIN lens endoscope (Insight Imaging, Inc.). Fig. 2 shows an example of the former as it is being constructed, in real-time, using the methods described above. This mosaicker was implemented in C and runs at 30 frames/sec.

3 Warping Models for Catadioptric Imaging

Recently, we have begun to investigate mosaicking for paracatadioptric sensors moving in tubular structures. As noted previously, catadioptric images cannot be registered by transformations that form a group. However, for motion that is largely axial through a tubular structure which provides a scene of roughly constant depth (as is the case in several types of diagnostic endoscopy), paracatadioptric images can be transformed into cylindrical representations. The transformations between these cylindrical images satisfy the condition of group membership. Therefore these representations can be registered to form a "tubular" mosaic. Fig. 5 shows a phantom setup we have developed for illustrating such an imaging system. A catadoptric imager is mounted on a linear stage and is positioned inside an empty clear cylindrical tube. Different textures are affixed to the clear tube and are imaged by moving the camera steadily in a straight line using the stage. This simulates the motion of an omnidirectional imager through different types of tissue.

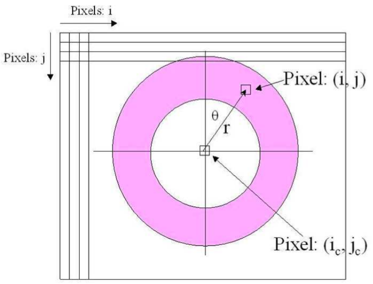

Fig. 3. Image Coordinates

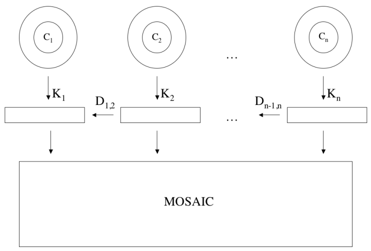

Fig. 4. Paracatadioptric Mosaicking

A raw image from this sensor appears as an annulus with image coordinates (i_c, j_c) as its center (Fig. 3). Points (i, j) from the raw image can be described in polar coordinates (r, θ) by: $r_{i,j} = ((i - i_c)^2 - (j - j_c)^2))^{1/2}$ and $\theta_{i,j} = \tan^{-1}\left(\frac{i_c - i}{j_c - j}\right)$. For the imaging geometry used in our system, the polar representation can now be related to the cylindrical surface with coordinates $(x_{i,j}, y_{i,j})$ as follows:

$$x_{i,j} = r_{i,j}\theta_{i,j}, \quad y_{i,j} = \left(\frac{-(r_{i,j} - c)}{a}\right)^{1/b} \tag{4}$$

where a, b, and c are power law coefficients characterizing the imaging geometry.

Once the images are in cylindrical coordinates, axial motion becomes translation in y, and rotation about the imager axis becomes translational motion in x (now viewed as wrapping at the edges of the image). In order to apply the framework of the previous section, the paracatadioptric mosaicking algorithm requires the extra step of changing from radial to cylindrical coordinates (Fig. 4). Let $T(i, j) \mapsto (x, y)$ denote this change of coordinates and let J_T denote the Jacobian matrix of this transformation. It follows then that

$$\nabla_{x,y} I = J_T \nabla_{i,j} I \tag{5}$$

relates image gradients in the raw ((i, j) coordinates) image to those in the cylindrical ((x, y) coordinates) image. The latter can now be used in (3) to solve for the cylindrical motion between images using the raw image as it is acquired.

4 Results

A preliminary validation of the method was conducted using the large scale paracatadioptric camera system described in Section 3. The purpose of this validation was to measure the accuracy of the reconstructed mosaic.

The first required step was calibration, in order to determine power fit coefficients a, b and c from (6). To solve for these, a uniform grid affixed to the imager was imaged and corners on the grid were automatically extracted (Fig. 5). Given this constant depth scene, the first grid circle could be chosen as the origin of the radial coordinate system allowing a solution for c. The distance between the first and second grid circle was taken as unit distance thus solving for

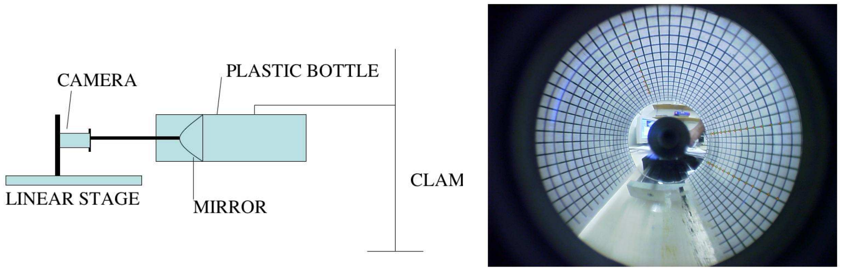

Fig. 5. Left:Large scale paracatadioptric simulator. Textures are affixed around the plastic bottle. The camera and mirror move in and out of the bottle during imaging. Everything else is stationary. Right:Calibration Grid

a. Finally, taking the log of both sides yielded: $b\log(y) = \log\left(\frac{r-c}{a}\right)$. This linear relationship easily solved for b using additional grid points. Once b was computed, a and c could be recomputed by standard linear regression. The values computed were as follows: a = 0.047, b = 1.51, c = 131.3. Tracking and mosaicking of uterine texture samples could then be implemented. Two types of data were collected. The first set of data was captured by imaging printed textures of normal and myomatous uterii affixed to the clear tube. For the second type of dataset, these printed textures were marked with pen mark fiducials placed at equal 1cm distances on a straight line. Image capture was then performed in the same manner as for the first type of dataset. The second set served as a validation tool. The uniformity of the reconstruction of the marks on the validation mosaics was used to determine the accuracy of the mosaicking algorithm. Fig 6(a) (top left) shows the original texture of a myomatous uterus which was pasted onto the tube. Fig 6(b) (top right) shows a reconstructed mosaic of the marked myoma texture. The distance between reconstructed marks ranged from 44 to 51 pixels (the image size is 551 X 661) and the variance of the distance between consecutive markers was 2.83 pixels. Given the ground truth distance of 1cm between markers, this gives a registration accuracy of 0.59mm. The ruler on the right side of the markers in Fig 6(b) provides a mark for every one half of the mean distance between reconstructed marks. Figs 6 (bottom two) show the mosaics generated from the dataset of the same texture without markers.

4.1 Experiments with Ex-Vivo Data

The above method was then applied to ex-vivo uterus images. Data was captured using a 4mm diameter hysteroscope with a paracatadioptric imager. The hysteroscope was moved at uniform 1mm intervals between consecutive images. Sample 1 is a set of 70 images of an endometrium with a myoma. Sample 2 is a set of 83 images of a normal endometrium. The resulting mosaics and motion plots are shown in Fig. 7. The computed motion in the y direction which corresponds to the dominant translational motion varies between 7 and 10 pixels between image pairs in the two samples. This is in accordance with the constant, purely translational motion of the imager. The computed motion in the x direction is very small as expected.

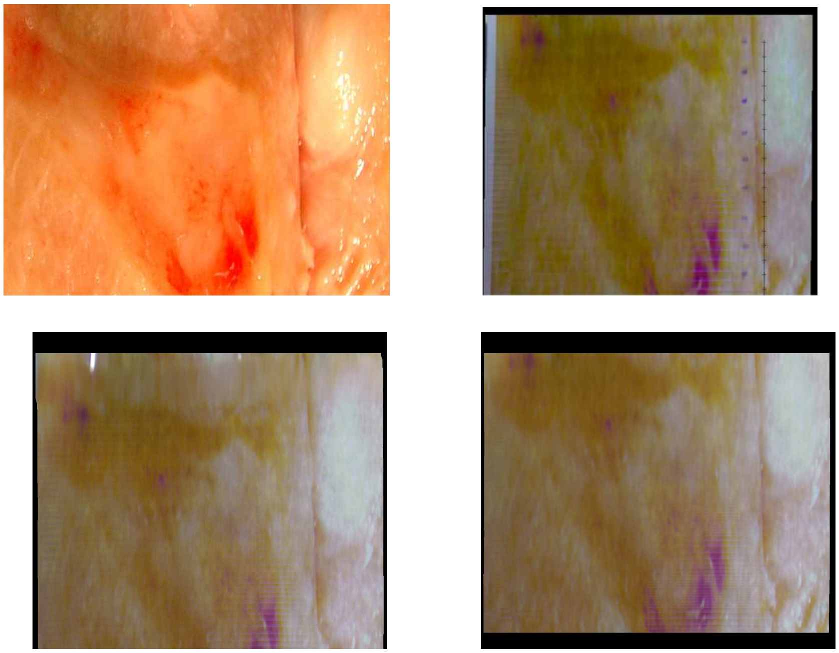

Fig. 6. Top Left: Original Myoma Texture Top Right: Mosaic of Myoma Texture with markers (ground truth) Bottom Left: Mosaic of Myoma Texture: Sample 1 Bottom Right: Mosaic of Myoma Texture: Sample 2

5 Discussion

We have presented a novel approach for online image tracking and mosaicking for the improved visualization of locally linear surfaces and closed tubular environments in the cases of microscopic and omnidirectional contact imaging. The ability to perform real-time processing is particularly important in the case of endoscopic mosaicking in order to provide real-time visualization. In general, image mosaicking is subject to some level of drift as there is no way to completely eliminate small incremental motion estimation errors. In order to reduce these registration errors a global block adjustment alignment can be applied to the whole sequence of images in offline processing, resulting in an optimally registered mosaic [2].

Future work will focus on extending the method to interpret lateral motion of the catadioptric imager, incorporation of tissue deformation and further ex-vivo experiments.

Acknowledgments

This research was supported by the National Institutes of Health and the CISST Engineering Research Center under the National Science Foundation Grant No. EEC9731748.

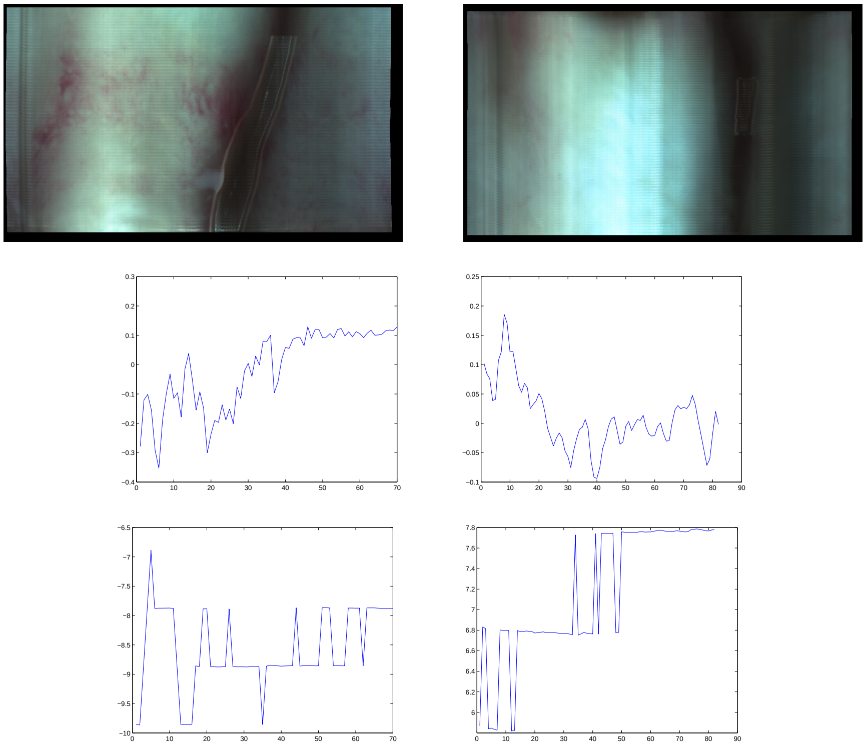

Fig. 7. Left and right columns show mosaics (first row), motion in x direction (second row) and motion in y direction (third row) of two different ex vivo samples.

References

1. Capel, D.: Image Mosaicing and Super-resolution. PhD thesis, Robotics Research Group, Department of Engineering Science, University of Oxford (2001)
2. Shum, H., Szeliski, R.: Construction of panoramic image mosaics with global and local alignment. IJCV **16** (2000) 63–84
3. D.Dey, Gobbi, D., an K.J.M Surry, P.S., Peters, T.: Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas. IEEE Transactions on Medical Imaging **21** (2002) 23–30
4. Dornaika, F., Elder, J.H.: Image registration for foveated omnidirectional sensing. In: ECCV (4). (2002) 606–620
5. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework: Part 1. Technical report, CMU-RI (2002)
6. Hager, G., Belhumeur, P.: Efficient region tracking with parametric models of geometry and illumination. IEEE PAMI **20** (1998) 1025–1039
7. Can, A., Stewart, C., Roysam, B., Tanenbaum, H.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE PAMI **24** (2002) 347–364
8. Lu, L., Dai, X., Hager, G.: Real time video mosaicing. cirl technical report. Technical report, Dept of Computer Science, Johns Hopkins University (2003)