Direct Global Adjustment Methods for Endoscopic Mosaicking

Sharmishtaa. Seshamani[a], Michael D. Smith[b], Jason J. Corso[b], Marcus O. Filipovich[b], Ananth Natarajan[b], Gregory.D Hager[a]

[a]Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[b]Infinite Biomedical Technologies, LLC, 3600 Clipper Mill Road, Baltimore, MD 21211, USA

ABSTRACT

Endoscopy is an invaluable tool for several surgical and diagnostic applications. It permits minimally invasive visualization of internal structures thus involving little or no injury to internal structures. This method of visualization however restricts the size of the imaging device and therefore compromises on the field of view captured in a single image. The problem of a narrow field of view can be solved by capturing video sequences and stitching them to generate a mosaic of the scene under consideration. Registration of images in the sequence is therefore a crucial step. Existing methods compute frame-to-frame registration estimates and use these to resample images in order to generate a mosaic. The complexity of the appearance of internal structures and accumulation of registration error in frame to frame estimates however can be large enough to cause a cumulative drift that can misrepresent the scene. These errors can be reduced by application of global adjustment schemes. In this paper, we present a set of techniques for overcoming this problem of drift for pixel based registration in order to achieve global consistency of mosaics. The algorithm uses the frame-to-frame estimate as an initialization and subsequently corrects these estimates by setting up a large scale optimization problem which simultaneously solves for all corrections of estimates. In addition we set up a graph and introduce loop closure constraints in order to ensure consistency of registration. We present our method and results in semi global and fully global graph based adjustment methods as well as validation of our results.

Keywords: Endoscopic Procedures, Data Integration for the Clinic/OR, Registration, Visualization

1. INTRODUCTION

In recent years, visual endoscopy has become an increasingly popular tool in diagnostic medicine. As a non-invasive visual procedure, it offers several advantages over blind biopsy methods such as increased comfort to the patient and convenience for the physician by providing opportunities for in-office procedures. Current techniques do however suffer from limitations in terms of visualization capabilities (due small fields of view and low video frame rates
) which make this type of diagnostic procedure inconvenient and long. Automated mosaicking offers an opportunity to tackle the problem of insufficient fields of view and resolution. Mosaicking procedures involve generation of an integrated picture or an environment map of a scene from a video sequence by stitching together multiple images of a scene.

The first step in the mosaicking procedure is registration, which involves computing a relationship between pixels in one image and another. Once pixel correspondence has been established, each image can be warped onto a common coordinate system to generate an environment map. Registration methods involve computation of a transformation that maps one image to another one. 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 [1][2], and feature based methods [8] [9], which use a sparse set of corresponding image features to estimate the image-to-image mapping. We [2] employ direct techniques on an endometrial mosaicking application whereas [10] [11] employ feature based registration techniques on a retinal and fetal mosaicking applications respectively. Once a registration is computed, the construction of a mosaic entails resampling the images to a common coordinate system so that they can be combined into a single image. The issues to consider in stitching are resampling and interpolation, blending, removal of artifacts due to motion (such as ghost effects) and generation of a seamless mosaic.

In most prior mosaicking work, registrations are computed by calculating relationships between consecutive frames. The use of frame to frame image registration methods is advantageous since they are simple and allow for rapid, real time processing of images for generation of a global environment map. However, long sequences of images accumulate misalignment errors when no further correction is applied to the registration. These errors could become large enough to misrepresent the scene under consideration. Furthermore, in medical imaging applications, the presence of deformable structures causes misalignment errors even within smaller sequences of images. Figure 1 shows some mosaics generated from images captured using a contact endoscopic system. Images are stitched together by computing frame to frame registration estimate between images. Figure 1A shows a mosaic of a regular grid generated using 80 frames and only frame to frame mosaicking. Figure 1B shows an image of an endometrial wall where the overlap clearly is incorrect.

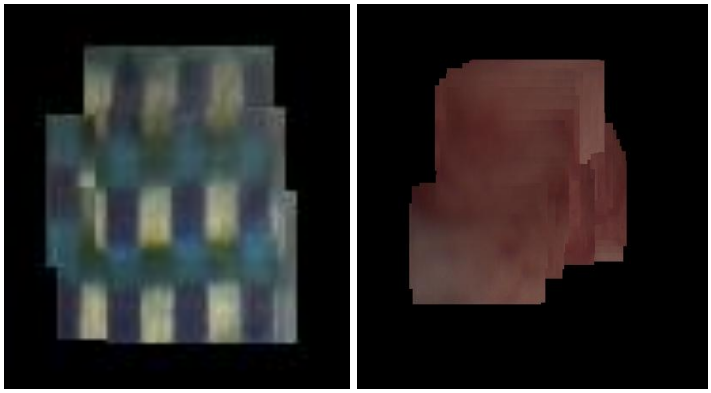

Figure 1 A: Mosaic of 80 image grid sequence using frame to frame registration,
B: Mosaic of 100 image grid sequence using frame to frame registration

There are generally two methods for enforcing global consistency: frame to mosaic methods and local to global methods. The first method involves estimation of an initial registration from frame to frame registration, followed by a local optimization of the frame to the mosaic at its current state. This method however is very sensitive to the appearance of the mosaic and the registration result can vary drastically with the choice of the blending method. The problem of global consistency can also be approached by enforcing photometric consistency (direct methods) or geometric consistency (feature based methods) between overlapping input frames. This is typically carried out by optimizing an objective function that describes the overall error of a set of registrations (bundle adjustment). Geometric methods [8][9] have been developed in the computer vision community where the objective function describes an error metric for reconstruction of the set of feature points in multiple views. In several cases of endoscopy however, image features are very weak and feature based registration methods cannot be applied. It is therefore useful to apply pixel based (direct) registration methods for such cases. Loewke et al [12] approach this problem by applying gaussian potentials to the direct case using the bundle adjustment framework. Vercauteren et al [6] apply a lie group formulation of the bundle adjustment problem and optimize on the cross correlation function.

The problem we address is the reduction of registration errors by performing direct global alignment in endoscopic video applications using a graph based method. In the next few sections, we describe our methods and results as well as the development of phantoms to validate our methods.

2. METHODS

2.1 Direct Registration of Two Images

Consider two images: I_0 the base image and I_1 which is to be registered and re-sampled back onto the coordinate frame of I_0. The registration estimate to be computed is the transformation that minimizes the SSD error criterion, with the assumption that pixel noise is gaussian [1]. The parameters of this transformation are denoted by p . For a general

motion model operating at image location x, the transformation function can be written as $f(x, p)$ and the SSD error for a parameter set p for all image locations is written as:

$$E(\partial p) = \sum_x \| I_1(f(x, p + \partial p)) - I_0(x) \|^2$$

Since the magnitude of components of ∂p is small, it is possible to apply a linearization of this expression using a first order Taylor expansion on $I(f(x, p + \partial p)$, ignoring higher order terms as

$$E(\partial p) = \sum_x \| I_1(f(x, p) + J \frac{\partial f}{\partial p} - I_0(x) \|^2$$

where J is the Jacobian matrix of the image I_1 with respect to image locations (spatial derivatives of Image I_1) and ∂f/ ∂p is the Jacobian of the registration transformation which is to be computed. A linear closed-form solution can therefore be obtained for the registration parameters p. In order to register images in a video sequence, this method can be used to compute registration estimates between every frame and the subsequent image in the video sequence. This technique is what is referred to as pair-wise registration. Once this registration is computed, the construction of a mosaic entails resampling all images to a common coordinate system so that they can be combined into a single image.

2.2 Global Mosaicking:

The proposed solution for global registration optimizes registration parameters by minimizing the total SSD error with respect to all pairs of images that overlap. This is a specific type of problem that falls under the general set of techniques known as bundle adjustment [7]. Bundle adjustment typically involves solving for a set of 3D structure and motion parameters that minimize the total reprojection error with respect to a set of feature points in many images. In our case, we formulate a solution to the global problem by extending pixel-based registration as follows. For a set of n Images in a video sequence $I_1 - I_n$:

$$E(\partial p) = \sum_{i,j=1:n} \sum_x \omega \| I_j(f(x, p_j) + J_j \frac{\partial f}{\partial p_j} - I_i(x) \|^2$$

where ω = 1 when I_i and I_j overlap and ω = 0 when there is no overlap. The parameter set p that is being solved for is the set of all registrations between every overlapping pair. The linear system that solves for the motion (registration) parameters in the direct optimization framework is:

$$J.p = -D_t$$

where J is the spatial derivative matrix, $D_t = I_1(f(x, p)) - I_0(x)$ is the time derivative vector and p is the motion vector to be solved for. In order to take advantage of different overlapping regions between images, we consider every overlapping pair of images, compute the region of overlap of this pair and then append the spatial and time derivatives of this region to the system above. The matrix J therefore is a block diagonal matrix with spatial derivatives between every image pair on the diagonal. (Pairs of images with a greater overlap have more rows in this matrix.)

Once this is setup, we can add additional constraints to the parameter set in order to ensure global consistency. This is carried out by setting up a graph system where each node of the graph represents an image and each edge represents the presence of overlap between two images. The parameter set being solved for is therefore the set of all edges of the graph. The metric of consistency defined to reduce drift error is the sum of motion parameters in a cycle. Since the elements of the parameter vector p that is being solved for include the motion parameters $p_{i,j}$, we can add the global consistency constraint to the system above as:

$$E(\partial p) = \sum_{i,j=1:n} \sum_{x} \omega \parallel I_j(f(x, p_j) + J_j \frac{\partial f}{\partial p_j} - I_i(x) \parallel^2 + \mu \sum_{k=1:q} \parallel c_k p \parallel^2$$

where q is the number of cycles and ck are the cycle bases. In order to add this to the linear system, we define C as the matrix of the sum of all cycle bases. The linear system that we solve for therefore becomes

$$(J + \mu C).p = -D_t$$

The added constraint enforces that all loops in the graph are closed (the sum of all registrations is 0), thus minimizing the accumulated error. In order to solve this sparse system efficiently in MATLAB, we use a conjugate gradients method.

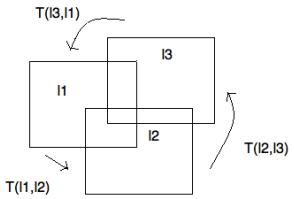

Figure 2: A system of 3 images where an overlap exists between every pair of images

In the case where there are 3 images and all images overlap with each other (Figure 2) , a fully connected graph representation is generated. A set of 6 registration parameters are to be solved for together. The constraint added therefore is:

[1 0 1 0 -1 0; 0,1,0,1,0,-1],

The algorithm can be summarized as follows:

- Generate all edges for the graph.
- Find all cycles in the graph and add this constraint to the system above.
- Solve the sparse system for registration parameters

Figure 3 shows mosaics generated from a sequence of 5 images. The images mosaicked here are the first image and the last (fifth) image. In this case, since all images overlap with each other, the graph is a fully connected graph. The mosaic on the left shows the first pass result where no global adjustment has been applied. The mosaic on the right shows the adjusted mosaic. The correction of drift error can be noted in this result.

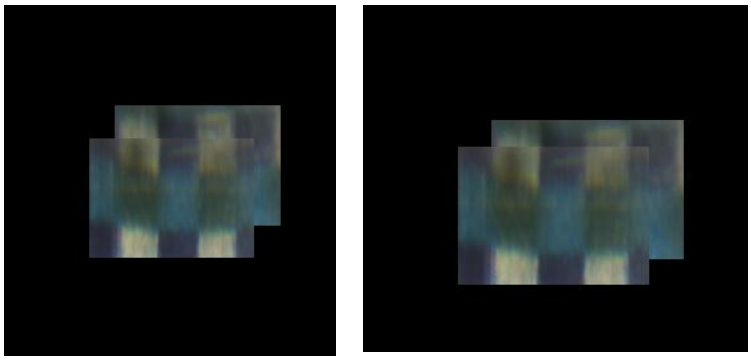

Figure 3: Images show mosaics of first and fifth image in the sequence. The result on the left is generated using only frame to frame registration. The result on the right shows the second pass result with graph based global adjustment

Semi global Registration: This algorithm was then extended to accommodate for a large video sequence where all images do not overlap with each other. Since the motion of the camera between image samples is small and the overlap between consecutive images is high, a small set of consecutive images can be considered to be fully overlapping (all images in the set overlap with every other image in the set). The Jacobian is the block diagonal matrix constructed using spatial derivatives between every image pair. For each of these sets, a fully connected graph structure can then be constructed. The advantage of this method is that it solves a much smaller system and therefore is more computationally efficient. The algorithm proceeds as follows:

- Suppose there are a total of N images to mosaic.
- Select the size of a subset n such that you can safely assume that any n consecutive images will have overlap.
- Using this set of n images, generate a fully connected graph structure where the nodes represent the images and the edges represent the registration between images. Now, use the current estimate of motion to warp images.
- Setup the Jacobian from derivatives and add graph constraints.
- Solve for all motion parameters using the optimization and update the registration parameters
- Repeat for all subsets

Fully Global Registration: Although this method is computationally efficient, it has the disadvantage of not being able to deal with the case where an image may overlap with another one that is not temporally close to it. In order incorporate all overlapping regions, we extend the algorithm to solve one large system which incorporates all overlapping image pairs. An initial frame-to-frame registration estimate is computed, and the overlap between all images is computed. The Jacobian is then constructed as the block diagonal matrix of spatial derivatives between every overlapping image pair. The algorithm is summarized as follows. Given a total of N images to mosaic:

- Find all image pairs that overlap
- Generate a graph structure where the nodes represent the images. An edge is added between two images only if there is overlap between the two.
- Setup the Jacobian from derivatives and add graph constraints.
- Solve for all motion parameters and update the registration parameters
- Repeat till convergence

3. EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental Setup and Data collection

Three types of data were used in our experiments. The first two are phantom datasets and the third is ex-vivo data of the endometrium. In order to obtain ground truth data in a reliable manner, the endoscope was attached to the steady hand robot [5], which has an accuracy of up to 10 microns. Various phantoms were positioned and the imager was moved in uniform steps using the robot. Figure 4 shows the system setup.

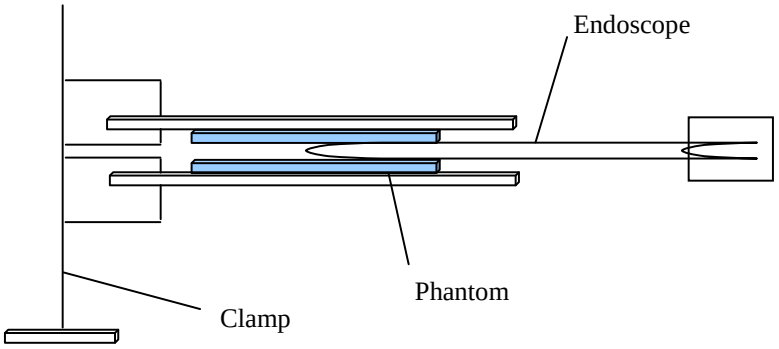

Figure 4: Experimental Setup

The first phantom used was a checkerboard pattern with squares of sides 1mm printed and pasted around a cylindrical pipette. The pipette allowed for a reduction in dewarping effects but did not allow for direct contact of the imaged checkerboard with the endoscope. Motions that could be applied using this phantom included insertion, retraction and rotation, as well as a combination of these which produced a spiral-like motion. The second type of phantom was a rubber phantom with similar visual feature properties as the endometrial wall. The phantom was prepared with a sample of pink silicone rubber that was painted with red silicone paint to generate features similar to the endometrial wall (Figure 5). The phantom was wrapped around the pipette during data collection. Motions that were applied using this phantom included insertion, retraction and rotation, as well as a combination of these which produced a spiral-like motion. The third set of data was collected on an ex-vivo uterus. Data was captured by a gynecologist using free hand motion. Motions that were applied include insertion, retraction, rotation and lateral motion.

Figure 5: Rubber Phantom

3.2 Validation

Validation of mosaics is a crucial aspect but this area has had very minimal work in medical imaging due to the difficulty of establishing a ground truth. In the endoscopic case it is usually not feasible to place fiducials on internal structures that need to be imaged to provide geometric ground truth. In addition, the field of view of the imagers is so small (usually around 4mm) that any type of marker that would have to be placed would have to be smaller than this in order to be useful. We propose a few methods for phantom based image mosaic validation in this proposal.Since the phantom does not have any regular features on its surface, it is not possible to extract higher level data for measurement of accuracy. However, since not all images are used to generate the mosaic, the regions of stitching can be compared with regions on corresponding regions of images not used for mosaicking to develop a measure of accuracy. The metric used for comparison is normalized SSD and the percentage error per pixel is reported. We refer to this metric as the SSD validation error. Further validation and accuracy measurements of results will be presented in the final paper.

3.3 Results

Semi Global Mosaicking: In this set of experiments, the semi global mosaicking described in section 2.2 was applied to generate the second pass mosaic. The images on the top show the frame to frame mosaic and the images on the bottom show the refined mosaic. In the grid case, a set of 5 consecutive images was used to refine the inter-frame registration. Figure 7 (top left) shows results on the 80 image grid sequence. The mean corner point error was 1.141 for the first pass mosaic and 1.1618 pixels for the adjusted mosaic. Figure 7(bottom left) shows results on phantom experiments. The

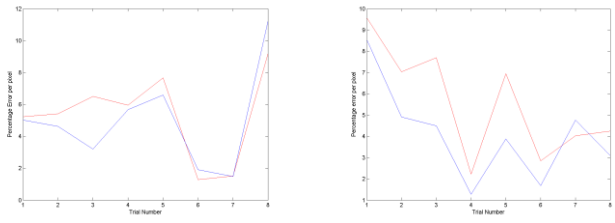

Figure 6:
Left - SSD Validation error for semi global phantom registration.
Right - SSD Validation error for fully global phantom registration.

input dataset was a sequence of 50 image. For validation, the method described in section 3.2 was applied. Figure 6 (Left) shows the result of 8 independent patches tested for percentage pixel error. The red and blue lines represent SSD validation error in the first and second pass mosaics respectively. Figure 7 (bottom right) shows the same algorithm applied to an endometrial image sequence, where the number of images in the subsequence forming the fully connected graph was 10.

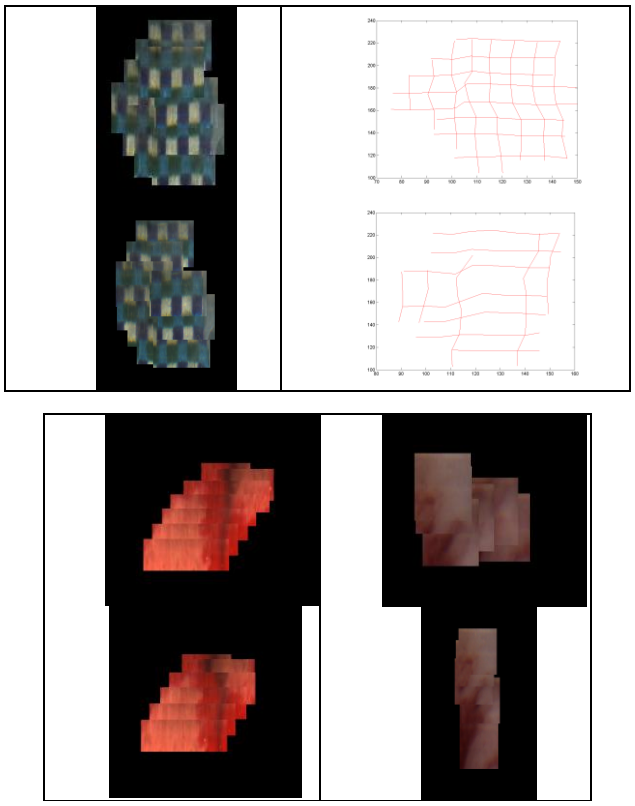

Figure 7: In each set the mosaic on the top is generated from frame to frame estimates and the mosaic on the bottom is generated from semi global adjustment. Top left: 80 grid images, Top right: Grid points plotted, Bottom left: Phantom mosaics, Bottom left: Ex-vivo endometrial mosaics

Fully Global Adjustment: The graph based fully global adjustment method described in section 2.2 was then applied. The mosaic on the top left in Figure 8 shows the result generated from the first pass frame to frame registration and the refined mosaic for the grid experiment. Validation was carried out by measuring the line fits of corners. The corner points on had an error of 1.141 pixels and the adjusted mosaic had an error of 0.6343 pixels. In the case of the phantom, the input dataset was the same used above. Eight independent feature rich regions were chosen on the mosaic for comparison with input images. A marked reduction of error can be observed between the first mosaic and the final mosaic. Figure 6 (Right) shows the result of 8 independent patches tested for percentage pixel error. The red and blue lines represent SSD validation error in the first and second pass mosaics respectively.

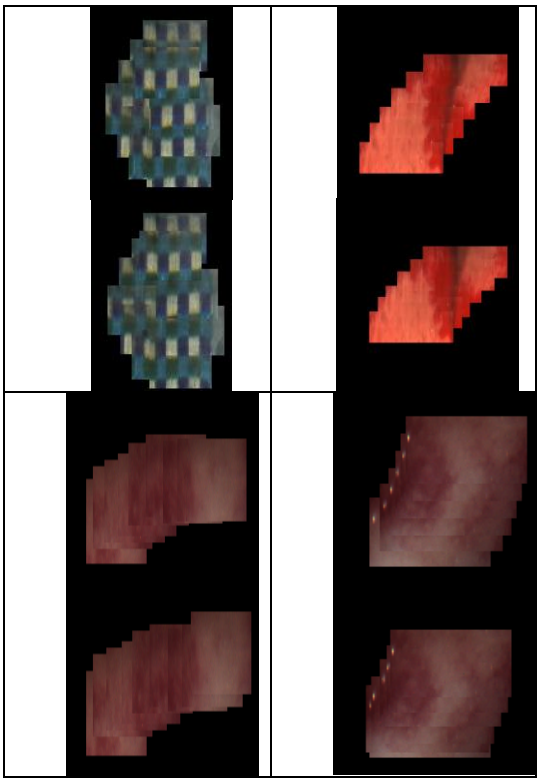

Figure 8: In each set the mosaic on the top is generated from frame to frame estimates and the mosaic on the bottom is generated from fully global adjustment. Top left: 80 grid images, Top right: Phantom Mosaics, Bottom: Two Ex-vivo endometrial mosaics

REFERENCES

[1] Hager, G., Belhumeur, P.: Efficient region tracking with parametric models of geometry and illumination. IEEE PAMI 20 (1998) 1025–1039

[2] Seshamani, S., Lau, W., Hager, G.D: Real-Time Endoscopic Mosaicking. MICCAI (1) 2006: 355-363

[3] Shum, H. and Szeliski, R. 1998. Construction and Refinement of Panoramic Mosaics with Global and Local Alignment. In Proceedings of the Sixth international Conference on Computer Vision (January 04 - 07, 1998). ICCV. IEEE Computer Society, Washington, DC, 953.

[4] Stewart, C., Tsai, C, Roysam, B: The Dual Bootstrap Iterative Closest Point Algorithm with Application to Retinal Image Registration. IEEE Trans. Med. Imaging 22(11): 1379-1394 (2003)

[5] Taylor, R.H., et al., A Steady-Hand Robotic System for Microsurgical Augmentation. International Journal of Robotics Research, 1999. 18(12).

[6] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. Medical Image Analysis 10 (2006)

[7] Triggs, B., McLauchlan P., Hartley R., Fitzgibbon A., (1999). "Bundle Adjustment — A Modern Synthesis". ICCV '99: Proceedings of the International Workshop on Vision Algorithms: 298-372, Springer-Verlag.

[8] Brown M. , Lowe D.G. : Recognizing Panormas ICCV 2003

[9] Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.:A feature based technique for joint linear estimation of high-order image-to mosaic transformations: mosaicing the curved human retina. IEEE PAMI 2004. 24 (2004), 412–419.

[10] Can, A., Stewart, C., Roysam, B., Tanenbaum, H.: The algorithm for registering pairs of images of curved human retina. IEEE PAMI 24 (2002) 347–364

[11] Reeff, M., Gerhard, F., Cattin, P., Székely, G.,: Mosaicing of Endoscopic Placenta Images[946] 2006, Informatik 2006. Informatik für Menschen, Band 1 (Lecture Notes in Informatics), vol. P-93

[12] Kevin E. Loewke, David B. Camarillo, Kenneth Salisbury, Sebastian Thrun: Deformable Image Mosaicing for Optical Biopsy. ICCV 2007: 1-8