Symposium - Original Research

# Global error minimization in image mosaicing using graph connectivity and its applications in microscopy

Parmeshwar Khurd, Leo Grady, Rafiou Oketokoun[1], Hari Sundar, Tejas Gajera[1], Summer Gibbs-Strauss[1], John V. Frangioni[1], Ali Kamen

Siemens Corporation, Corporate Research, Princeton, NJ, [1]Beth Israel Deaconess Medical Center, Boston, MA, USA

E-mail: *Parmeshwar Khurd - parmeshwar.khurd@siemens.com
*Corresponding author

## Abstract

Several applications such as multiprojector displays and microscopy require the mosaicing of images (tiles) acquired by a camera as it traverses an unknown trajectory in 3D space. A homography relates the image coordinates of a point in each tile to those of a reference tile provided the 3D scene is planar. Our approach in such applications is to first perform pairwise alignment of the tiles that have imaged common regions in order to recover a homography relating the tile pair. We then find the global set of homographies relating each individual tile to a reference tile such that the homographies relating all tile pairs are kept as consistent as possible. Using these global homographies, one can generate a mosaic of the entire scene. We derive a general analytical solution for the global homographies by representing the pair-wise homographies on a connectivity graph. Our solution can accommodate imprecise prior information regarding the global homographies whenever such information is available. We also derive equations for the special case of translation estimation of an $X$-$Y$ microscopy stage used in histology imaging and present examples of stitched microscopy slices of specimens obtained after radical prostatectomy or prostate biopsy. In addition, we demonstrate the superiority of our approach over tree-structured approaches for global error minimization.

**Key words:** Image mosaicing, Whole-slide scanning in digital pathology, Graph Connectivity

## 1. INTRODUCTION

Several applications such as multiprojector displays, underwater sea-floor mapping, microscopy, etc., require the mosaicing of images (tiles) acquired by a camera as it traverses an unknown trajectory in 3D space. A homography relates the image coordinates of a point in each tile to those of a reference tile provided the imaged 3D scene is planar.[1] One approach[2] in such applications is to first perform pair-wise alignment of the tiles that have imaged common features (points) in order to recover a homography relating the tile pair. We then find the global set of homographies relating each individual tile to a reference tile such that the homographies relating all tile pairs are kept as consistent as possible. Using these global homographies, one can generate a mosaic of the entire scene. The present paper is concerned with a method to compute the global homographies from the pair-wise homographies via global error minimization (GEM) under the assumption that the 3D object being imaged is planar. We first describe our algorithm in the general case in section 2 and then describe a specific

solution for microscopy in section 3. We demonstrate the performance of our algorithm on real microscopy data in section 4 and conclude with a discussion in section 5. Our contributions in this paper are threefold: (1) We provide a novel objective function and error minimization technique for global homography estimation in section 2 that can accommodate prior information regarding the global homographies whenever such information is available. (2) For the previously investigated case of global translation estimation in microscopy, we provide novel techniques to incorporate prior information. (3) We demonstrate the reduced error of the GEM techniques with respect to tree-structured approaches in section 4 (Results) and also consider application of the general solution in section 2 for the microscopy application.

Relationships between our algorithms and the existing literature are pointed out in sections 1.1 and 3. Applications of image mosaicing algorithms to histology are separately discussed in section 1.2, wherein we also discuss our primary applications, namely histopathological imaging after radical prostatectomy and prostate biopsy.

## 1.1. Related Algorithms

Several existing approaches have focused on fast incremental mosaicing by imposing a suitable tree-structure on the pairwise homography estimates.[3-6] These approaches do not try to minimize global mosaicing error by integrating information from all pairwise estimates. To our knowledge, the only techniques attempting to reduce a global measure of mosaicing error.[7-13] However, the global sum of Euclidean distance errors between common feature points across image pairs is considered[9,10] rather than our metric in (1), whereas Kang E,[7] uses the sum of intensity errors between such feature points as their metric. The homogeneous nature of microscopic tissue, e.g., lumen or stroma, can make it difficult to identify a sufficient number of common salient feature points for such approaches, rendering the metrics[7,9,10] harder to use. The focus is on reducing the cycle error from the pairwise 3D rotations and translations (they do not consider homographies) by back propagating it into the individual pairwise estimates.[11] Unlike us, they do not try to estimate homographies with respect to a single reference. The mosaic is built incrementally and translation and rotation errors in cycles are corrected whenever a cycle is detected.[12] However, this approach is not guaranteed to minimize a global error metric. The work[8,13] does minimize a global error metric and has inspired us to extend their approach in several ways. Note that they only address the case of translation estimation discussed in section 3 and we additionally present our generalization to global homography estimation. The work[13] uses the same objective function for GEM as in (5), but without the prior information. The work[8] uses the SVD to solve a linear system involving the edge-node incidence matrix discussed in section 3, but this is equivalent to the solution in Preibisch S *et al*.[13] Neither Emmenlauer M *et al*[8] nor Preibisch S *et al*[13] present our alternative technique with prior information or discuss the connections with the graph Laplacian and the related efficient linear system solver presented in section 3.

## 1.2. Applications of Image Mosaicing Algorithms in Histology and Prostate Histopathology

As discussed in section 1.1, the work[8,13] applied GEM techniques for 3D translation estimation in microscopy. That work was applied for 3D stitching of confocal microscopy images in nonpathological histology applications such as zebrafish embryo development,[8] cerebellum imaging of transgenic mice,[13] and central nervous system imaging in drosophila.[13] The work[6] mentioned above was concerned with 2D translation estimation in generic hematoxylin and eosin (H and E) stained histopathological imaging,whereas the work[5] was concerned with 2D rigid motion estimation for a variety of histological applications (both pathological and nonpathological) involving *in vivo* fibered microscopy videos.

The major clinical motivation for developing our novel mosaicing system for the assembly of whole-slide histopathological scans is the ability to utilize digital zoom in the context of feature recognition for the rapid automated prediagnosis of disease with the ability for pathologists to confirm suspicious areas rapidly. We have investigated two usage scenarios for our novel mosaicing algorithms. The first use case involves the online stitching of image tiles obtained by scanning a clinical histopathology slide using a microscope equipped with a motorized stage.[14] We have put our algorithm to this particular use as part of a special-purpose whole-slide scanner being developed in our laboratory for optical imaging in both the visible and near-infrared (NIR) spectra. However, the current computational implementation of our algorithm is suboptimal and hence slow, for the reasons described in section 4. Nevertheless, we are currently using it for prostate histopathology imaging, as described below, as part of a clinical research study. Our whole-slide scanner design falls under the progressive tile scanning category, and hence our image mosaicing algorithm can also be incorporated into other scanners that are similarly designed.[14] However, it is not useful in designs employing line-scanning cameras (that are currently limited to optical imaging in the visible spectrum), such as those manufactured by Aperio.[14] The second use case involves the offline correction of mosaics that have been generated by other techniques used in other tile-based whole-slide scanners, and that have stitching errors. The correction of stitching errors from whole-slide scanners is not a major requirement in most applications, but apart from being visually disturbing, such errors can

adversely affect the performance of automated image analysis algorithms in niche applications such as rare event counting, e.g., the counting of mitotic events in breast cancer histopathological imaging. However, we shall focus on the first use case in the remainder of this paper.

Our fully automated image mosaicing system employing the global energy minimization algorithm described in this paper has been used to yield qualitatively pleasing light microscopy mosaics of several $10\times$ images of radical prostatectomy specimens and $40\times$ images of prostate biopsy specimens. However, we do not address the specific problem of stitching quadrant prostatectomy specimens, which have specific problems associated with tearing and folding of tissue and require manual annotation of landmark points.[15] Our algorithm has only been used to stitch entire nonquadrant prostatectomy specimens acquired via a special microtome and processing techniques. However, it could potentially be used to stitch the individual quadrants in quadrant prostatectomy as well.

We also note that our prostate scanning procedures use an autofocusing technique, wherein the stage moves in the Z-direction for each tile to adjust focus while acquiring different tiles by moving in the X-Y plane. Therefore our stage motion differs slightly from the 2D translations considered.[6] and provided the motivation for developing our homography-based GEM algorithm in section 2. However, the comparison between the general homography model and the 2D translation model in section 4.3 shows that we can obtain very good stitching results solely with the 2D translation model and that the scaling issues can be ignored in our datasets.

## 2. MOSAICING VIA GLOBAL ERROR MINI-MIZATION USING GRAPH CONNECTIVITY: GENERAL CASE

Let us assume that N tiles were acquired by the camera. Mathematically, the global homography for tile $i$ can be expressed as $M_i = K(R_i - \frac{1}{d_i}p_i n^T)K^{-1}$ [1]. Here $K$ denotes the intrinsic camera matrix,[1] $R_i$ denotes the rotation of the camera axis of tile $i$ with respect to a reference tile, $p_i$ denotes the translation of tile $i$'s center with respect to a reference tile's center, $n$ and $d_i$ are the normal vector of the plane and the distance to the plane respectively. We can form a connectivity graph between all acquired images (tiles) by considering each tile, with its associated global homography, as a node, and by placing an edge between tiles that share a border to have some overlap. Let $N(i)$ denote the set of tiles that share any overlap with tile $i$. Let $M_{ij}$ denote the pairwise homography between tiles $i$ and $j$ found via pairwise alignment and let the weight $w_{ij}$ denote a measure of

confidence in this pairwise estimate. To determine the global homographies, we need to minimize the following objective:

$$\Phi(M_1, M_2, \cdots, M_N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)} w_{ij}\text{dist}(M_i, \tilde{M}_{ij} M_j)^2 \quad (1)$$

We take the squared distance between any two homography matrices $A$ and $B$ as $\sum_m\sum_n(A_{mn}-B_{mn})^2 = \text{Trace}(A-B)(A-B)^T$, i.e., the squared Frobenius norm of $A$-$B$, although alternative matrix distance measures such as those based upon the Log-Euclidean transform or the Baker-Campbell--Hausdorff formula may also be used for special cases involving positive-definite homographies.[5] Note that it is desirable to use to symmetric distance measures s.t. $\text{dist}(M_i, \tilde{M}_{ij} M_j)^2 = \text{dist}(M_j, \tilde{M}_{ji} M_i)^2$, a condition that is satisfied by the squared Frobenius norm.

In certain instances, a noisy estimate $M_i$ of the global homography for each tile may be available as prior information, e.g., from an electromagnetic tracking sensor. In such situations, we prefer to minimize the alternative objective function:

$$\Phi(M_1, M_2, \cdots, M_N) = \frac{\lambda}{2}\sum_{i=1}^{N}\text{dist}(M_i, \tilde{M}_i)^2 +$$

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)} w_{ij}\text{dist}(M_i, \tilde{M}_{ij} M_j)^2 \quad (2)$$

In Equation (2), the weight $\lambda$ controls the trade-off between the usages of the information in the pairwise estimates and the prior information. The above objective can be interpreted from a Bayesian viewpoint with the prior information corresponding to a uniform Gaussian prior probability. Strictly speaking, since matrices $\tilde{M}_i$ are derived from sensors during acquisition, they represent side information rather than prior knowledge. We are not restricted by symmetry considerations in choosing the distance measure for the prior information terms, but we continue to use the Frobenius norm.

The *unique* minimum of the objective function in Equation (2) can be found by setting its gradient to zero. The gradient of the objective function in Equation (2) with respect to the homography matrix $M_i$ is

$$\nabla_{M_i}\Phi(M_1, \cdots, M_N) = \frac{1}{2}\nabla_{M_i}[\text{Fro}(M_i - \tilde{M}_i)]^2 +$$

$$\frac{1}{2}\sum_{j\in\mathcal{N}(i)} w_{ij}\nabla_{M_i}[\text{Fro}(M_i - \tilde{M}_{ij} M_j)]^2. \quad (3)$$

Note that $\nabla_{M_i}[\text{Fro}(M_i - \tilde{M}_{ij} M_j)]^2 = \nabla_{M_i}\text{Trace}((M_i - \tilde{M}_{ij} M_j)(M_i - \tilde{M}_{ij} M_j)^T) = 2(M_i - \tilde{M}_{ij} M_j)$

Setting the gradient in Equation (3) to zero, we get the

following set of $N$ simultaneous matrix equations:

$$\left(\sum_{j \in \mathcal{N}(i)} w_{ij} + \lambda\right)M_i = \lambda \tilde{M}_i + \sum_{j \in \mathcal{N}(i)} w_{ij} \tilde{M}_{ij} M_j \qquad (4)$$

The equations in (4) imply that each global homography matrix can be obtained by updating the prior estimate with a linear combination (with matrix-valued coefficients involving the unknown global homography matrices) of the neighboring global homography matrices. Each matrix equation in (4) is equivalent to a set of nine scalar equations for the individual matrix elements. We can concatenate the $(a,b)^{th}$ elements $M_i$ of the set of the $3 \times 3$ matrices $M_i$ into a vector **x** and then compute this vector by solving the linear system $A\mathbf{x} = \mathbf{y}$, where $A$ is a $9N \times 9N$ matrix . The exact form of $A$ and **y** is given in Appendix A.

In case prior information is not available, we assume that a reference homography $M_N$ is known and solve for the remaining $N\text{-}1$ homographies in (4) with $\lambda = 0$. Note that knowing a single-reference homography may not guarantee a unique solution to these matrix equations, especially when the connectivity graph is disconnected. We shall revisit this issue in the case of translation estimation in section 3. For the case of a single-reference homography $M_N$, the matrix $A$ in the linear system now becomes a $9(N - 1) \times 9(N - 1)$ matrix and details are again provided in Appendix A.

Note that the above derivation assumes a general $3 \times 3$ homography matrix without any constraints on its form. However, general homography matrices have at most 8 degrees of freedom[1,16] (up to scale), with 3 degrees of freedom each for the rotations and translations, and additional degrees of freedom arising from the variations in the intrinsic camera matrix during the scanning process, e.g., due to a changing focus. We can deal with this 8-degree homography problem by setting $M_i=1$ for all homographies[17] and drop the corresponding rows and columns of $A$ to get an updated $8(N\text{-}1) \times 8(N\text{-}1)$ matrix for $A$. The vector **y** would also get updated with the appropriate terms transferred from the original LHS. Note that since $M_i$ is nonzero for all tiles in our microscopy application (since the origin does not project to a line at infinity), we do not need to use the minimum norm constraints discussed.[1] In case the intrinsic matrix is fixed, but both translations and rotations are allowed, the homography matrix has the form $M_i = (R_i - t_i n^T)$ and thus has only 6 degrees of freedom (the scalar factor $d_i$ has been absorbed into $t_i$). In this case, using the method.[12] $R_i$ and $t_i$ can be recovered from the $M_i$ found by the $8(N\text{-}1) \times 8(N\text{-}1)$ linear system solver. For other special cases of homographies such as pure rotations (which would occur, e.g., when the camera is rotated from a single-vantage point to cover a large wall painting), we can impose these constraints on the solution found via the $8(N\text{-}1) \times 8(N\text{-}1)$ linear system solver using the SVD-based rotation extraction approach described.[16] However, pure rotations are not likely to happen in our microscopy applications. For the case of 2D translations considered in section 3, we obtain only two independent equations per matrix equation and do not need to impose any additional constraints on the resulting linear system solution, as demonstrated by the first-principles derivation in that section. In that section, we also describe how to deal with the case of multiple connected components in the graph connectivity (whenever they arise).

## 3. MICROSCOPY *X-Y* STAGE MOTION ESTIMATION AND IMAGE MOSAICING

We now describe how our image mosaicing method from section 2 can be used in histological microscopic imaging. Since the motion reported by the slide-scanning microscopy *X-Y* stage in histopathology imaging is typically unreliable, our image-based mosaicing method is particularly useful in this application. Assuming that the microscope moves in a plane parallel to the planar specimen (we may assume this plane to be normal to the Z-axis) and $K = I$, the homography $M_i$ thus reduces to $M_i = (I - \frac{1}{d}p_i e_z^T)$. Then $M_i - \tilde{M}_{ij}M_j = -\frac{1}{d}(p_i - p_j - \tilde{p}_{ij})e_z^T$, $M_i - \tilde{M}_i = -\frac{1}{d}(p_i - \tilde{p}_i)e_z^T$ and the objective function to be minimized can be rephrased as

$$\Phi(\mathbf{p}_1, \cdots, \mathbf{p}_N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j \in \mathcal{N}(i)} w_{ij} \left\| \mathbf{p}_i - \mathbf{p}_j - \tilde{\mathbf{p}}_{ij} \right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\| \mathbf{p}_i - \tilde{\mathbf{p}}_i \right\|^2 \qquad (5)$$

where $\mathbf{p}_i$ denote the coordinates available as prior information and $\mathbf{p}_{ij}$ denote the pairwise estimates. The new linear system to be solved in order to minimize (5), in lieu of (8), is derived below in Appendix A, and is much simpler.

Our method thus reduces to a simple three step noniterative approach similar:[4,10] (1) Perform pairwise registration for all neighboring tiles by maximizing a similarity measure such as the normalized cross-correlation (NCC) function. (2) Obtain absolute locations for each tile by solving a linear system such that the global measure of registration in (5) is minimized. Unlike,[8,13] our method can accommodate imprecise prior information about the prior tile locations. (3) Perform blending to obtain a stitched image with nonoverlapping tiles.

We again present a first-principles derivation of the second step similar to the one in section 2. The tile connectivity and the homographies now have a special form. We describe the second step in detail in section 3 in order to point out the connections with the graph

Laplacian and because our method of solution differs from the SVD-based solution reported.[8]

## 3.1. First-Principles Derivation Based Upon the Graph Laplacian

Assuming a rectangular scan with $R$ rows and $C$ columns, we now form a graph with $N = RC$ vertices with edge connectivity based upon eight-connected neighborhoods because only the eight nearest neighbors (NN) have overlapping regions. An example of this connectivity graph is shown in Figure 1a. We assume equal (unity) weights for all connections. Technically, we could choose weights based upon the NCC values, but we avoid doing this so that our solution is unaffected by inconsistencies in the NCC values across tile pairs. Such inconsistencies arise because NCC is not a true measure of pairwise registration accuracy, e.g., a given tile pair containing featureless regions has a high NCC value although the error in NCC peak estimation could be high on account of the relatively flat peak. In fact, evaluating registration accuracy is an important topic in its own right.[18] Note that we could have assigned smaller weights to the diagonal edges within the 8NN connectivity, but we do not do this because the diagonal input translations appear to be just as reliable as the vertical or horizontal translations.

The optimal locations for the tiles are found using a procedure similar to that in section 2. The partial derivative of the above objective in (5) w.r.t. $\mathbf{p}_i$ is

$$\frac{\partial \Phi}{\partial p_i} = \lambda(p_i - \tilde{p}_i) + p_i \left( \sum_{j \in \mathcal{N}(i)} 1 \right) - \sum_{j \in \mathcal{N}(i)} p_j - \sum_{j \in \mathcal{N}(i)} \tilde{p}_{ij} \qquad (6)$$

By setting this gradient to zero, we can see that the global position of each tile is the weighted sum of three terms, namely the average global positions of the neighboring tiles, the average of the pairwise neighboring translations and the prior position. Note that the $X$ and $Y$ coordinates are decoupled in this set of equations. If $\mathbf{x} = [p_1^X, \cdots, p_N^X]$ represents the concatenation of all the unknown $X$-coordinates, we again need to solve a linear system $A\mathbf{x} = \mathbf{y}$, the details for which are provided in Appendix A. The optimal $Y$-coordinates can also be obtained by solving a similar linear system. In case no prior information is available, and reference $\mathbf{p}_N$ is fixed to a known value, the dimension of $\mathbf{x}$ becomes $(N-1)$ and the resulting linear system is also provided in Appendix A.

The objective in Equation (5) is also decoupled in $X$, $Y$ and the part corresponding to the $X$-coordinates alone, $\Phi^X$, can be rephrased in terms of the graph Laplacians and the edge-node incidence matrix.[19] The $N \times N$ connectivity graph Laplacian is defined as $L = D\text{-}W$, where $D$ is a diagonal matrix such that $D_{ii} = \sum_{j \in \mathcal{N}_i} 1$ and $W$ is a binary adjacency matrix such that $W_{ij} = 1$ if $j \in N(i)$ and $j \neq i$. Assuming $N_e$ total edges in the graph, the $N_e \times N$ edge-node incidence matrix $E$ is defined to have the nonzero row entries $E_{ki} = 1$ and $E_{kj} = -1$ if $i < j$ and $j$

$\in N_i$. Note that $L = E^T E$. If $\mathbf{t}$ is the concatenation of all the $N_e$ pairwise translations ($X$-components), then

$$\Phi^X = \mathbf{x}^T L \mathbf{x} - 2\mathbf{t}^T \mathbf{x} + \mathbf{t}^T \mathbf{t} + \lambda (\mathbf{x} - \overline{\mathbf{x}})^T (\mathbf{x} - \overline{\mathbf{x}}), \qquad (7)$$

and the linear system in Equation (10) corresponds to $(L + \lambda I)\mathbf{x} = \lambda \overline{\mathbf{x}} + E^T \mathbf{t}$. In the absence of prior information, the linear system corresponds to $A\mathbf{x} = F^T \mathbf{t} - p_N^X \mathbf{1}_N$, where the matrix $A$ is obtained by dropping the last row and column of $L$, $F$ is obtained by dropping the column of $E$ corresponding to the last tile and $l_N$ denotes the vector of the first $(N-1)$ rows of the last column of $L$. We note that $A = F^T F$ in this case. Since matrix $(L + \lambda I)$ is positive definite,[19] we can use the Cholesky decomposition or the preconditioned conjugate gradient (PCG) algorithm to solve the corresponding linear system. When $\lambda$ is large (indicating more confidence in the prior information), fewer PCG iterations are required for convergence.

With full 8NN connectivity (as described above), in the absence of prior information, matrix A of dimension $(N - 1) \times (N - 1)$ is invertible because we include the Nth tile in the neighborhood connectivity. Also, note that this linear system can be easily modified to accommodate *holes* (i.e., constant intensity tiles which cannot be aligned via pairwise estimation) within the specimen. We simply drop the corresponding images, appropriately change the 8NN connectivity and modify the matrix A and the column vector **y**. If the holes end up disconnecting the specimen, then we need to supply reference pixel coordinates for each connected component and solve one linear system per component.[19] Note that we can also accommodate partial holes that only remove a part of the 8NN connectivity. Using W, we can find all connected components in N steps using depth-first-search or breadth-first-search (regardless of the actual number of components) and compute the matrices $A^c$ corresponding to each component. Since each matrix $A^c$ is also positive definite, we can again use the Cholesky decomposition or the preconditioned conjugate gradient algorithm to solve the corresponding linear system. For each connected component, we select a reference tile and use the following technique to pick reference tile coordinates: We specify the reference coordinates for only one component (the largest one) using the X-Y stage report and use linear extrapolation in order to propagate reference coordinates from the largest component to all others. This choice is motivated by the need for automation and the various stage error considerations discussed in section 4.

## 4. RESULTS AND DISCUSSION

Our microscopy data consists of hematoxylin and eosin (H and E) stained images of sliced prostate specimens imaged with a light microscope after radical

prostatectomy or after a prostate biopsy. After radical prostatectomy, each image tile was acquired using a Nikon Eclipse TE2000 microscope equipped with a Velmex BiSlider motorized stage at 10× resolution with 0.64 micron pixel size and was of size 1392 × 1040 pixels. In addition to problems associated with CCD/stage misalignment and backlash, this Velmex stage has a repeatability error of 5 microns, which translates to roughly 8 pixels at 10×. Following prostate biopsy, each image tile was acquired using a Nikon Eclipse 55i microscope equipped with a PRIOR H101A motorized stage at 40× resolution with 0.16 micron pixel size and was of size 1360 × 1024 pixels. This PRIOR stage does not suffer from backlash problems, but we still have some irremovable CCD/stage misalignment and a repeatability error of 0.75 microns, which translates to roughly 5 pixels at 40×. For pairwise homography estimation, the translation estimate found via NCC was used to identify rough correspondences which were then refined using local normalized cross-correlation and subsequently used for homography estimation.

## 4.1. Global Translation Estimation Results on Radical Prostatectomy Data

Figure 1 shows the global positions computed using the GEM algorithm in section 3 without any prior information, for the largest component of a specimen comprised 79 × 34 tiles that contained both partial and full holes. On account of the holes, only 7942 pairwise translations out of a total possible 11207 8NN-connectivity edges were fed to our system and our system first eliminated more than 50 tiny components. Note the missing tiles (holes) in the center of the specimen in Figure 1. Also note the slight tilt in the X-coordinates possibly on account of the backlash that occurs when the X-Y stage returns to its leftmost X-coordinate after scanning an entire row. This indicates imperfect backlash correction by our stage. Part of this tilt is due to CCD/stage misalignment as well. The complete stitched image is shown in Figure 2. We currently observe a run-time of a few seconds on 79 × 34 images (tiles) for the global registration and a run-time of around 1 hour to obtain all pairwise translations. Therefore, speed is not a critical issue for the global registration. The net computation time of about 1 hour for the global coordinates calculation is reasonable for this application since this acquisition also required about 1 hour (on account of our autofocusing procedure) and the subsequent mosaic pyramid creation also required about an hour (on account of the large image sizes). In addition, since the pairwise registration step is highly parallelizable on account of registrations being independent, greater speed-up can be obtained by using more processor cores. Moreover, the pairwise registration step can be simultaneously
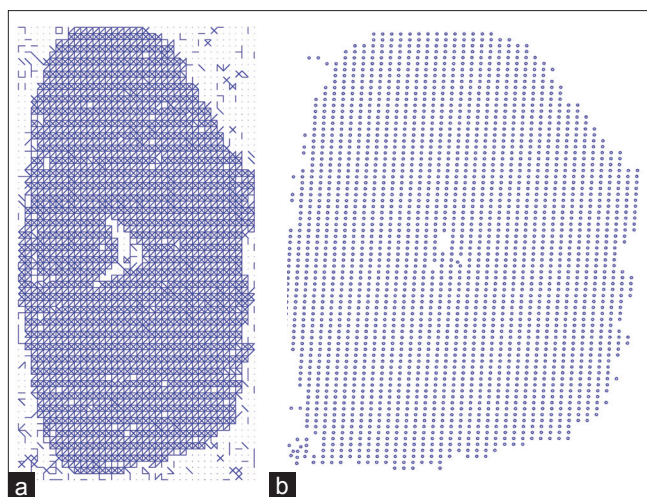


**Figure 1: (a)** Connectivity graph, **(b)** global X-Y coordinates of the largest connected component, for a specimen comprised of 79×34 tiles



**Figure 2: Stitched image of the largest connected component from the 79×34 tiles in Fig. 1**

performed during the hour-long acquisition thus significantly reducing the net computation time for calculation of global coordinates. Since each image tile needs to be registered to four previously acquired tiles (except for tiles located on some of the borders), the parallelization potential is present even when the pairwise registration is performed in conjunction

with acquisition. Additionally, individual pairwise registrations can themselves be sped up, e.g., via parallelization on GPU's (graphical processing units).

## 4.2. Comparison of Methods on Simulated Data

We conducted an experiment with noisy simulated translations (not derived from images) on a 3 × 3 mosaic in order to confirm the error-minimizing advantages of our approach with respect to tree-structured approaches [Figure 3]. We considered equi-spaced square tiles with 10% overlap and added independent zero-mean Gaussian noise ($\sigma_t$=2% of the square image tile size) to each coordinate of the 20 pairwise translations. We chose the central tile as the reference tile for our GEM approach as it minimized trace ($A^{-1}$) and hence the positioning error in Table 1. When averaged over 5000 runs, we obtained an average positioning error of 1.34% [row 5 of the Table in Figure 3] with our GEM approach using the central (5th) tile. For comparison, we also show the average positioning error corresponding to tiles 1 (left top) and 2 (middle top) in rows 3 and 4 respectively. This central tile is also the optimal reference for the tree-structured approaches considered below. We obtained an error and 2.70% (row 1) with the fish-bone approach.[3] Since we allow diagonal connections and each connection is associated with the same noise level, the star graph (rather than the fishbone graph) is the ideal tree for our connectivity graph because it minimizes the sum of all path lengths from the reference node to all other nodes. The star tree yielded an average positioning error of 2.23% (row 2). The mean sum of the squared positioning errors (last column) for the GEM approaches (rows 3-5), the fishbone approach and the star approach assumed their expected values: $\frac{2}{9}\text{Trace}(A^{-1})\sigma_t^2$, $\frac{2}{9}\times 12\sigma_t^2$, and $\frac{2}{9}\times 8\sigma_t^2$, respectively, in this simulation. The GEM approach (rows 3-5) clearly reduces the overall positioning error although the need to obtain all pairwise translations increases its computation time. In the last two rows, we also show the results using prior

information. Zero-mean Gaussian noise with $\sigma_p$=2% and 4% was added to the X, Y coordinates of the true locations and used as prior information in rows 6 and 7, respectively. The parameter λ was set to 8 and 32 in rows 6 and 7, respectively, which represented near-optimal values in this simulation. The squared error with prior information as precise as the translation estimates (row 7) was greatly superior to the best result without prior information (row 5) and even with very noisy prior information, the advantage of using prior information persisted (row 6).

## 4.3. Comparison of Methods on H and E Data

We shall now compare our methods on H and E microscopy images acquired with motorized stages yielding different levels of accuracy in the prior coordinates. Unlike section 4.2, we no longer possess knowledge of the true global coordinates, i.e., the ground truth. We start by presenting a 40× cut-out partial mosaic from a prostate biopsy specimen

**Table 1: Positional errors for different global position estimators (GEM-1, GEM-2, GEM-5 represent our global error minimization technique without prior information with tiles 1, 2, and 5 as references and GEM-P represents our GEM technique with prior information)**

| Method description | Average position error | Average position -squared error |
|---|---|---|
| Fishbone | 2.70 ± 0.64 | 10.75 ± 5.03 |
| Star | 2.23 ± 0.42 | 7.15 ± 3.56 |
| GEM-1 | 1.70 ± 0.58 | 4.21 ± 2.84 |
| GEM-2 | 1.49 ± 0.42 | 3.20 ± 1.76 |
| GEM-5 | 1.34 ± 0.31 | 2.59 ± 1.15 |
| GEM-P, $\frac{\sigma_p}{\sigma_t}=2$ | 1.33 ± 0.32 | 2.27 ± 1.04 |
| GEM-P, $\frac{\sigma_p}{\sigma_t}=1$ | 0.91 ± 0.17 | 1.07 ± 0.38 |
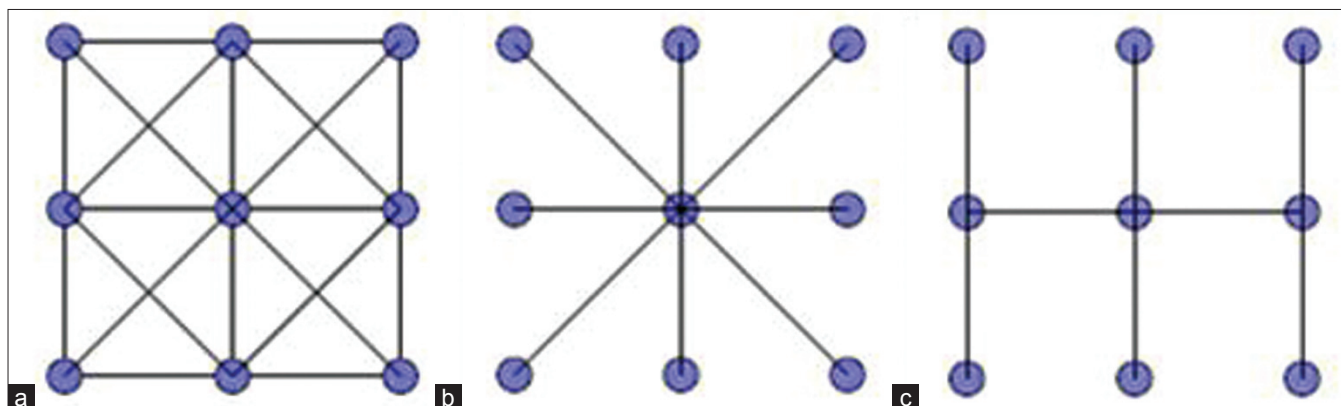


**Figure 3: (a) 3 x 3 connectivity graph used during global error minimization, (b) Star graph (optimal tree, please see text for explanation), (c) Fishbone graph[3]**

obtained with a newly-purchased, freshly calibrated, encoded X-Y stage (PRIOR H101A) providing extremely reliable prior information. There was a 15% overlap between neighboring tiles during acquisition. There was no backlash with this stage, but we note that there was an alignment error between the camera's CCD and the X-Y stage motion which could not be easily removed by rotating the camera on account of camera rotation accuracy limitations. We estimated the true motions in the X and Y directions by performing multiple pairwise registrations and subsequently using a line-fitting procedure. The beneficial effect of using the corrected vs. the original prior coordinates is shown in Figure 4. (Please see insets 4(c)-(f) of highlighted areas to notice the presence of the seams of the tiles in Figure 4a and their absence in Figure 4b. We note that there were no visible stitching errors within this entire mosaic consisting of 72 × 168 tiles. Given this accurate corrected prior information, there is no need to use our GEM algorithms. It also indicates that the Z-motion due to autofocusing can be ignored during stitching.

We now present a 10× partial mosaic from a radical prostatectomy specimen acquired with an inexpensive stage (Velmex BiSlider) providing highly unreliable prior information, with 20% overlap between neighboring tiles. In Figure 5, we show a comparison of a partial 2 × 3 mosaic, independently stitched from the corresponding 2 × 3 tiles, via homography estimation and pure translation estimation with and without prior information. The prior information was obtained via line fitting [Figure 4] and linear extrapolation to coordinates obtained without prior information and was hence biased and not very reliable. As seen in this example, the differences between our three GEM approaches on this partial dataset are not very significant. Figure 5c shows extremely minor errors (please see inset 5e-f of yellow highlighted areas) although we set λ to 2.0 (a lower value than any diagonal Laplacian entry), indicating the need for more reliable prior information. By manual placing landmarks, we obtained ground truth translations between neighboring pairs and compared the results of our automatic pairwise alignment procedure with the ground truth. The median (maximum) Euclidean pairwise alignment error for the 11 2D translations was 1.0 (3.6) pixels, indicating near-perfect pairwise alignment. Some of the homography matrices indicate that the X-Y stage motion does not seem to be perfectly parallel to the plane of the specimen. However, we have observed our pairwise homography estimation to be less stable (please see errors in insets 5g-h of red highlighted area in Figure 5d than the pairwise translation estimation on account of the small overlap and the occasional almost feature-less overlapping regions and are currently in the process of improving our pairwise homography estimation algorithm. Moreover, at this moment, our visualization software

builds multiresolution pyramids computed from 2D tile coordinates. Therefore, we have not attempted to stitch a full slide using homographies as in Figure 2. Note that in Figure 5, we have used image replacement rather than blending in order to highlight the differences between the different GEM approaches.

Finally, in Figure 6, we present an example of a 40× cut-out partial mosaic from a prostate biopsy specimen (56 × 156 tiles) obtained with the same reliable PRIOR stage, as in Figure 4, but after several months of usage. There was a 10% overlap between neighboring tiles during acquisition. The prior information was no longer as reliable and the benefit of using our 2D translation-based GEM algorithm with a prior weight of 0.01 is clearly visible in Figure 6. According to the manufacturer, this stage does not require any recalibration procedure and the exact cause for this drop in accuracy with respect to Figure 4 is not yet clear. Nevertheless, given the fact that the quoted repeatability error of this stage is roughly 5 pixels, we cannot always expect the near-perfect scans corresponding to Figure 5.

We note that although the GEM algorithm without prior information could be readily applied to our radical prostatectomy data obtained with large overlaps (e.g., 20%), the use of GEM algorithms with prior information is more desirable for prostate biopsy scans with smaller overlaps (e.g., 10%). This is because biopsy scans with smaller overlaps lead to more errors in the pairwise alignment on account of the more frequent feature-less (often nearly empty) overlapping regions. (We note that smaller overlaps are desirable on account of the resulting faster acquisition times.) Also, biopsy scans involve thin individual structures and contain more large connected components than radical prostatectomy specimens which typically contain only one or two large components.

## 5. CONCLUSION

We have presented a novel general solution for image mosaicing using graph connectivity and demonstrated its applicability for homography estimation and 2D translation estimation in histological microscopy imaging. Our solution can accommodate imprecise prior information regarding the unknown homographies or translations whenever such information is available. The general solution of section 2 has been applied to the microscopy data in section 4 and the X-Y stage motion estimation algorithm from Section 3 has been used to stitch full slides obtained after radical prostatectomy. Note that the same graph connectivity plays an important role in the solutions for both the generic and the microscopy scenarios, as seen by equations (8) and (10). Through experiments with simulated data, we have shown that prior information can significantly
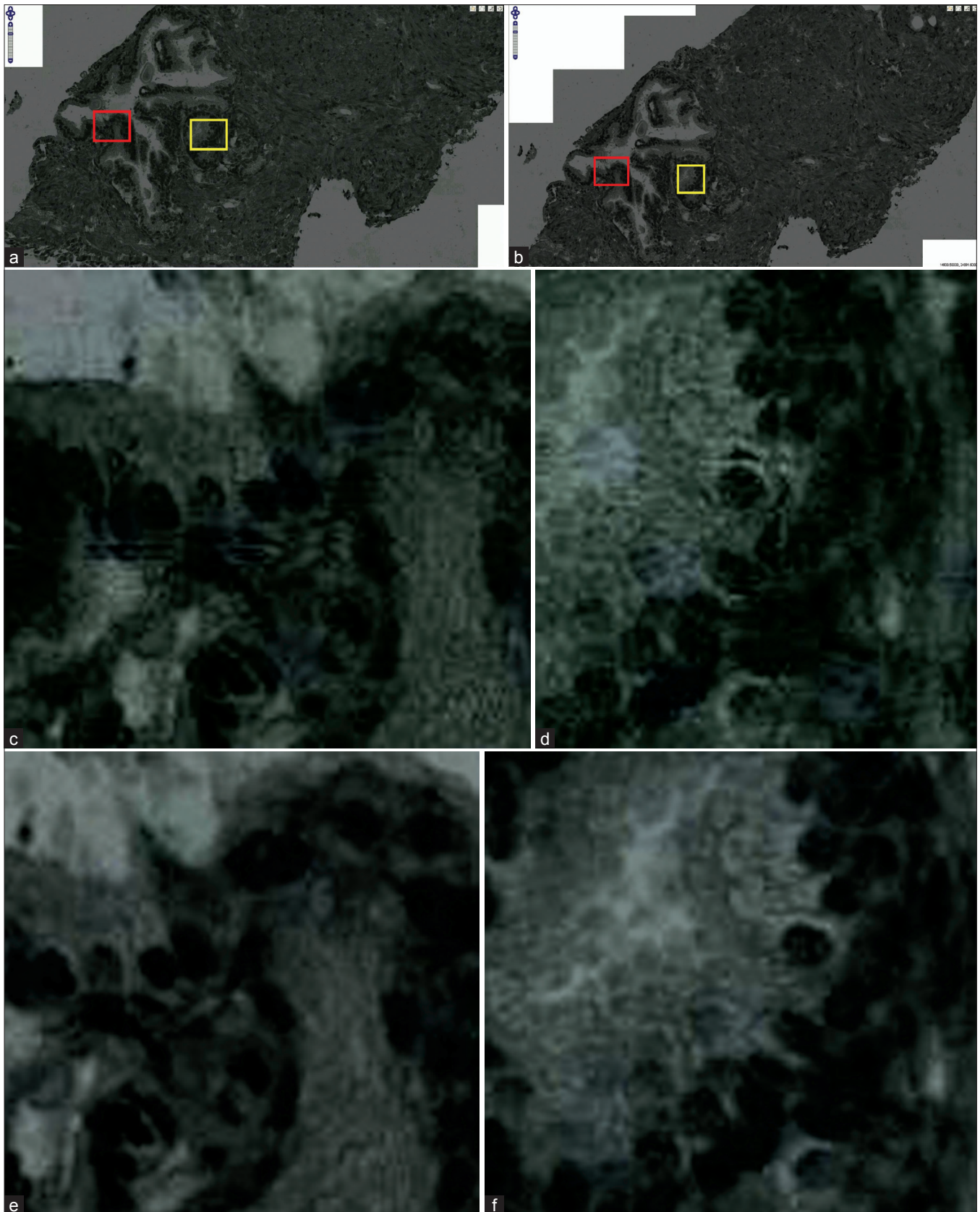
**Figure 4: Precise prior information: (a) Translation-based mosaic stitched using the original prior coordinates, (b) Translation-based mosaic stitched using the corrected prior coordinates, (c) Red inset from 4(a), (d) Yellow inset from 4(a)**

improve the accuracy of the GEM approaches and have also shown that the GEM approaches are superior to tree-structured approaches. We could also use an algorithm similar to the one in section 3 to compute 3D global positions from 3D translations as in the confocal microscopy volumetric stitching application considered in Ref.[8] We also note that having provided a general energy minimization framework incorporating prior information in equation (2), we could replace the Frobenius norm in equation (2) with outlier-resistant norms and perform the optimization in an iterative manner. Another interesting avenue for future work would be to investigating the trade-off between computational speed and accuracy in our GEM approach with prior information by selecting fewer pairwise connections, e.g., those corresponding to an optimal tree. This would reduce the number of pairwise registrations required and hence increase speed.

## ACKNOWLEDGEMENT
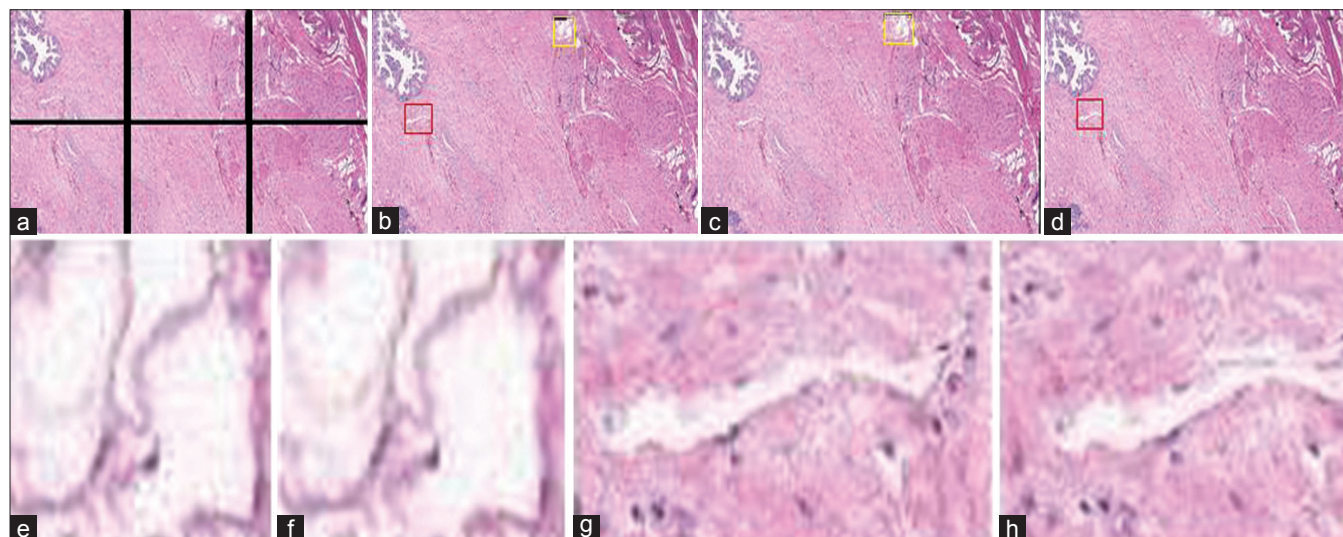
Figure 5: Imprecise prior information: (a) Neighboring 2x3 image tiles (b) Translation-based mosaic stitched using our GEM approach without prior information (Sec. 3) (c) Translation-based mosaic using our GEM approach with prior information (Sec. 3) (d) Homography-based mosaic stitched using our GEM approach without prior information (Sec. 2) (e) (e) Yellow inset from 5(b) (f) Yellow inset from 5(c) [The difference between (e) and (f) is minor, but note that the size of the central purple nucleus slightly reduces in (f)] (g) Red inset from 5(b) (h) Red inset from 5(d)
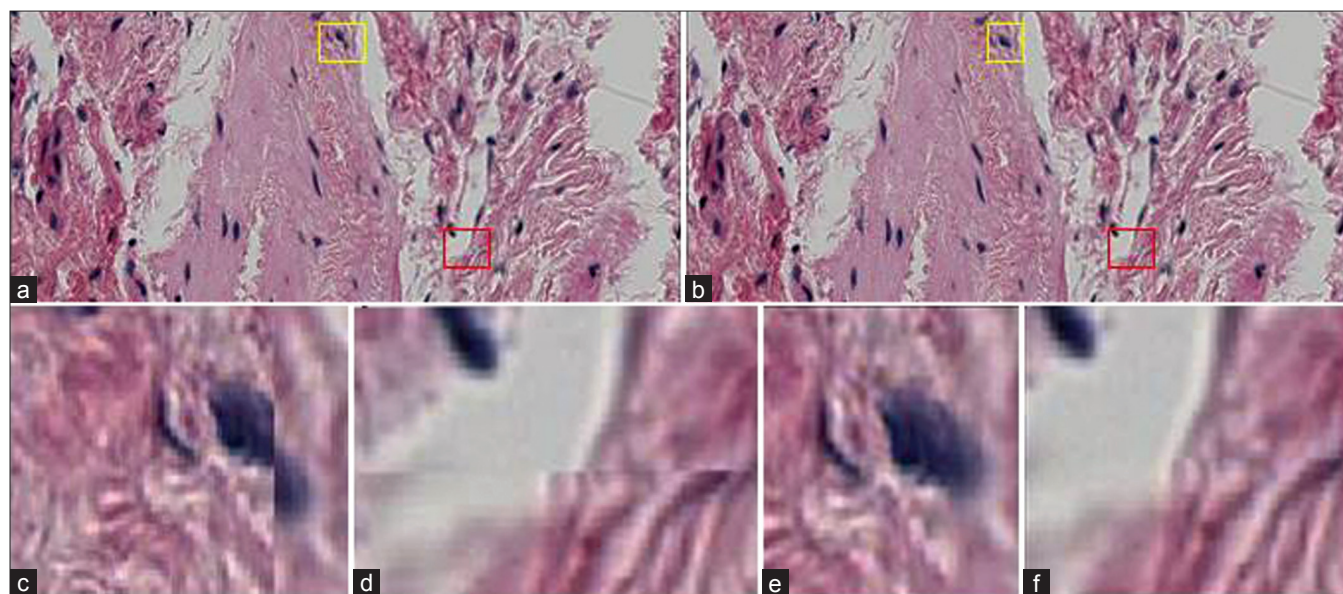


Figure 6: Semi-precise prior information: (a) Translation-based mosaic stitched using the corrected prior coordinates (b) Translation-based mosaic using our GEM approach with prior information (Sec. 3) (c) Yellow inset from 6(a) (d) Red inset from 6(a) (e) Yellow inset from 6(b) (f) Red inset from 6(b)

## APPENDIX A

The matrix A and vector y in the linear system corresponding to equation (4) are as follows:

$$\forall i \in \{1, \cdots, N\},$$

$$A_{iab;iab} = \lambda + \sum_{j \in \mathcal{N}(i)} w_{ij} \quad \forall a,b \in \{1,2,3\}$$

$$A_{iab;icd} = 0, \qquad \forall a,b,c,d \in \{1,2,3\},$$
$$c \neq a, d \neq b$$

$$A_{iab;jcb} = -w_{ij} \tilde{M}_{ij}^{ac}, \quad \text{if } j \in \mathcal{N}(i), j \neq i,$$
$$\forall a,b,c \in \{1,2,3\}$$

$$A_{iab;jcd} = 0, \qquad \text{if } j \in \mathcal{N}(i), j \neq i,$$
$$\forall a,b,c,d \in \{1,2,3\}, d \neq b$$

$$A_{iab;jcd} = 0, \qquad \text{if } j \notin \mathcal{N}(i), j \neq i,$$
$$\forall a,b,c,d \in \{1,2,3\}$$

$$y_{iab;iab} = \lambda \tilde{M}_{ij}^{ab}, \qquad \forall a,b \in \{1,2,3\} \tag{8}$$

When no prior information is available and $M_N$ is known, we set $\lambda$ to 0 in (8), the index $i$ ranges from *1* to *N-1* and y becomes

$$y_{iab} = \sum_{c=1}^{3} w_{iN} \tilde{M}_{iN}^{ac} M_N^{cb} \qquad \text{if } N \in \mathcal{N}(i),$$

$$y_{iab} = 0 \qquad \text{if } N \notin \mathcal{N}(i) \tag{9}$$

The matrix A and vector y in the linear system corresponding to (6) are as follows:

$$A_{ii} = \lambda + \sum_{j \in \mathcal{N}(i)} 1$$

$$A_{ij} = -1 \qquad j \in \mathcal{N}(i), j \neq i$$

$$A_{ij} = 0 \qquad j \notin \mathcal{N}(i), j \neq i \tag{10}$$

$$y_i = \lambda \tilde{p}_i^X + \sum_{j \in \mathcal{N}(i)} \tilde{p}_{ij}^X$$

When no prior information is available and $p_N^X$ is known, we set $\lambda$ to 0 in (10), the index $i$ ranges from *1* to *N -1*, and y becomes:

$$y_i = \sum_{j \in \mathcal{N}(i)} \tilde{p}_{ij}^X \qquad \text{if } N \notin \mathcal{N}(i)$$

$$y_i = p_N^X + \sum_{j \in \mathcal{N}(i)} \tilde{p}_{ij}^X \qquad \text{if } N \in \mathcal{N}(i) \tag{11}$$

## REFERENCES

1. Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press, ISBN-10: 0521540518; 2003.
2. Marks R, Rock S, Lee M. Real-time video mosaicking of the ocean floor. J Oceanic Eng 1995;20:229-41.
3. alignment of large-format multi-projector displays using camera homography trees. Boston, USA: Proceedings of the IEEE Conf. on Visualization; 2002. p. 339-46.
4. Proceedings of the European Conf. on Computer Vision, Lecture Notes in Computer Science, Vol. 1407, pp. 103-119, Freiburg, Germany, 1998.
5. 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. Med Image Anal 2006;10:673-92.
6. Xie YB, Yang P, Gong Y. On the graph-based panorama construction for 2D large-scale microscope images. Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B5. 2008. p. 711-8.
7. Kang E, Cohen I, Medioni G. A graph-based global registration for mosaics. Brisbane, Australia: Proceedings of the IEEE Int. Conf. on Pattern Recognition; 1998. p. 257-60,
8. Emmenlauer M, Ronneberger O, Ponti A, Schwarb P, Griffa A, Filippi A, *et al.* Xuvtools: Free, fast and reliable stitching of large 3D datasets. J Microsc 2009;233:42-60.
9. Marzotto R, Fusiello A, Murino V. High-resolution video mosaicking with global alignment. Washington, DC, USA: Proceedings of the IEEE Int. Conf. on Computer Vision and Pattern Recognition; 2004. p. 692-8.
10. Shum H, Szeliski R. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. Int J Comput Vis 2004;36:101-30
11. Sharp GC, Lee SW, Wehe DK. Multiview registration of 3D scenes by minimizing error between coordinate frames. IEEE Trans Pattern Anal Mach Intell 2004;26:1037-50.
12. Zhang P, Milos E, Gu J. Graph-based automatic consistent image mosaicking. In IEEE Int Conf Robot Biomim 2004.
13. Preibisch S, Saalfeld S, Tomancak P. Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics 2009;25:1463-5.
14. Rojo M, Garcia G, Mateos C, Garcia J, Vincente M. Critical comparison of 31 commercially available digital slide systems in pathology. Int J Surg Pathol 2006;14285-305.
15. Chappelow J, Tomaszewski JE, Feldman M, Shih N, Madabhushi A. Histostitcher©: An interactive program for accurate and rapid reconstruction of digitized histological whole sections from tissue fragments. Comput Med Imaging Graph 2011;35:557-67.
16. Zhang Z. A flexible new technique for camera calibration. IEEE Pattern Anal Mach Intell 2000;22:1330-4.
17. Bradski G, Kaehler A. Learning Open CV. O'Reilly Media. ISBN-10: 0596516134; 2008.
18. Vetter C, Kamen A, Khurd P, Westermann R. A learning-based approach to evaluate registration success. Vol. 6326. Beijing, China: Proc. Int. Conf. on Med Imaging Augmented Reality, Lecture Notes in Computer Science; 2010. p. 429-37.
19. Grady L, Polimeni J. Discrete Calculus: Applied Analysis on Graphs for Computational Science. Berlin, Heidelberg: Springer; 2010.