Feature Based Registration of Fluorescent LSCM Imagery Using Region Centroids

Sang-Chul Lee and Peter Bajcsy

The National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

ABSTRACT

We present a novel semi-automated registration technique for 3D volume reconstruction from fluorescent laser scanning confocal microscope (LSCM) imagery. The developed registration procedure consists of (1) highlighting segmented regions as salient feature candidates, (2) defining two region correspondences by a user, (3) computing a pair of region centroids, as control points for registration, and (4) transforming images according to estimated transformation parameters determined by solving a set of linear equations with input control points. The presented semi-automated method is designed based on our observations that (a) an accurate point selection is much harder for a human than an accurate region (segment) selection, (b) a centroid selection of any region is less accurate by a human than by a computer, and (c) registration based on structural shape of a region rather than on intensity-defined point is more robust to noise and to morphological deformation of features across stacks. We applied the method to image mosaicking and image alignment registration steps and evaluated its performance with 20 human subjects on LSCM images with stained blood vessels. Our experimental evaluation showed significant benefits of automation for 3D volume reconstruction in terms of achieved accuracy, consistency of results and performance time. In addition, the results indicate that the differences between registration accuracy obtained by experts and by novices disappear with an advanced automation while the absolute registration accuracy increases.

1. INTRODUCTION

The problem of 3D volume reconstruction can be found in multiple application domains, such as, medicine, mineralogy, or surface material science. In almost all applications, the overarching goal is to automate a 3D volume reconstruction process while achieving at least the accuracy of a human operator. The benefits of automation include not only the cost of human operators but also the improved consistency of reconstruction and the eliminated training time of operators. Thus, in this paper, we study the performance of fully automated, semi-automated and manual 3D volume reconstruction methods in a medical domain. Specifically, we conduct experiments with fluorescent laser scanning confocal microscope imagery used for mapping the distribution of extracellular matrix proteins in serial histological sections of uveal melanoma. Fluorescent laser scanning confocal microscope is abbreviated as LSCM[1] although other abbreviations, such as CSLM or SCLM, are also found in the literature[2].

We define the 3D reconstruction problem as a registration problem[3]. The goal of 3D reconstruction is to form a high-resolution 3D volume with large spatial coverage from a set of spatial tiles (small spatial coverage and high-resolution 2D images or 3D cross section volumes). 3D volume data are acquired from multiple cross sections of a tissue specimen by (a) placing each cross section under a laser scanning confocal microscope, (b) changing the focal length to obtain an image stack per cross section, and (c) moving the specimen spatially for specimen location. The set of spatial tiles is acquired by LSCM and consists of images that came from (1) one cross section or (2) multiple cross sections of a 3D volume. Our objectives are to (1) mosaic (stitch together) spatial tiles that came from the same cross section, (2) align spatial tiles from multiple cross sections, and (3) evaluate the accuracy of 3D volume reconstruction using multiple techniques. An overview of the 3D volume reconstruction problem is illustrated in Figure 1. Our assumption is that there is no prior information about (a) tile locations and their spatial overlap, (b) cross section feature correspondence and transformation model, and (c) evaluation methodology and metrics.

In general, 3D volume reconstruction without *a priori* information requires performing the following steps. First, select a reference coordinate system or a reference image. Second, determine location of salient features in multiple data sets. This step is also denoted as finding spatial correspondences. Third, select a registration transformation model that will compensate for geometric distortions. Fourth, evaluate registration accuracy with a selected metric. Regardless of the automation category (manual or semi-automated), these selections and evaluations are needed to perform 3D volume reconstruction. The challenges lie not only in making appropriate selections in the aforementioned steps but also in defining optimality criteria for any made selection. In many cases, it is very hard to assess objectively the registration
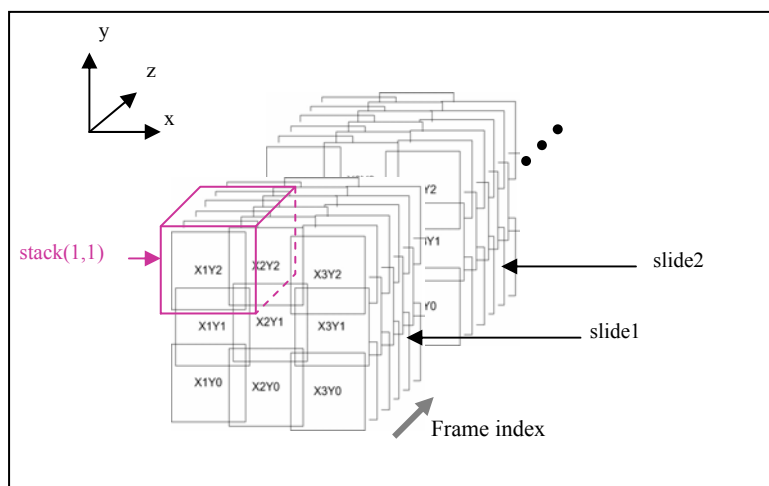
Figure 1. An overview of 3D volume reconstruction from fluorescent laser scanning confocal microscope images.

accuracy due to a lack of *a priori* information about data sets. While the selection challenges are one part of each registration technique, the accuracy assessment challenge is addressed in the experimental evaluation.

There exist many published techniques for 3D volume reconstruction and many commercial tools from multiple vendors that could be used for manual registration[5-9]. An overview of 3D registration tools for MRI, CT, confocal, and serial-section data for medical/life-sciences imaging is provided at the Stanford[i] or at the NIH[ii] web sites. One could list a few software tools that have been developed specifically for LSCM, for example, 3D-Doctor, Science GL, MicroVoxel, 3DVIEWNIX or Analyze[iii]. Most of these tools use manual registration methods and users have to make manual selections described as steps 1, 2, 3 and 4 in the previous paragraph, before any particular software reports registration error associated with registered images. Some software packages include semi-automated or fully automated 3D volume reconstruction for specific imaging modalities under the assumption that visually salient markers have been inserted artificially in imaged specimens. For instance, 3D-Doctor provides a maximum likelihood algorithm for aligning slices[iv] under such assumption.

It is apparent that using artificially inserted fiduciary markers in the registration step 2 allows automating 3D volume reconstruction while keeping the registration error low. However, there still exist medical experiments with LSCM, where fiduciary markers cannot be inserted into a specimen. For example, the placement of fiduciary markers in paraffin-embedded tissue is problematic. The introduction of markers internally may distort tissue and areas of interest. On the other hand, markers placed outside the tissue may migrate during sectioning or expansion of the paraffin. The composition of the marker also poses challenges. Rigid material, such as suture, may fragment or distort the tissue when sections are cut. In addition to attempting to locate fiduciary markers into tissues using the aforementioned techniques, we also attempted to insert small cylindrical segments of "donor tissue" from paraffin-embedded tissues according to the techniques used to construct tissue microarrays[10]. We discovered that the round outlines of donor tissue cores were inconsistent between tissue sections, making it impossible to use these donor samples as reliable internal fiduciary markers.

To our knowledge, there have not been developed robust and accurate automated 3D volume reconstruction methods from a stack of LSCM images without artificially inserted fiduciary markers. Furthermore, there has not been proposed a methodology for assessing the accuracy of already existing, mostly manual, 3D volume reconstruction methods. These facts motivated us to develop (a) software tools for assisting during 3D volume reconstruction (also called semi-

[i] http://biocomp.stanford.edu/3dreconstruction/software/
[ii] http://www.mwrn.com/guide/image/analysis.htm
[iii] *3D-Doctor*, Able Software Corp, Lexington, MA, URL: ttp://www.ablesw.com/3d-doctor/3ddoctor.html
Software tools, Science GL, Quincy, MA, URL:http://www.sciencegl.com/volume_4d/volume.html
MicroVoxel Indec Systems, Inc., Capitola, CA, URL: http://biocomp.stanford.edu/3dreconstruction/software/microvoxel.html
3DVIEWNIX, University of Pennsylvania, Department of Radiology, Philadelphia, PA, URL:
http://biocomp.stanford.edu/3dreconstruction/software/3dviewnix.html
Analyze, Medical Ventures, Rochester, MN, URL: http://biocomp.stanford.edu/3dreconstruction/software/analyze.html
[iv] http://www.ablesw.com/3d-doctor/align.html

automated 3D volume reconstruction) and (b) methodology for assessing inaccuracies of 3D volume reconstruction techniques. Although we are addressing the 3D volume reconstruction problem without artificially inserted fiduciary markers into paraffin-embedded tissue, we still need to identify an internal specimen structure for registration that would be visually salient. For this purpose, tonsil tissue was selected because it contained structures of interest, blood vessels. The tonsillar crypts provided a complex edge against which alignment was possible, and the epithelial basement membrane followed its contour. We stained the blood vessels with an antibody to laminin that also stained the epithelial basement membrane. Therefore, by using the epithelial basement membrane – a normal constituent of the tissue – as the visually salient registration feature in the input LSCM image, we were able to align the tissue sections. Thus, LSCM images of tonsil tissue sections were used for 3D volume reconstruction accuracy evaluations.

Our proposed work aims at estimating upper error bounds for automated, semi-automated and manual 3D volume reconstruction techniques. To achieve our aim, we have developed three mosaicking methods (registration of tiles) and two alignment algorithms (registration across multiple slides). Next, we designed an experimental evaluation methodology that addresses the issues of (a) defining optimality criteria for assessing registration accuracy, and (b) obtaining the ground truth (or reference) images, as encountered in real medical registration scenarios. After conducting experiments with human subjects consisting of experts and novices, we drew conclusions about the 3D reconstruction methods and thoroughly analyzed the driving factors behind our results.

This paper is organized in the following way. Section 2 introduces all image mosaicking and alignment registration methods developed for the accuracy assessment study. Section 3 presents our evaluation methodology for multiple registration methods. Finally, all experimental results are documented and analyzed in Section 4, and our work is summarized in Section 5.

2. REGISTRATION METHODS

As we described in the introduction, there are four registration steps. While certain parameters are defined once during registration of a batch of images, such as a reference coordinate system and a registration transformation model, other parameters have to be determined for each image separately, for example, locations of salient features and their spatial correspondences. Thus, since our goal is to determine the most cost-efficient registration technique in terms of automation/labor and accuracy/time, we have to automate selection of image specific parameters. This leads us to the development of manual, semi-automated and fully automated registration techniques. One should be aware of the fact that although the design of robust automated algorithms is very challenging, there are also challenges in the development of user-friendly software interfaces, for instance, for manual selection of image registration points and their correspondences.

There exist image mosaicking and alignment constraints that have been included in the software development as well. The current software has been developed for mosaicking problem constrained to spatial translations of image tiles and for image alignment problem constrained to affine transformation between two cross sections. The description of the methods developed and evaluated in this work follows next.

2.1 Image Mosaicking

Image mosaicking can be performed by visually inspecting two images, selecting one pair of corresponding points in the overlapping image areas and computing transformation parameters for stitching together image tiles. This approach is denoted as *manual mosaicking* and is supported with computer software that enables (a) pixel selection of matching pairs of points and (b) computation of transformation parameters from a set of control points. If images are stitched together without any human intervention then we refer to the method as *automated mosaicking*. If a computer program pre-computes salient feature candidates and a user interaction specifies correspondences between any two features, then the method is referred to as *semi-automated mosaicking*. Based on the underlying registration mechanism, we also denote manual registration as the pixel-based method and semi-automated registration as the feature-based method.

First, we developed a manual mosaicking method that presents two spatially overlapping image tiles on a computer screen to a user. A user selects two matching pixels and image tiles are stitched. In the next step, a user is presented with the already stitched image and a new tile to select matching pixels. Manual mosaicking is performed in this way till all images are stitched together and the final mosaicked image can be viewed for verification purposes. Second, we have developed a novel semi-automated method that (1) highlights segmented vascular regions (closed contours) as salient feature candidates and (2) computes a pair of region centroids, as control points for registration, after a user defined two region correspondences. This new semi-automated method is designed based on our observations that (a) an accurate point selection is much harder for a human than an accurate region (segment) selection, (b) a centroid selection of any

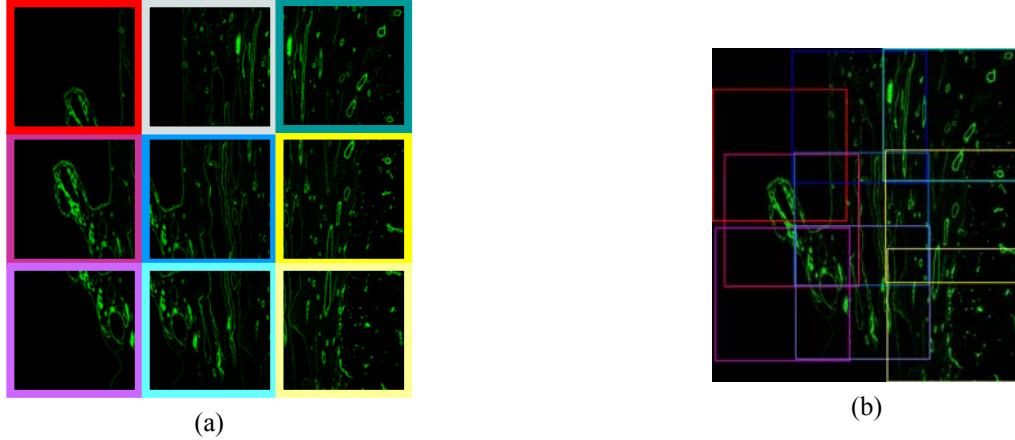(a)   (b)

Figure 2: Image mosaicking problem: (a) Image tiles with colored borders and (b) mosaicked image showing where each tile belongs in the final image basedon its color.

region is less accurate by a human than by a computer, and (c) registration based on structural shape of a region rather than on intensity-defined point is more robust to noise. Third, we present a normalized correlation-based fully automated mosaicking algorithm. This is one of the standard methods used in pattern recognition and computer vision[11,12,] where all possible translational tile overlaps are evaluated based on pixel comparisons according to the Equation (1).

$$E = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(X(i,j)-\mu_X)(Y(i,j)-\mu_Y)}{\sqrt{\sum_{j=1}^{m}\sum_{i=1}^{n}(X(i,j)-\mu_X)^2 \sum_{j=1}^{m}\sum_{i=1}^{n}(Y(i,j)-\mu_Y)^2}} \tag{1}$$

where X and Y are images to be compared (adjacent tiles in image mosaicking), and μ is the mean value of the image. The largest correlation coefficient E indicates the best match of two tiles and provides the sought translational offset for tile stitching. The main advantages of this method are (a) its relatively low computational cost for translation only, (b) robust performance for image tiles acquired with the same instrumentation setup, and (c) no user interaction (full automation). For example, Figure 2 shows how a high-resolution mosaicked image is constructed from nine image tiles.

2.2 Image Alignment

The transformation technique and model selections are two of the challenges of image alignment. We have not applied the normalized correlation based, e.g., fully automated, registration technique developed for image mosaicking to the image alignment problem because it is usually limited to a rigid transformation model, e.g., translation and rotation[13]. Furthermore, the problem of image alignment (or registration along z-axis) is much harder to automate than the problem mosaicking because cross section images are much more dissimilar, for instance, due to the process of cross section specimen preparation (sample warping due to slicing), intensity variation, and structural changes (bifurcating structures).

In terms of transformation model selection, higher order (elastic), local or global models would be preferable to achieve smooth transition of images across slides (higher order continuity). However, the difficulty with higher order models is in robust parameter estimation due to intensity variation (noise), deformation exceeding the order of the chosen model, or bifurcation (appearing and disappearing structures). The transformations using higher order models can be very erroneous due to inaccuracy of parameters, and ultimately distort the 3D anatomical structures (features) by matching accurately small regions but significantly distorting other regions. Rigid transformation model is one of the most popular lower order transformation models designed for rigid structures like bones. However, in our case, the paraffin-embedded tonsil tissue represents a non-rigid structure and has to include deformation like shear due to tissue slicing. Based on the medical specimens of our interest, we chose an affine transformation for modeling distortions between two adjacent cross sections and expected to measure only a small amount of scale and shear deformations. We plan to research automated registration techniques using other transformation models in future.

Given the affine transformation model, the image alignment can be performed by selecting at least three pairs of corresponding points and computing six affine transformation parameters shown below.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{2}$$

The $(x', y')=U(u(X), v(X))$ values are the transformed coordinates of $X=(x,y)$. The four parameters, a_{00}, a_{10}, a_{01} and a_{11}, represent a 2 by 2 matrix compensating for scale, rotation and shear distortions in the final image. The two parameters, t_x and t_y, represent a 2D vector of $X=(x,y)$ translation.

The manual and semi-automated methods for image alignment differ from the methods described for image mosaicking by the need to select at least three pairs of corresponding registration point as opposed to one pair of points sufficient in the case of image mosaicking. The affine transformation parameters are computed by solving six or more linear equations.

3. EVALUATION METHODOLOGY FOR REGISTRATION ACCURACY

In this section, we outline our methodology for assessing upper error bounds of automated, semi-automated and manual 3D volume reconstruction techniques. Our experimental variables include (1) the type of registration problem (image mosaicking and alignment), (2) the type of registration method (automated, semi-automated and manual), and (3) the type of human subject (experts and novices) doing registration. We decided to consider the level of medical expertise as one of the experimental variables of human subjects performing registration based on our hypothesis that the understanding of imaged specimen and its anatomical/structural properties while establishing feature correspondences is significant for obtaining accurate registration results. Thus, human subjects were classified as experts and novices according to their medical understanding of imaged specimen.

Our primary evaluation criterion is registration accuracy with an auxiliary measure of performance time. The challenges of registration evaluations are usually in defining optimality criteria for assessing registration accuracy and in knowing the ground truth (or a reference image). The two fundamental questions that arise during registration accuracy evaluations are (1) what to compare the registered (mosaicked or aligned) image to, and (2) how to compare two images. Next, we describe how these challenges were overcome for image mosaicking and image alignment accuracy evaluations.

3.1 Image Mosaicking Accuracy Evaluation

In the case of image mosaicking, we could carve out several spatially overlapping tiles from one large image and use the original image as the reference (ground truth) image. However, this evaluation setup would not simulate the real problem of mosaicking multiple tiles acquired at different time instances, therefore would not represent unpredictable intensity variations due to fluorescent imaging physics. Thus, we chose to establish the ground truth image and the locations of all nine tiles in this image (denoted as $\vec{T}^{GT} = \{(t_{ix}^{GT}, t_{iy}^{GT})\}_{i=1}^{9}$) in the following way.

First, we took an overview image of a specimen at 20x optical magnification and 3x3 image tiles at 63x optical magnification. The overview image became the ground truth image. Second, tile images (63x magnification) are sub-sampled to match the resolution of the overview image (20x magnification). Third, we find the best match between a sub-sampled tile and the overview image with a template-based search technique using a normalized correlation metric[12]. Fourth, the location of the best tile match is re-scaled to the original tile resolution. Fifth, steps one through four are repeated for all nine tiles to obtain a matrix of tile locations $\vec{T}^*$. Sixth, the matrix $\vec{T}^*$ is normalized with respect to the tile location in the upper left corner (t_{1x}, t_{1y}) of the final mosaic image according to Equation (3). We denote the normalized matrix as the ground truth matrix $\vec{T}^{GT}$ of tile locations.

$$\vec{T}^* = \begin{pmatrix} t_{1x} & t_{1y} \\ t_{2x} & t_{2y} \\ . & . \\ t_{9x} & t_{9y} \end{pmatrix}; \quad \vec{T}^{GT} = \vec{T}^* - \begin{pmatrix} t_{1x} & t_{1y} \\ t_{1x} & t_{1y} \\ . & . \\ t_{1x} & t_{1y} \end{pmatrix} \tag{3}$$

Any other result of mosaicking is represented by a matrix of tile locations $\vec{T}$ and compared with $\vec{T}^{GT}$. The mosaicking registration error E is computed as an average error distance according to the formula in Equation (4). The smaller the error, the better mosaicking accuracy.

$$E = \frac{1}{9} \sum_{i=1}^{9} \sqrt{(t_{ix}^{GT} - t_{ix})^2 + (t_{iy}^{GT} - t_{iy})^2} \tag{4}$$

The proposed mosaicking evaluation methodology using (1) the overview image acquired at low optical magnification as the true reference image and (2) the normalized correlation biased estimation of tile locations $\vec{T}^{GT}$,

simulates more closely real image tile data than a set of carved out tiles from one image. Furthermore, the bias of tile locations $\bar{T}^{GT}$ coming from normalized correlation based matching can be qualitatively expressed by the correlation values in the vicinity of the best tile match with the overview image. Our final remark is related to the selection of the error metric E. Due to the intensity variations of LSCM images, it is preferable to assess accuracy based on spatial matches of salient structures rather than on pixel intensity matches. Thus, an error metric based on tile locations is more appropriate than a metric based on intensity comparisons.

3.2 Image Alignment Accuracy Evaluation

Similar to the case of image mosaicking, we could create a pair of misaligned images by applying a known affine transformation to any image and presenting the original and transformed images to a user for accuracy evaluation purposes. However, this evaluation setup would not simulate the real problem of image alignment where two cross sections might have missing or new or warped structures with a priori unknown intensity variations. Thus, we chose to establish the reference image and its corresponding affine transformation parameters in the following way.

First, we acquired a stack of LSCM images that are co-registered along z-axis because a specimen has not moved while the focal depth of LSCM has varied during image acquisition. Second, multiple stacks of LCSM images are aligned by a manual alignment method and the representative of all resulting affine transformations is recorded. Third, a pair of misaligned images is constructed for accuracy evaluations by taking the first and last image along the z-axis of one LSCM stack and applying the representative affine transformation to the last image. The first and the last transformed images become the evaluation images with the *known* ground truth transformation parameters α^{GT}. All pixel coordinates of the transformed (ground truth) image $\mathbf{P}^{GT} = \{p_1^{gt}, p_2^{gt}, ..., p_n^{gt}\}$ are then defined by the affine transformation α^{GT}; $\mathbf{P}^{GT} = \{p_1 * \boldsymbol{\alpha}^{GT}, p_2 * \boldsymbol{\alpha}^{GT}, ..., p_n * \boldsymbol{\alpha}^{GT}\}$. Based on user's registration input, a computer program computes a set of affine transformation parameters α^{USR} and the corresponding set of transformed pixel coordinates $\mathbf{P}^{USR} = \{p_1^{usr}, p_2^{usr}, ..., p_n^{usr}\} = \{p_1 * \boldsymbol{\alpha}^{USR}, p_2 * \boldsymbol{\alpha}^{USR}, ..., p_n * \boldsymbol{\alpha}^{USR}\}$. The final image alignment registration error E is then calculated as an average error distance over all pixels coordinates according to Equation (5), where n is the number of transformed pixels. Once again, the smaller the error E, the better image alignment accuracy is achieved.

$$E = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(p_{ix}^{gt} - p_{ix}^{usr})^2 + (p_{iy}^{gt} - p_{iy}^{usr})^2} \qquad (5)$$

The proposed image alignment evaluation methodology utilizes (a) confocal imaging to obtain required image frames, and (2) empirically observed affine distortions to prepare test alignment data as close to real data as possible. The justification for choosing the alignment error metric E is similar to the explanation provided for the choice of the mosaicking error metric. An error metric based on pixel locations seems more appropriate than a metric based on intensity comparisons.

3.3 Statistical Comparison of Multiple Registration Methods

Now we describe a statistical test method to evaluate accuracy improvement of the feature-based approach against pixel-based approach. Let $\{E_i^P\}$ and $\{E_i^F\}$ be two paired sets of m measured error values for the pixel-based method and the feature-based method respectively obtained with the same data. In our experiments, the size of the set is relatively large ($m = 50$ for mosaicking and $m = 78$ for alignment). We assume that the paired error values are independent and follow a Normal distribution. The null hypothesis in our tests states that there is no improvement of the feature-based registration approach in comparison with the pixel-based registration approach. We perform the Student's t-test[14] to prove or reject the null hypothesis. We compute $\hat{E}_i^P = (E_i^P - \bar{E}_i^P)$ and $\hat{E}_i^F = (E_i^F - \bar{E}_i^F)$, where $\bar{E}_i^P$ and $\bar{E}_i^F$ are the average errors of each set. Then, we calculate the t value for the paired t-test according to the equation below.

$$t = (\bar{E}^P - \bar{E}^F)\sqrt{\frac{n(n-1)}{\sum_{i=1}^{n}(\hat{E}_i^P - \hat{E}_i^F)^2}} \qquad \textbf{(6)}$$

Given the t value from Equation (6), we obtain the confidence interval (p-value[14]) to prove or reject the null hypothesis (no improvement) using one-tailed cumulative probability distribution function $P(X \leq t)$ with $n-1$ degrees of freedom. The results of statistical comparisons are shown in the next section.
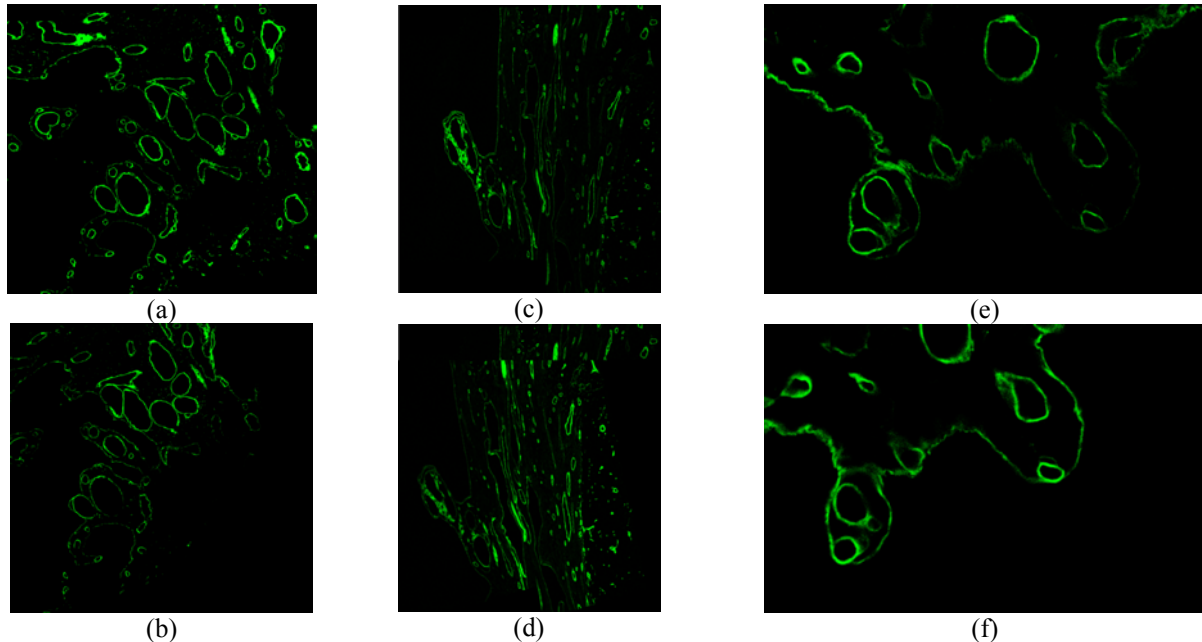
4. EXPERIMENTAL RESULTS

Figure 3. Three pairs ((a)-(b), (c)-(d), and (e)-(f)) of image examples used for alignment evaluation.

The goal of experimental evaluations was to assess the accuracy of multiple 3D volume reconstruction approaches to determine the most cost-efficient (automation/labor, accuracy/time) combination of registration techniques. Furthermore, the results obtained from experts and novices were analyzed to understand the commonalities and difference in registration performance, as well as, registration accuracy improvements due to automation.

The overall experiments consisted of mosaicking 3x3 image tiles (see Figure 2), and aligning three pairs of different cross-sections (see image examples in Figure 3). We report results obtained from twenty human subjects (thirteen experts and seven novices) who participated in our study, and performed manual and semi-automated image mosaicking and alignment registrations. To assess registration consistency, every human subject performed registration three times with any given data set.

4.1 Image Mosaicking

Figure 4 (a) shows the user interface for selecting matching points in two image tiles. Users selected one pair of feature points, one from each tile. Figure 4 (b) illustrates the interface for selecting regions that would be used for centroid calculation. In order to construct a mosaicked image (as shown in Figure 2), eight pairs of points or regions had to be selected. Our experimental results are summarized in Figure 5 and Table 1, and the t-test result comparing the pixel-based and feature-based mosaicking is shown in Table 2.
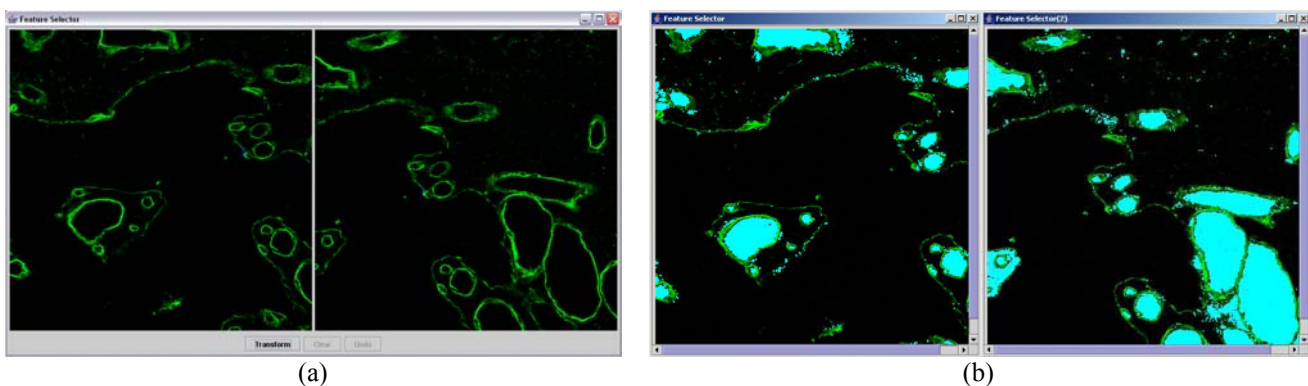

Figure 4: Software interface for (a) manual mosaicking, and (b) semi-automated mosaicking with highlighted regions for selection
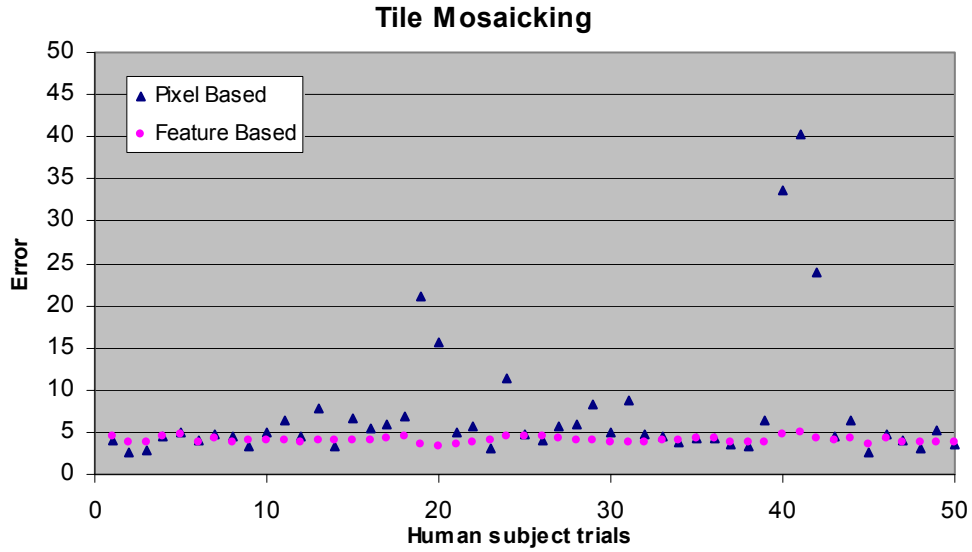
Figure 5. Mosaicking registration errors for all human subjects performing pixel-based (manual) and feature-based (semi-automated) tile mosaicking computed according to Equation (4).

Table 1.   A summary of mosaicking experiments.

	Error (pixels)					Time (seconds)				
	Pixel-Based		Feature-Based		Auto	Pixel-based		Feature-based		Auto
	expert	novice	Expert	novice		expert	novice	expert	novice	
Average	5.97	8.80	4.08	4.05	4.12	221.77	151.10	131.05	97.19	68
Standard deviation	3.58	10.36	0.33	0.40	0	138.26	83.75	57.50	43.18	0
Total average	6.96		4.07		4.12	197.03		119.2		68
Total std. deviation	6.82		0.35		0	125.88		55.01		0
Upper bound (99.73%)	27.42		5.12		4.12	574.67		284.23		68

Table 2. The paired t-test result for errors of the pixel-based and the feature-based methods.

Paired t-test for error in Table 1	The pixel-based and the feature-based method
Degrees of freedom	49
t value	3.019
p value	0.998

   Table 1 and Table 2 lead to the following conclusions. First, fully automated mosaicking is the fastest method, followed by semi-automated (feature-based) and manual mosaicking. Second, manual pixel based image mosaicking is the least accurate with the highest standard deviation among all methods. Third, semi-automated and fully automated mosaicking methods are approximately equally accurate. Fourth, experts using the manual (pixel-based) mosaicking method selected one pair of points/regions more accurately (small average error) and consistently (small standard deviation) than novices although it took them more time. Fifth, the difference in mosaicking average errors and their standard deviations between experts and novices using the pixel-based method disappears when human subjects start using the feature-based mosaicking method. Sixth, the upper error bound of each mosaicking method can be estimated in pixels as the average plus three times standard deviation (99.73% confidence interval), which leads to about 4.12, 5.12 and 27.42 pixel error for the fully automated, semi-automated and manual methods respectively. Seventh, the t-test result in Table 2 shows that the null hypothesis (no improvement) is rejected with 99.8% confidence. Finally, the timesaving for experts and novices using semi-automated method with respect to manual method is 41% and 36% respectively.
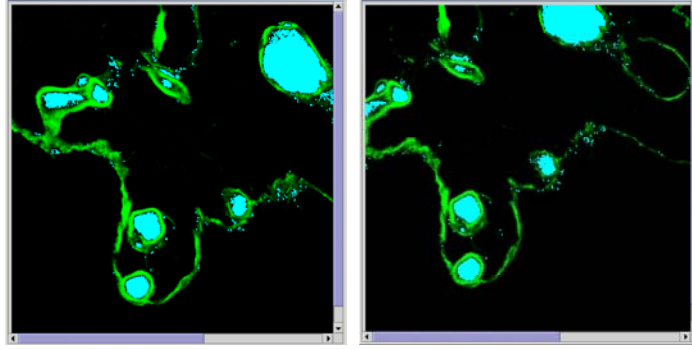
Figure 6. An example of a region candidate (right image, upper border) that would lead to an inaccurate centroid value.

4.2 Image Alignment

For the image alignment experiments, we used the same user interfaces for selecting multiple points and regions as shown in Figure 4. We recommended that human subjects select at least three points or regions, in such a way that they would be well spatially distributed in each image but would not be collinear. If points are close to be collinear then the affine transformation parameters cannot be uniquely derived from a set of linear equations (more unknowns than the number of equations), which leads to large alignment errors. If points are locally clustered and do not cover an entire image spatially then the affine transformation is very accurate only in the proximity of the selected points. However, the affine transformation inaccuracy increases with the distance from the selected points, which leads to large alignment error since the error metric takes into account errors across the entire image area. In order to assess the points selected by a user in terms of their distribution and collinear arrangement, we have designed a spread measure defined as a ratio of the entire image area divided by the largest triangular area formed from the points (see Equation. (7)). One can also view the one-over-spread measure as the point compactness metric.

$$Spread\ Measure = a_{TRIANGLE} / A_{IMAGE} \qquad (7)$$

During the analysis of preliminary experimental results, we encountered three issues. First, it was the existence of region candidates that are close to image borders and are partially outside of image area (see Figure 6). These regions would lead to inaccurate centroid values. We eliminated the border regions from the list of candidates by padding each image with its background color and detecting region connectivity with the background (or the image border). Second, sometimes mouse clicks were not reliably recorded which led to invalid regions. We corrected these three alignment trials assuming the subject could establish region correspondence. Third, we observed large alignment error when human subjects selected almost collinear points or locally clustered points regardless of our recommendations. The spread measure values for all human subjects are shown in Figure 7. The alignment error results of all experiments as a function of human subject trials are shown in Figure 8 and summarized in Table 3. The t-test values for comparing the pixel-based and feature-based mosaicking are shown in Table 4.

The image alignment results in Table 3 and Table 4 lead us to the following conclusions. First, manual (pixel-based) image mosaicking is less accurate and consistent (large standard deviation) than the semi-automated (feature-based) alignment. Based on the t-test result in Table 4, the null hypothesis (no improvement) can be rejected with 99.9% confidence. Second, selection of (a) collinear features, or (b) spatially dense points or regions, can have a detrimental effect on alignment accuracy. Third, novices achieved higher average alignment accuracy than experts. Finally, the difference in alignment errors between experts and novices using the pixel-based method is significantly smaller when human subjects start using the feature-based alignment method. We should also mention that the majority of human subjects selected only three points or regions for aligning two images.

4.3 Discussion of Mosaicking and Alignment Results

We investigate the main factors behind the summarized experimental results and present them in this section. First, based on our scrutiny of the point selection, we observed that experts tend to find one matching structure with a high confidence and select points from the identified structure when using pixel-based registration. In contrary, novices other than the algorithm developers select points randomly and spatially scattered as long as they find an approximate match. This explains why experts outperform novices in the case of mosaicking (one point selection – accurate vs. inaccurate) while novices outperform experts in the case of image alignment (at least three point selection – randomly distributed vs. spatially dense) when using pixel-based registration. This observation is supported by Figure 7 where the
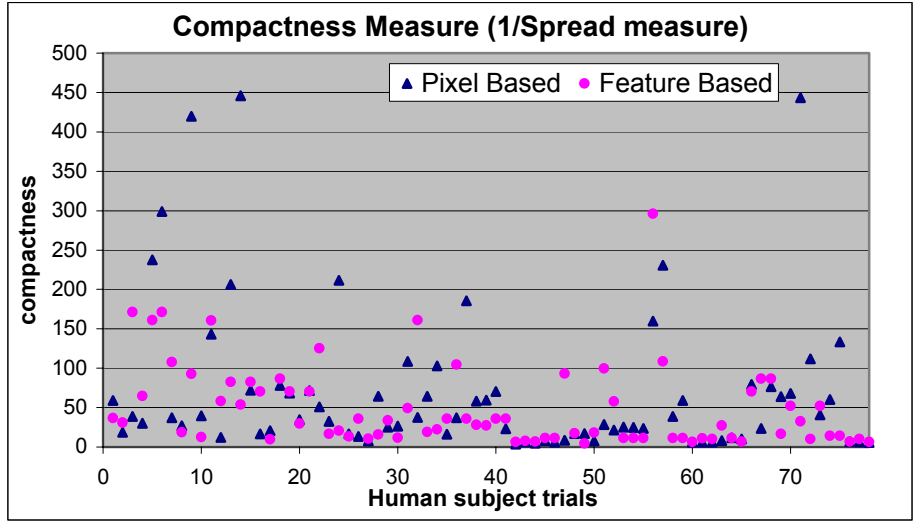
Figure 7. The graph illustrates the assessment of selected points for each human subject by using the compactness measure. The larger compactness measure implies more collinear and spatially dense points.
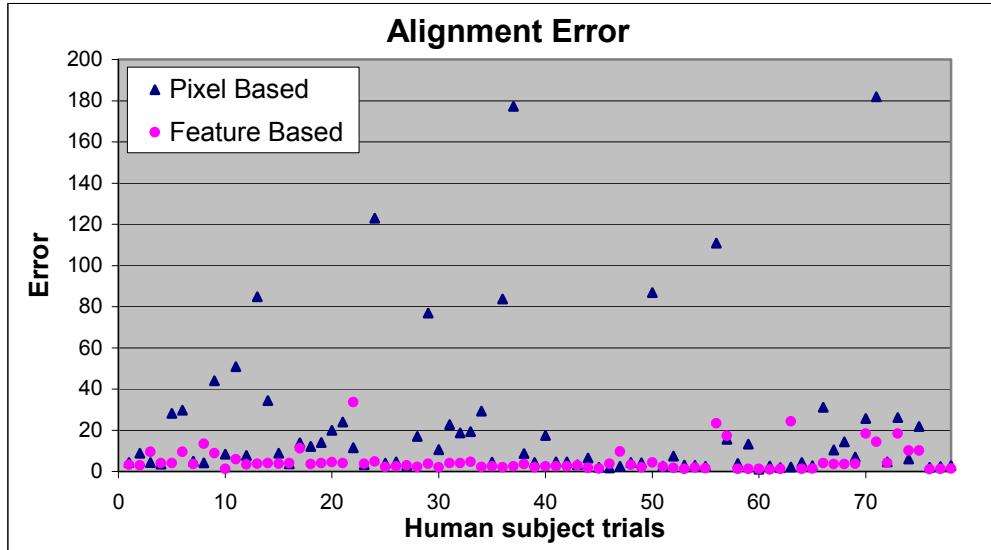

Figure 8. Alignment errors for all human subject trials including pixel-based (manual) and feature-based (semi-automated) image alignment (expert trials are indexed with the trial number less than 24 and novice trials are from [25,78].

expert human subjects indexed between 1 and 24 have on average larger values of area ratios than the novice human subjects indexed between 25 and 78. To circumvent selections of collinear and spatially dense points in future, a computer program has to validate selections in addition to introducing the selection rules for image alignment.

Second, it is apparent that the feature-based registration is faster and more accurate than pixel-based registration for both mosaicking and alignment problems. Our confidence in accuracy improvement is supported by the paired t-test result. We did not report time measurements for the alignment problem because the experiments were conducted on multiple computers with different operating speeds and the reported numbers for mosaicking provide only indications of true comparative values.

Third, the image alignment upper bound errors (23.12 for semi-auto and 129.42 for manual) are much higher than the mosaicking upper bound errors (4.12 for auto, 5.12 for semi-auto and 27.42 for manual). We believe that the main factors behind these differences are (1) one order higher complexity of the alignment problem (intensity and spatial structure variations across slides) in comparison with the mosaicking problem (intensity variations across tiles), (2) a

Table 3.   A summary of image alignment

Error (pixels)	Pixel-Based		Feature-Based	
	expert	novice	expert	novice
Average	23.04	21.04	6.45	4.75
Standard deviation	28.75	38.92	6.48	5.68
Total average	21.66		5.27	
Total std. deviation	35.92		5.95	
Upper bound (99.73% confidence)	129.42		23.12	

Table 4. The paired t-test result for errors of the pixel-based and the feature-based methods.

Paired t-test for error in Table 3	The pixel-based and the feature-based method
Degrees of freedom	77
t value	4.109
p value	0.999

larger degree of freedom in occurring image alignment transformations (rotation, scale, shear and translation) than in mosaicking transformations (translation), and (3) significantly larger sensitivity to human inconsistency in selecting points (attention level, skills, fatigue, display quality) that is qualitative expressed by a much larger standard deviation of errors in the case of alignment (35.92 for manual and 5.95 for semi-auto) than in the case of mosaicking (6.82 for manual and 0.35 for semi-auto).

In addition, we would like to add a few comments about the performance robustness of fully automated and semi-automated methods. Fully automated mosaicking method based on normalized correlation might not achieve the best performance when corresponding salient features have spatially mismatched intensity variations. Semi-automated method based on region centroids might not be used when closed regions cannot be detected due to the spatial structure of an imaged specimen or a very low image quality, for instance, a small signal-to-noise (SNR) ratio and a large amount of intra-region noise. We will investigate in future how to predict accurately centroids of partially open regions and closed regions with speckle noise internal to a region.

5. SUMMARY

We presented an accuracy evaluation of 3D volume reconstruction from LSCM imagery that consists of image mosaicking and image alignment registration steps. The contribution of this paper is not only in developing three registration methods having different levels of automation but also in proposing a methodology for conducting realistic evaluations and performing a thorough analysis of the experimental results. We report accuracy evaluations for (1) three registration methods including manual (pixel-based), novel semi-automated (region centroid feature-based) and fully automated (correlation-based) registration techniques, (2) two groups of human subjects (experts and novices) and (3) two types of registration problems (mosaicking and alignment). Our study demonstrates significant benefits of automation for 3D volume reconstruction in terms of achieved accuracy, consistency of results and performance time. In addition, the results indicate that the differences between registration accuracy obtained by experts and by novices disappear with an advanced automation while the absolute registration accuracy increases.

ACKNOWLEDGEMENT

REFERENCES

1. C. L. Collins, J. H. Ideker and K. E. Kurtis, "Laser scanning confocal microscopy for in situ monitoring of alkali-silica reaction," *Journal of Microscopy*, **vol 213,** no. 2, p. 149, 2004.
2. J.B. Pawley, *Handbook of Biological Confocal Microscopy*, Plenum, New York, 1990.
3. J.B.A. Maintz and M.A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, **vol. 2**, no. 1, pp. 1-36, 1998.
4. P. J. Besl,, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**: 239-55, 1992.
5. D. Hill, P. Batchelor, M. Holden and D. Hawkes, "Medical Image Registration," *Phys. Med. Biol.* **46** R1-R45, 2001.
6. L. G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, 1992.
7. L.D. Cohen, and I. Cohen, "Deformable models for 3D medical images using finite elements and balloons," *IEEE Computer Vision and Pattern Recognition*, 1992.
8. A. Goshtasby, "Registration of images with geometric distortions," *IEEE Transactions on Geoscience and Remote Sensing* **26**: 60-4, 1988.
9. M. Touhy, et al., "Computer-assisted three-dimensional reconstruction technology in plant cell image analysis: applications of interactive computer graphics," *Journal of Microscopy* **147**: 83-8, 1987.
10. A. Nocito, J. Kononen, O. P. Kallioniemi, and G. Sauter, "Tissue microarrays (TMAs) for high-throughput molecular pathology research," *Int.J.Cancer* **94** (1):1-5, 2001.
11. W. K. Pratt, "Correlation techniques for image registration," *IEEE Trans. on Aerospace Engineering Systems,* **10**:353-358, 1974.
12. R. Duda, P. Hart, and D. Stork, *Pattern Classification, Second Edition*, Wiley-Interscience, New York, N.Y., 2001
13. P. A. Van den Elsen, E. D. Pol, and M. A. Viergever, "Medical Image Matching: A review with classification," *IEEE Eng. Med. Biol.*, **12**:26-39,1993.
14. C. H. Goulden, *Methods of Statistical Analysis, 2nd ed.* New York: Wiley, pp. 50-55, 1956.