Published in final edited form as:
Comput Methods Programs Biomed. 2009 December ; 96(3): 182–192. doi:10.1016/j.cmpb.2009.04.012.

Feature-Based Registration of Histopathology Images with Different Stains: An Application for Computerized Follicular Lymphoma Prognosis

Lee Cooper[a], **Olcay Sertel**[a], **Jun Kong**[a], **Gerard Lozanski**[b], **Kun Huang**[a], and **Metin Gurcan**[a]

Lee Cooper: cooperl@ece.osu.edu; Olcay Sertel: osertel@bmi.osu.edu; Jun Kong: kongj@ece.osu.edu; Gerard Lozanski: gerard.lozanski@osumc.edu; Kun Huang: khuang@bmi.osu.edu; Metin Gurcan: gurcan@bmi.osu.edu

[a]Department of Biomedical Informatics, The Ohio State University, 3190 Graves Hall, 333 W. 10th Avenue, Columbus OH 43210

[b]Department of Pathology, The Ohio State University Medical Center, 129 Hamilton Hall, 1645 Neil Avenue, Columbus OH 43210

Abstract

Follicular lymphoma (FL) is the second most common type of non-Hodgkin's lymphoma. Manual histological grading of FL is subject to remarkable inter-and intra-reader variations. A promising approach to grading is the development of a computer-assisted system that improves consistency and precision. Correlating information from adjacent slides with different stain types requires establishing spatial correspondences between the digitized section pair through a precise nonrigid image registration. However, the dissimilar appearances of the different stain types challenges existing registration methods.

This study proposes a method for the automatic nonrigid registration of histological section images with different stain types. This method is based on matching high level features that are representative of small anatomical structures. This choice of feature provides a rich matching environment, but also results in a high mismatch probability. Matching confidence is increased by establishing local groups of coherent features through geometric reasoning. The proposed method is validated on a set of FL images representing different disease stages. Statistical analysis demonstrates that given a proper feature set the accuracy of automatic registration is comparable to manual registration.

Keywords

Image registration; image analysis; feature extraction; follicular lymphoma grading; digital pathology

1. Introduction

Histopathological examination is a crucial step in cancer prognosis. Pathological analysis of biopsy samples is necessary to characterize the tumor for treatment planning. Cancer prognosis that relies on this qualitative visual examination may have significant inter- and intra-reader variability due to due to several factors, such as experience or fatigue at the time of examination (1; 2). Poor reproducibility of histological grading may lead to inappropriate clinical decisions on the timing and type of therapy, and may result in under- or over-

Correspondence to: Lee Cooper, cooperl@ece.osu.edu.

treatment of patients with serious clinical consequences. A computer system capable of extracting quantitative, and thereby more precise and objective prognostic clues, may provide more accurate and consistent evaluations. For this reason we are developing a computer-assisted grading system for one particular cancer type, *Follicular Lymphoma* (FL) (3; 4).

FL is the second most common type of non-Hodgkin's lymphoma that consists of a group of cancers developing from the lymphatic system. The word of "follicular" is derived from round-shaped biological structures, namely "follicles", which are visible under microscope. In current clinical practice, the risk stratification and subsequent choice of therapy for FL mainly depends on the histological grading process that involves computing the average number of centroblasts (CBs), i.e., malignant follicle center cells, as recommended by World Health Organization (5; 6; 7). Due to the large number of follicles usually exhibited in biopsy samples, only ten follicle regions equivalent to a microscopic high power field (HPF) of $0.159mm^2$ are randomly sampled to make this process feasible in practice. Performing CB count over a limited number of follicles can introduce a considerable sampling bias as the selected follicles may not be representative of other sample regions, especially in heterogeneous tumors (2).

With sampling regions identified, centroblasts are then manually counted in HPFs of the selected follicle regions. FL cases are classified into three histological grades based on the centroblast average count: grade I (0–5 CB/HPF), grade II (6–15 CB/HPF) and grade III (>15 CB/HPF) (5). Grade I is usually associated with indolent disease and not treated, while Grade III is associated with aggressive disease and treated aggressively. A multi-site study reported only 61% ~ 73% grading agreement across expert pathologists (1). In addition to this inter-observer variation, the manual counting of centroblasts is very time-consuming, especially when a large number of biological samples need to be examined.

In the current follicle grading processes, pathologists usually resort to using pairs of adjacent slides dyed with different stains to enhance visual contrasts. For example, immunohistochemical (IHC) stains, e.g., CD3 and CD20, provide a clear visual contrast for the follicle structures at low magnifications, e.g. 2×, 4× and 8×. By comparison, Hematoxylin and Eosin (H&E) stain enhance the contrast of the cytological components, and provide better cellular-level detail at higher magnifications, e.g. 20× and 40×. Two representative sample image regions from IHC and H&E stained images captured at 2× magnification are shown in Fig. 1, where follicle boundaries are clearly visible in the IHC (CD3 stain in this specific example) stained image, but are not clearly discernible in the H&E stained counterpart. The proposed computer-assisted system mimics the manual grading procedure, working jointly with pairs of images with IHC and H&E stains. The flowchart of this hybrid FL grading system is presented in Fig. 2.

One of the key steps in this system is to map the spatial coordinates of the detected follicle positions from the IHC stained image to the H&E counterpart image where the centroblast detection will occur. In order for the IHC image analysis to be able to interact with the H&E analysis process, an image registration algorithm is required that allows the output of IHC follicle detection to be fed into the H&E centroblast detection stage. In this paper, such a methodology and its implementation on clinical cases are reported.

Image registration for biological applications has been studied extensively (8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19). Registration can be considered as an optimization problem, posed as finding the optimal transformation $\mathcal{T}$ between two images I_1 and I_2 to maximize a defined similarity measure such as mutual information (20). Registration may also be formulated as a problem of feature matching: finding correspondence between sets of

representative features using descriptors and spatial relations (21). The space of transformations includes rigid, that deals with only rotation and translation, and nonrigid, that compensates for deformations such as bending, stretching, shearing and warping (22; 23; 14). Like most optimization processes, a good initialization is critical for a global optimum outcome. In many cases, a good rigid registration serves as an ideal initialization for non-rigid registration (13). For large images with conspicuous deformations, hierarchical multi-resolution registration methods have also been widely used in medical imaging applications (24; 25).

The key challenge for the registration of sectioned histopathological images is to compensate for distortion introduced by slide preparation. The input slide pairs are cut with a 5 μm thickness from adjacent locations so that the morphological structures vary minimally between image pairs. However, there are discernible global and local deformations between these neighboring tissue sections due to the slide preparation procedure (i.e., sectioning, fixation, embedding, and staining). The preparation process can introduce a variety of nonrigid deformations including bending, shearing, stretching, and tearing. At micron resolutions, even minor deformations become conspicuous and may prove problematic when accuracy is critical to the end application. In order to compensate for such deformations, a nonrigid registration is essential and success depends on establishing a large number of precise spatial correspondences throughout the extent of the image.

An additional challenge for the registration of histopathological images exists when the images to be registered are stained with different stain types, and consequently have dissimilar appearances. An approach based on intensity values requires the ability to resolve similarity between intensity signals using a measure such as mutual information. Such similarity is not necessarily guaranteed for combinations of stain pairs, since for some stain combinations only complex high-order perceptual qualities will be consistent. If the images do exhibit a significant visual similarity, then an approach exists that uses correlation sharpness as a means for classifying local similarity between intensity information (26). However, in the case of follicular lymphoma images with H&E and IHC staining, content at local scales appears as a uniform texture of cellular components, certainly not an ideal condition for intensity comparison between distinct sections. Another approach exists that uses a segmentation of tissue types as input to a registration process (13). The registration reconciles differences in the segmentation by calculating a displacement field that is used for nonrigid registration. Again, this approach is not reasonable in the case of follicular lymphoma, where the content is textural and segmentation is the original problem that a registration is intended to aid.

To address these challenges, this paper proposes a registration approach based on the matching of small salient anatomical features. Small features such as blood vessels appear universally in most tissues and have a common appearance in many stains, making their extraction and matching feasible. These features are used to establish spatial correspondences and register the images in two stages: first rigidly, to roughly align the images, then nonrigidly, to correct for elastic distortions introduced by preparation. The first stage uses a previously established mismatch-tolerant voting procedure (17). With the rough alignment of the images calculated, the second stage establishes coherent local networks of matched features between the images to enhance the confidence of matching and reduce the probability of mismatch and provide a set of spatial correspondences that is satisfactory for nonrigid registration.

The outline of the remaining paper is organized as follows. Section 2 describes the proposed algorithm for registering multi-stained consecutive histopathological FL images. Two

components, including the feature extraction and the actual transformation, are presented. In Section 3, extensive experimental results and the validation processes are presented. Conclusions are presented in Section 4.

2. Methods

To address the challenges of comparing content from consecutive slides stained with different stain types, nonrigid distortion, and feature-rich content, a two stage algorithm is proposed that consists of rigid initialization followed by nonrigid refinement. Both stages operate by matching *high level features*, image regions that correspond to distinct and anatomically significant features such as blood vessels, other ductal structures, or small voids within the tissue area. These matches serve as the control points for calculating spatial transformations to register the image pair. Rigid initialization estimates the rigid alignment of the image pair from the loose consensus of correspondences between anatomical features, following the method presented in (17). The nonrigid stage refines the initialization, by establishing a more accurate set of feature correspondences at a local scale. Initialization reduces the search for matching in the refinement stage, resulting in a lower likelihood of erroneous matches and less computation.

2.1. Data

The input images of FL tissue slides are digitized using a Scope XT digitizer (Aperio, San Diego, CA) at 40× magnification. Tissue slides are collected from the Department of Pathology, The Ohio State University in accordance with an IRB (Institutional Review Board) approved protocol. Slides are prepared by slicing the biopsy specimen in 5 micrometer sections. Adjacent sections are stained pairwise, one of each pair with CD3 and the other with H&E. In this study five pairs of whole-slide biopsy specimens associated with multiple FL patients having different grades of the disease were used.

2.2. Measure for Evaluating Image Registration

For images with the same stain type, an ideal registration would be expected to match the areas of corresponding follicles with perfect overlap, natural morphological differences aside. However, this expectation does not apply to the scenario of images with different stain types, as the difference in appearance of corresponding follicles in each stain type results in significantly different follicle boundaries. In general, the follicles in CD3-stained images appear smaller than their H&E counterparts due to the preparation process (the tissue is boiled or microwaved), and so when correctly registered the CD3 follicles only cover the interior “kernel” regions of those follicle regoins in the H&E images. As illustrated in Fig. 3, this fact implies a possible ambiguity in evaluating registration accuracy from a ground truth perspective in that a decision cannot be made on which result is more optimal. However, since the aim is to identify regions of interest in the H&E image, this ambiguity will not compromise accuracy evaluation from the perspective of follicular lymphoma grading. Therefore, we propose the performance measure as the ratio between the overlap area of the registered CD3 and H&E follicles and the area of the CD3 follicle as follows:

$$r = \frac{Area(\mathcal{T}(S_{CD3}) \cap S_{H\&E})}{Area(S_{CD3})}, \tag{1}$$

where S_{CD3} and $S_{H\&E}$ are follicle regions detected in the CD3 and H&E images and $\mathcal{T}$ is the transformation between the two images.

This quantity is measured for multiple manually marked follicles in each image as described in Section 3.

2.3. Feature Extraction

Extraction of high level features is a simple process as for most types of stains these features correspond to large contiguous regions of pixels with a common color characteristic. For each stain type, a particular color segmentation followed by morphological operations for cleanup usually suffices. Morphological opening is performed to reduce small noisy features resulting from the color segmentation, and morphological closing follows to fill in small gaps. The computational cost of these operations can be significantly reduced by performing the extraction on down-sampled versions of the original images without compromising the quality of the final nonrigid result. Fig. 4 demonstrates sample input and output of the extraction process.

Given the base image B, and float image F, we extract their respective feature sets $\mathcal{B} = \{b_i\}$ and $\mathcal{F} = \{f_j\}$ according to the process described above. Each feature has associated with it a set of *descriptors* used for the matching processes, $\mathbf{b_i}=(\vec{\mathbf{x}}_i^b, s_i^b, e_i^b, \varphi_i^b)$ and $\mathbf{f_j}=(\vec{\mathbf{x}}_j^f, s_j^f, e_j^f, \varphi_j^f)$, where $\vec{\mathbf{x}} = (x, y)$ is the feature centroid, s the feature area in pixels, e the feature eccentricity, and φ the feature semimajor axis orientation.

2.4. Feature Matching

Both the initialization and refinement stages use feature matching schemes to establish correspondences between the base and float images. The following describes the conventions used for feature matching in both stages. Matches between individual features are referred to as *match candidates* if their size and eccentricity descriptors are *consistent*. That is, given the feature sets $\mathcal{B}$, $\mathcal{F}$, a match candidate (b_i, f_j) is established if the descriptors of size s_i^b, s_j^f and eccentricity e_i^b, e_j^f are consistent within given percent difference thresholds $\varepsilon_s, \varepsilon_e$

$$(b_i, f_j) \iff \begin{cases} \frac{|s_i^b - s_j^f|}{min(s_i^b, s_j^f)} \leq \varepsilon_s \\ \frac{|e_i^b - e_j^f|}{min(e_i^b, e_j^f)} \leq \varepsilon_e \end{cases} . \quad (2)$$

If the base and float images are already roughly aligned then φ-consistency may also be enforced in the identification of match candidates.

Both stages also use feature matches to generate model rigid transformations $(\tilde{\theta}, \tilde{T}_x, \tilde{T}_y)$ as part of their matching schemes. Generating a model rigid transformation requires, at minimum, a pair of match candidates. To identify models originating from coherent pairs of match candidates, geometric consistency criteria are used to ensure consistent intra-image distances between feature centroids and also consistent feature orientations. For a pair of match candidates to form a *candidate pair*, $\{(b_i, f_j), (b_k, f_l)\}$, the intra-image centroid-to-centroid distances between features b_i, b_k and f_j, f_l are required to be consistent within the percent difference threshold $\varepsilon_{\vec{x}}$. Additionally, for the initialization stage, the orientations of the feature semimajor axes must be consistent with the model transformation angle $\tilde{\theta}$

$$\{(b_i, f_j), (b_k, f_l)\} \iff \begin{cases} \frac{\left|\left\|\vec{x}_i^b - \vec{x}_k^b\right\|_2 - \left\|\vec{x}_i^f - \vec{x}_l^f\right\|_2\right|}{\min\left(\left\|\vec{x}_i^b - \vec{x}_k^b\right\|_2, \left\|\vec{x}_j^f - \vec{x}_l^f\right\|_2\right)} \leq \varepsilon_{\vec{x}} \\ |\varphi_i^b - \varphi_j^f - \tilde{\theta}| < \varepsilon_\varphi \\ |\varphi_k^b - \varphi_l^f - \tilde{\theta}| < \varepsilon_\varphi \end{cases} . \quad (3)$$

The model transformation $(\tilde{\theta}, \tilde{T}_x, \tilde{T}_y)$ for the candidate pair $\{(b_i, f_j), (b_k, f_l)\}$ is calculated by first solving for the angle $\tilde{\theta} = tan^{-1}((y_i^f - y_k^f)/(x_i^f - x_k^f)) - tan^{-1}((y_j^b - y_l^b)/(x_j^b - x_l^b))$, corrected to the interval $[-\pi, \pi]$. The translation components $\tilde{T}_x$, $\tilde{T}_y$ are calculated using $\tilde{\theta}$ and least squares.

The match candidate and candidate pair concepts are illustrated in in Fig. 5.

2.5. Rigid Initialization

Determining an estimate for rigid registration from a set of feature matches requires a method that is robust to erroneous matchings. This is especially true in microscope images where many features are indistinguishable, and a substantial amount of mismatches are inevitable. The fundamental idea of the method presented in (17) is the recognition that any candidate pair $\{(b_i, f_j), (b_k, f_l)\}$ defines a model rigid transformation $(\tilde{\theta}, \tilde{T}_x, \tilde{T}_y)$, and for carefully chosen candidate matches and candidate pairs, a large portion of the concomitant model transformations will concentrate around the desired parameters in the Euclidean transformation space. Careful choice of matches and match pairs is achieved with a set of consistency criteria enforced at two levels: between feature descriptors for matches between individual base and float features, and geometrically between pairs of such matches. With a set of model transformations identified from consistent candidate pairs, a histogram voting scheme is used to estimate the initialization parameters (θ, T_x, T_y).

The details of rigid initialization are previously published in (17). Sample voting results from a follicular lymphoma image pair are presented in Fig. 6. The associated parameter values are presented in Table 1.

2.6. Nonrigid Refinement

The challenge in nonrigid registration is the sensitivity of computed non-rigid transformations to errors in matching, a consequence of the freedom of such transformations to accommodate distortion. In computing a relatively constrained transformation such as a rigid transformation, the effect of mismatches can be mitigated through the constraints of the transformation and least squares. For most common nonrigid transformation types the effect of a mismatched feature is certainly strong locally, and depending upon the number of matches used may also affect the registration quality globally.

For this reason the standard for establishing matches to compute a non-rigid transformation must be strict to achieve a low probability of mismatch. In the rigid stage, feature comparisons are made globally to accommodate the possibly gross misalignment of the image pair. The rigid transformation is inferred from the modes of the collection of model transformations resulting from the set of all possible candidate pairs (which inevitably includes a large proportion of mismatches). Due to the presence of mismatches from model transformations surrounding these modes these candidate pairs are not appropriate input for computing the transformation of the nonrigid stage. However, the rigid initialization provides a starting point that can reduce the search area for a stricter feature matching procedure that can reduce the likelihood of mismatching and also computation.

Given the rigid initialization, the problem of matching individual features with high confidence can be formulated as a pattern matching problem. Instead of comparing individual features solely via their descriptors, the spatial patterns formed by the collection of features within their neighborhood can be compared to increase the matching confidence. Features that match with a high degree of confidence will have similar spatial patterns of neighboring features with consistent descriptors. Since these neighborhood comparisons are made at a local scale nonrigid distortion is usually mild and local rigidity can be assumed.

Procedurally, the nonrigid matching scheme is as follows: Given feature sets $\mathcal{B}$ and $\mathcal{F}$, for each base feature b_i, the surrounding features in the R^b-neighborhood are identified. Match candidates for b_i are located in the float image within the $\mathcal{S}$-neighborhood centered at $\vec{x}_i^b$, and are matched to b_i based on size s_j^f, eccentricity e_j^f, and orientation φ_j^f (orientation can be used as criteria now that the images are rigidly aligned). For each match candidate f_j, the surrounding features are identified within the R^f-neighborhood of $\vec{x}_j^f$, and match candidates other than (b_i, f_j) are identified. From these other match candidates, candidate pairs are formed with (b_i, f_j), and pairs with model rotation angle $|\tilde{\theta}| > \tau$ are eliminated. The model for each of the remaining candidate pairs is used to transform the two neighborhoods, and the number of base features in R^b that fall within δ of an s, e, φ-consistent float feature are counted. A match (b_i, f_j) is established if the maximum count exceeds the pattern match threshold v and $|\mathcal{R}_i^b|/2$. This process is illustrated in Fig. 7 and summarized in Algorithm Table 1.

Parameters for the nonrigid matching procedure have clear interpretations and can be selected by examining the features for a particular dataset. Neighborhood size R^b is chosen to capture small local networks of features, and depends on the density of features and scan magnification. The match candidate search neighborhood, S, is selected to account for error in the rigid alignment. The match neighborhood size, δ, is chosen to account for physical distortion and noise due to feature extraction including natural morphological differences. Parameter values for the dataset used in this paper are presented in Section 3.

2.7. The polynomial transformation

The collection of point correspondences generated by nonrigid matching provides the information needed to form a mapping that transforms the float image into conformation with the base. A variety of nonrigid mappings are used in practice, differing in computational burden, robustness to erroneous correspondences, and existence of inverse form (22; 23; 14).

The desired transformation qualities include not only the capability to correct nonrigid distortions, but also robustness to match errors, closed inverse form, and computationally reasonable calculation and application. Of the commonly used nonrigid mapping types such as thin-plate spline, local weighted mean, affine, polynomial, and piece-wise variations, polynomial offers a good compromise between warp complexity and the aforementioned qualities. Thin plate spline provides a minimum energy solution which is appealing for problems involving physical deformation, however perfect conformity at correspondence locations can potentially cause large distortion in other areas and excess error if an erroneous correspondence exists. The lack of an explicit inverse form means the transformed image is calculated in a forward direction, likely leaving holes in the transformed result. Methods such as gradient search can be used to overcome the inverse problem, but at the cost of added computation, which can become astronomical when applied to each pixel in a gigapixel image. Kernel-based methods such as local weighted mean require a uniform

distribution of correspondences. Given the heterogeneity of tissue features this distribution cannot always be guaranteed.

Polynomial warping admits an inverse form, is fast in application, and is capable of satisfactorily correcting the mild distortion encountered in sectioned images. Polynomial warping parameters can be calculated using least squares or its variants which can mitigate the effect of matching errors. Affine mapping offers similar benefits but is more limited in the complexity of the warping it can represent.

Second degree polynomials are used for the results in this paper. Specifically, for a point (x, y) in the base image, the coordinate (x', y') of its correspondence in the float image is

$$\begin{cases} x' = a_1x^2 + b_1xy + c_1y^2 + d_1x + e_1y + f_1, \\ y' = a_2x^2 + b_2xy + c_2y^2 + d_2x + e_2y + f_2, \end{cases} \quad (4)$$

Since each pair of matched correspondences provide two equations, we need at least six pairs of correspondences to solve for the coefficients in (4).

2.8. Experimental Procedures

To demonstrate the effectiveness of the automatic nonrigid registration method, the feature extraction and registration algorithms were applied to the five image pairs described in Section 2.1. Magnification was reduced from 40× to 4× using Aperio's ImageScope software, resulting in images roughly 10, 000 × 7500 pixels in size. For feature extraction, the same parameters for color segmentation and morphological operations were used for all image pairs. The automatic registration parameters, presented in Table 1, were also identical for all image pairs. For comparison, manual rigid and manual nonrigid registrations were also performed to the five image pairs, using eight manually selected control point pairs per image pair. A simple Euclidean transformation was used for the rigid registrations. A second degree polynomial transformation was used for the nonrigid registrations.

All computations were carried out on a dual core 2.6 GHz AMD Opteron system with 8 Gigabytes of RAM. Software was developed using a combination of Matlab, and Matlab's C/C++ interface MEX. With the RGB images loaded into memory, the entire process executes in two minutes for a single image pair. Less than one second of that is devoted to the nonrigid matching procedure.

Visual inspection of the feature extraction results revealed that features in two of the five image pairs are not uniformly distributed, being concentrated almost entirely in one half of the tissue area in each case. In regions where features are sparse, nonrigid refinement matches are hard to establish since it is difficult to identify coherent networks of features at a local scale. This can result in spatially clustered control points, and depending on the severity of distortion between the slides, a transformation that is significantly biased to the feature-rich areas of the tissue. The validation analysis that follows is carried out separately on these challenging image pairs and the feature regular image pairs, to illustrate the importance of feature input and the expected outcome if a sufficient feature set can be identified.

2.9. Validation

The procedure for registration validation was motivated by the application of automated FL grading. The goal in this application is to correctly register follicle regions so that follicle segmentations from the CD3 image can be used to direct grading analysis in the counterpart H&E image. To evaluate registration performance in the context of this application, the

overlap of manually identified follicle regions was compared for different registration methods.

For each H&E/CD3 image pair, five corresponding test follicle pairs were selected. The boundaries of each of these test follicle pairs were then marked by five different observers. The same test follicle pairs were marked by each observer, generating a total of 25 follicle pair markings per observer. The overlap ratio demonstrated in Figure 3 was then computed for every follicle test pair marking using the manual rigid, manual nonrigid, and automatic nonrigid registrations for each image pair. We denote these overlap ratios for observer i, image pair j, and follicle test pair k as $\mathbf{Rigid}_i(j, k)$, $\mathbf{Manual}_i(j, k)$, and $\mathbf{Auto}_i(j, k)$ respectively. The feature regular image pairs are the set $j \in \{1, 2, 3\}$ and the challenge image pairs are the set $j \in \{4, 5\}$.

This validation aims to illustrate two points: 1. that nonrigid registration is beneficial in terms of follicle overlap and 2. that the automatic nonrigid registration is comparable to a reasonable manual nonrigid registration. We address these points with three statistical analyses: the boxplot graphical analysis, significance testing by paired t-test, and the Bland-Altman graphical analysis.

The boxplot is a graphical analysis that presents the distributions of the overlap ratios for feature image pairs, separated by both registration method and observer. The median, inner-quartile range, and outliers are plotted for each observer-method set, $\{\mathbf{Method}_i(j, k)\}$, $\forall(j, k) \in \{1, 2, 3\} \times \{1, \ldots, 5\}$, for some i.

To demonstrate the similarities of manual nonrigid registrations, significance testing was performed on these observer-method sets using the *paired t-test*. For each observer i, the overlap ratios were paired by method for all follicles in the feature regular image pairs, $\{(\mathbf{Manual}_i(j, k), \mathbf{Auto}_i(j, k))\} \forall(j, k) \in \{1, 2, 3\} \times \{1, \ldots, 5\}$. The t-statistic was calculated for these method-pair sets,

$$t_i = \overline{D}_i \sqrt{\frac{15}{\sigma_i^2}}, \tag{5}$$

where $\bar{D}_i$ and σ_i^2 are the mean and variance

$$\overline{D}_i = \sum_{j=1}^{3} \sum_{k=1}^{5} (\mathbf{Auto}_i(j, k) - \mathbf{Manual}_i(j, k)),$$
$$\sigma_i^2 = \frac{1}{15-1} \sum_{j=1}^{3} \sum_{k=1}^{5} (\mathbf{Auto}_i(j, k) - \mathbf{Manual}_i(j, k))^2.$$

The t-statistic t_i was compared against the Student's t distribution to compute the p-value p_i.

To further illustrate the similarities between automatic and manual non-rigid registrations, a Bland-Altman graphical analysis was performed. The Bland-Altman analysis is commonly used in biostatistics to examine the extent of agreement between to distinct measurement methods (27). It is included here because it illustrates the performance of the automatic and manual methods well. We note, however, that comparing the overall performance of two registration methods is fundamentally different from the assessment of the agreement of measurement methods. In the case of measurement assessment, agreement between individual samples is critical, since the measurements intended to provide the same

information about some underlying physical state. In registration, follicle overlaps may disagree individually between methods, but the collection of overlaps may still indicate comparable performance.

For each observer i, the difference $d_{j,k}$ and mean $\mu_{j,k}$ were computed

$$\mu_{j,k}=\frac{\mathbf{Auto}_i(j,k)+\mathbf{Manual}_i(j,k)}{2} \tag{6}$$

$$d_{j,k}=\mathbf{Auto}_i(j,k)-\mathbf{Manual}_i(j,k), \tag{7}$$

and the mean and difference tuples $(\mu_{j,k}, d_{j,k})$ were plotted for all follicles in the feature regular image pairs. Along with the mean and difference tuples, the average-difference and 95% confidence intervals are plotted to provide information on the mean performance of the methods and their range of agreement.

Finally, a simple analysis is performed to demonstrate the spatial variation of registration quality in the challenge image pairs. For each follicle k, the overlap ratios $\mathbf{Auto}_i(j,k)$, $j \in \{4, 5\}$ are averaged over observer i.

3. Results

The boxplot is presented in Figure 8. The corresponding means and standard deviations of the observer-method sets are presented in Table 2. Comparing manual rigid and manual nonrigid registrations, the nonrigid registration improves the mean overlap ratio for all markings except those of observer two, demonstrating the benefit of correcting nonrigid distortion. Mean overlap ratios for automatic nonrigid registration are comparable to manual nonrigid, with slight improvements noted for the markings of three observers.

The p-values for the t-statistics of the method-pair sets are presented in Table 3. These p-values range from 0.79 to 0.93 indicating no statistically significant difference between the manual and automatic methods.

The Bland-Altman plot is presented in Figure 9. Tuples plotted above zero indicate better performance for the automatic method. The average-difference is nearly zero for all observers. Most tuples are clustered tightly in the center right of their plot, indicating a high average overlap and small difference for the manual and automatic methods. Each observer has at least one outlier tuple with a difference beyond the 95% confidence limits. For each outlier tuple indicating superior performance for the manual registration, there is a complementary tuple indicating superior performance for the automatic method.

The overlap results from the challenge image pairs illustrate the impact of feature input to the automatic nonrigid registration. Where the test follicle pairs were chosen uniformly throughout the extent of the tissue, the features in the challenge image pairs were not uniformly distributed, resulting in a transformation that is biased to feature-rich areas. The overlap ratios of Table 4 demonstrate this point, where test follicles located in feature rich regions show comparable quality and others apparently suffer from a lack of proximal feature matches.

4. Conclusion and Discussion

This paper presents a method for the nonrigid registration of distinctly stained follicular lymphoma section images. As a key step for fusing the information extracted from images of two different stains, i.e., IHC and H&E, computerized registration serves as a bridge that allows for the combination of valuable information otherwise unique in each resource in a meaningful way. In this particular study, the registration step makes it possible to recognize salient features from both stained images and map the follicle boundaries detected in IHC images to appropriate locations in H&E images. As a consequence, further grading analysis can proceed with H&E counterparts where cellular level analysis is favorable. In the end, by providing accurate follicle boundaries on the H&E images, the registration contributes to more precise CB count, the essential step in the FL grading process

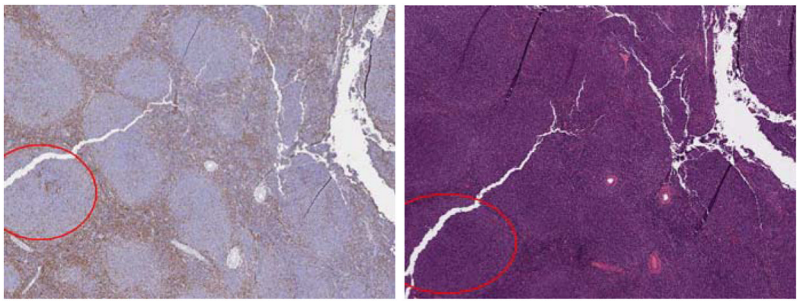
The automatic matching method presented in this paper offers a solution for applications such as microscopy imaging, where a large number of nondescript features are to be matched with high-fidelity. Matching such features individually is a high probability-of-error endeavor, and matching errors can result in poor conformation between the registered image pair due to the freedom of nonrigid transformations. Here, confidence in matches between individual features is enhanced by verifying the existence of coherent networks of features in the surrounding areas.

In terms of registration accuracy, the quality of transformations derived from automatic matching depends on the ability to extract features throughout the extent of the tissue area. When excluding the image pairs where extracted features are sparse and highly spatially clustered, the registrations based on automatic matching are indistinguishable from those based on the manual nonrigid method. This suggests that the registration framework could benefit from a more sophisticated feature extraction process. However, in practice, poorly registered follicles located in feature sparse areas could possibly be avoided by analyzing the spatial distribution of feature matches and their proximities to each follicle.

References

1. Dick F, VSBP. Use of the working formulation for non-hodgkin’s lymphoma in epidemiological studies: agreement between reported diagnoses and a panel of experienced pathologists. Journal of National Cancer Institue. 1987; 78:1137–1144.
2. Metter GE, Nathwani BN, Burke JS, Winberg CC, Mann RB, Barcos M, Kjeldsberg CR, Whitcomb CC, Dixon DO, Miller TP, Jones SE. Morphological subclassification of follicular lymphoma: variability of diagnoses among hematopathologists, a collaborative study between the repository center and pathology panel for lymphoma clinical studies. Journal of Clinical Oncology. 1985; 3(1): 25–38. [PubMed: 3965631]
3. Sertel, O.; Kong, J.; Lozanski, G.; Catalyurek, U.; Saltz, JH.; Gurcan, M. Computerized microscopic image analysis of follicular lymphoma. Proc. of SPIE Medical Imaging; 2008.
4. Sertel O, Kong J, Catalyurek UV, Lozanski G, Saltz JH, Gurcan M. Histopathological image analysis using model-based intermediate representation and color texture: Follicular lymphoma grading. Journal of Signal Processing Systems for Signal, Image, and Video Technology Accepted.
5. Jaffe, ES.; Harris, NL.; Stain, H.; Vardiman, JW. Tumors of haematopoietic and lymphoid tissues. IARC; Lyon, France: 2001.
6. Hiddemann W, Buske C, Dreyling M, Weigert O, Lenz G, Forstpointner R, Nickenig C, Unerhalt M. Treatment strategies in follicular lymphomas: current status and future perspectives. Journal of Clinical Oncology. 2005; 23(26):6394–6399. [PubMed: 16155025]
7. Mann RB, Berard CW. Criteria for the cytologic classification of follicular lymphomas: a proposed alternative method. Hematology Oncology. 1983; 1:187–192.
8. Levinthal C, Ware R. Three-dimensional reconstruction from serial sections. Nature. 1972; 236:207–210.

9. Capowski J. Computer-aided reconstruction of neuron trees from several sections. Computational Biomedical Research. 1977; 10(6):617–629.
10. Moss V. The computation of 3-dimensional morphology from serial sections. European Journal of Cell Biology. 1989; 48:61–64.
11. Brandt R, Rohlfing T, Rybak J, Krofczik S, Maye A, Westerhoff M, Hege H-C, Menzel R. A three-dimensional average-shape atlas of the honeybee brain and its applications. The Journal of Comparative Neurology. 2005; 492(1):1–19. [PubMed: 16175557]
12. Hajnal, J.; Derek, H.; Hawkes, D. Medical Image Registration. CRC Press; Boca Raton, FL: 2001.
13. Braumann U, Kuska J, Einenkel J, Horn L, Luffer M, Huckel M. Three-dimensional reconstruction and quantification of cervical carcinoma invasion fronts from histological serial sections. IEEE Transactions on Medical Imaging. 2005; 24(10):1286–1307. [PubMed: 16229416]
14. Crum W, Hartkens T, Hill D. Non-rigid image registration: Theory and practice. The British Journal of Radiology. 2004; 77(2):S140–S153. [PubMed: 15677356]
15. Wenzel P, Wu L, Sharp R, de Bruin A, Chong J, Chen W, Dureska G, Sites E, Pan T, Sharma A, Huang K, Ridgway R, Mosaliganti K, Machuraju R, Saltz J, Yamamoto H, Cross J, Robinson M, Leone G. Rb is critical in a mammalian tissue stem cell population. Genes & Development. 2007; 21(1):85–97. [PubMed: 17210791]
16. Cooper, L.; Huang, K.; Sharma, A.; Mosaliganti, K.; Pan, T. Registration vs. reconstruction: Building 3-d models from 2-d microscopy images. Proceedings of the Workshop on Multiscale Biological Imaging, Data Mining and Informatics; Santa Barbara, CA. 2006. p. 57-58.
17. Huang, K.; Cooper, L.; Sharma, A.; Pan, T. Fast automatic registration algorithm for large microscopy images. Proceedings of the IEEENLM Life Science Systems & Applications Workshop; Bethesda, MD. 2006. p. 1-2.
18. Prescott, J.; Clary, M.; Wiet, G.; Pan, T.; Huang, K. Automatic registration of large set of microscopic images using high-level features. Proceedings of the IEEE International Symposium on Medical Imaging; Arlington, VA. 2006. p. 1284-1287.
19. Mosaliganti, R.; Pan, T.; Sharp, R.; Ridgway, R.; Iyengar, S.; Gulacy, A.; Wenzel, P.; de Bruin, A.; Machiraju, R.; Huang, K.; Leone, G.; Saltz, J. Registration and 3d visualization of large microscopy images. Proceedings of the SPIE Medical Imaging Meeting; San Diego, CA. 2006. p. 923-934.
20. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging. 1997; 16(2): 187–198. [PubMed: 9101328]
21. Kim T, Im Y-J. Automatic satellite image registration by combination of matching and random sample consensus. IEEE Transactions on Geoscience and Remote Sensing. 2003; 41(5):1111–1117.
22. Bookstein F. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989; 11(6):567–585.
23. Rohr, K. Landmark-Based Image Analysis: Using Geometric and Intensity Models. Springer; New York, NY: 2007.
24. Bajcsy R, Kovacic S. Multiresolution elastic matching. Computer Vision, Graphics, and Image Processing. 1989; 46:1–21.
25. Lefebure, M.; Cohen, L. A multiresolution algorithm for signal and image registration. Proceedings of IEEE International Conference on Image Processing; Washington, D.C. 1997. p. 252-255.
26. Cooper, L.; Naidu, S.; Leone, G.; Saltz, J.; Huang, K. Registering high resolution microscopic images with different histochemical stainings - a tool for mapping gene expression with cellular structures. Proceedings of the Workshop on Microscopic Image Analysis with Applications in Biology; Piscataway, NY. 2007.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]



(a) (b)

Figure 1.

Sample image regions from CD3 and H&E stained FL slides captured at 2× magnification. (a) and (b) correspond to adjacent sections from the same specimen and demonstrate local and global deformations and the difficulty of identifying follicles from H&E-stained slides. Sample regions corresponding to the same follicle are highlighted in red.

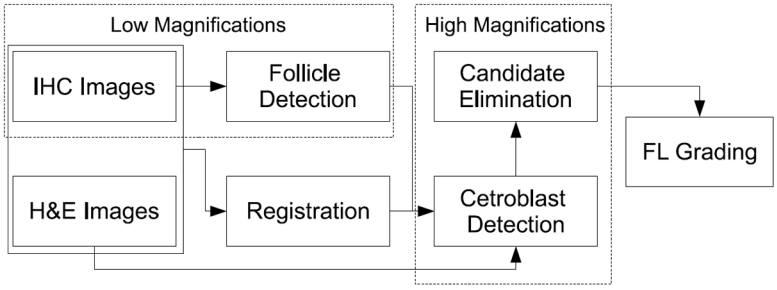
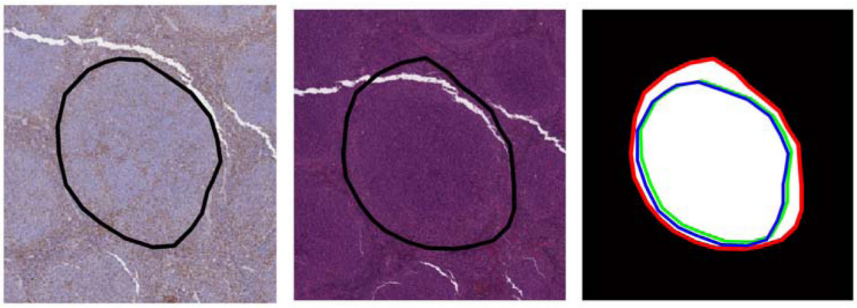


Figure 2.
Flowchart of the computer-aided FL grading system.



(a) (b) (c)

Figure 3.

Overlap ratio score. The corresponding boundaries of a follicle from the CD3 image (a) and it H&E counterpart (b). As shown in (c), different registration results can produce a perfect overlap ratio score due to the differences in follicle appearance between the CD3 and H&E stains. In (c) The red line indicates the H&E follicle boundary, and the green and blue lines indicate different manual registrations of the CD3 follicle boundary to the H&E.

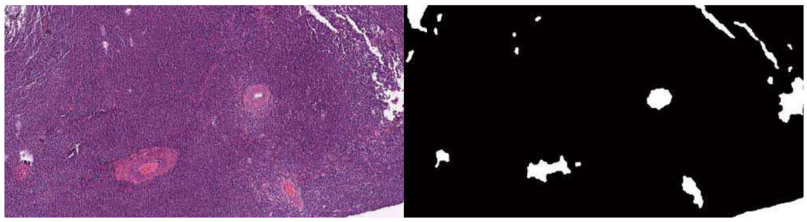


Figure 4.
Feature extraction. This figure contains high-level feature extraction results from a typical H&E image (left). Extracted features, shown in a binary image(right), represent regions such as blood vessels recognized by the use of a combination of color segmentation and morphological operations. Descriptions of centroid location, size, eccentricity, and major-axis orientation are calculated for each feature.

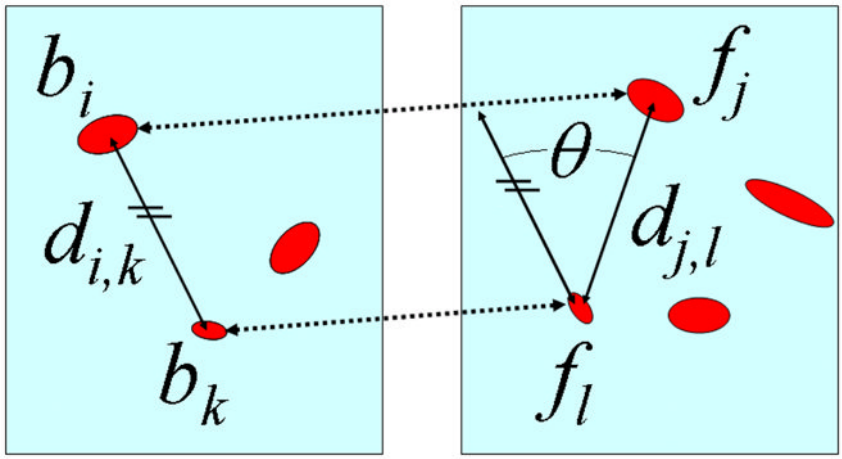


Figure 5.

Rigid feature matching. Features are matched between the base and float images based on size and eccentricity to form *match candidates* (b_i, f_j), (b_k, f_l). Intra-image distance between pairs of match candidates are compared to identify *candidate pairs*. A model rigid transformation, ($\tilde{\theta}$, $\tilde{T}_x$, $\tilde{T}_y$), is defined for candidate pairs with consistent distances.

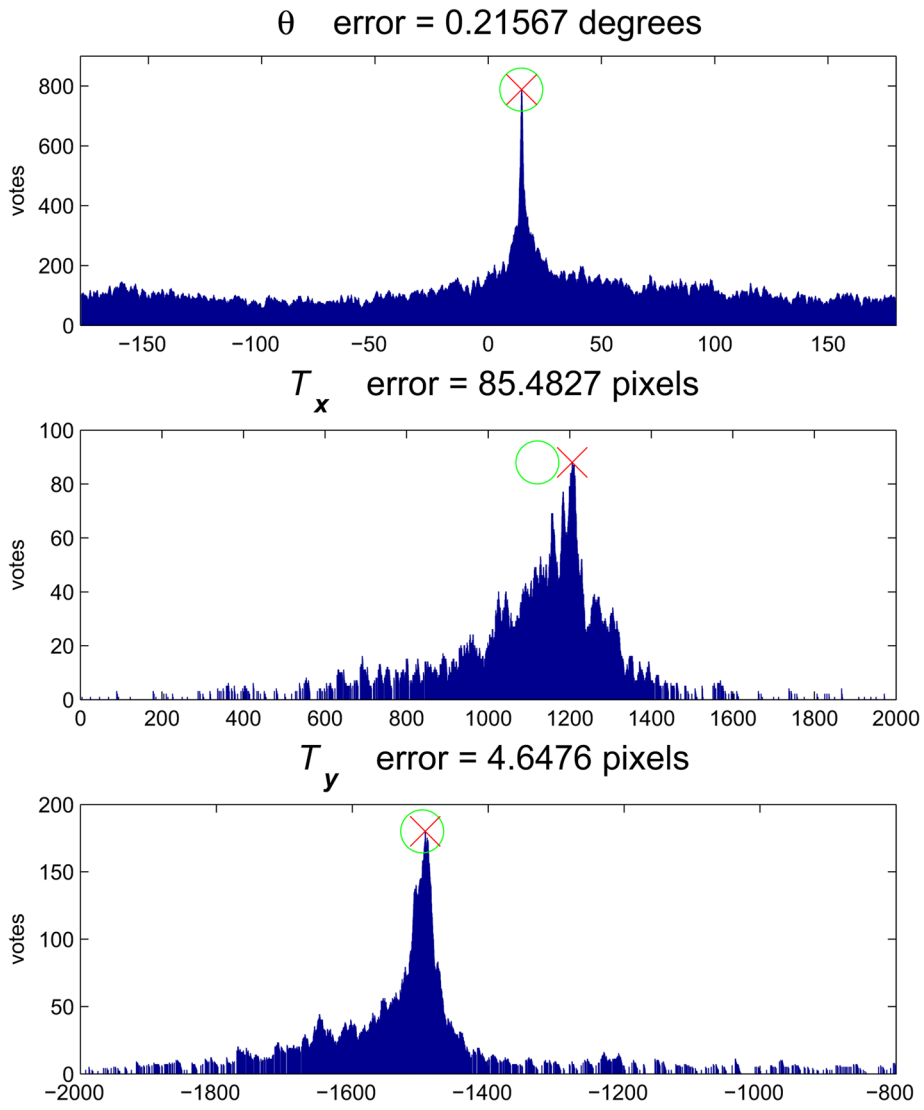


Figure 6.
Sample histogram voting result for rigid initialization of follicular lymphoma image pair. Manual parameter results are shown in red and automatic results in green.

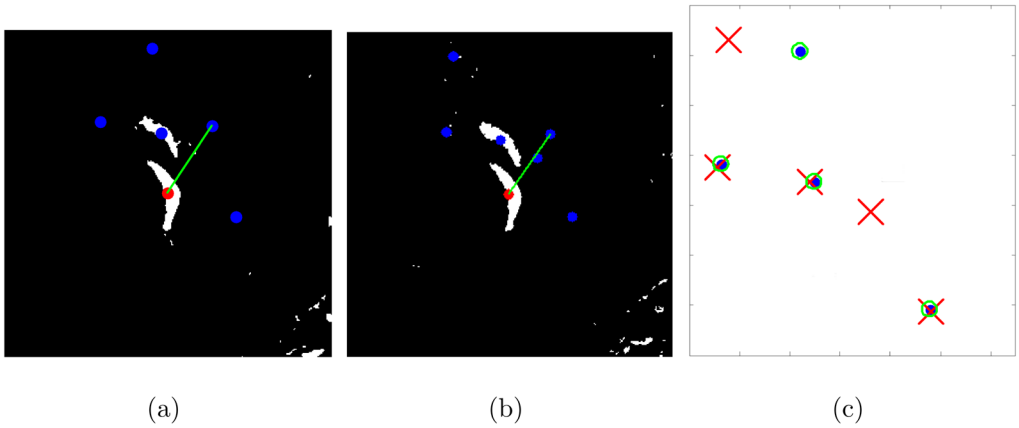


Figure 7.

Nonrigid feature matching. **(a)** Locations of feature b_i (red) and surrounding features in R_i^b-neighborhood (blue). **(b)** Match candidate f_j (red) and surrounding features in the R_j^f-neighborhood (blue). Green lines in (a) and (b) indicate the pairings that generate a model local rigid transformation. **(c)** The float features of $\mathscr{R}_j^f$ (red x's) are transformed onto $\mathscr{R}_i^b$ features (blue dots). In this case, the number of base features with a consistent transformed float feature within its δ-neighborhood (green circle) is three.

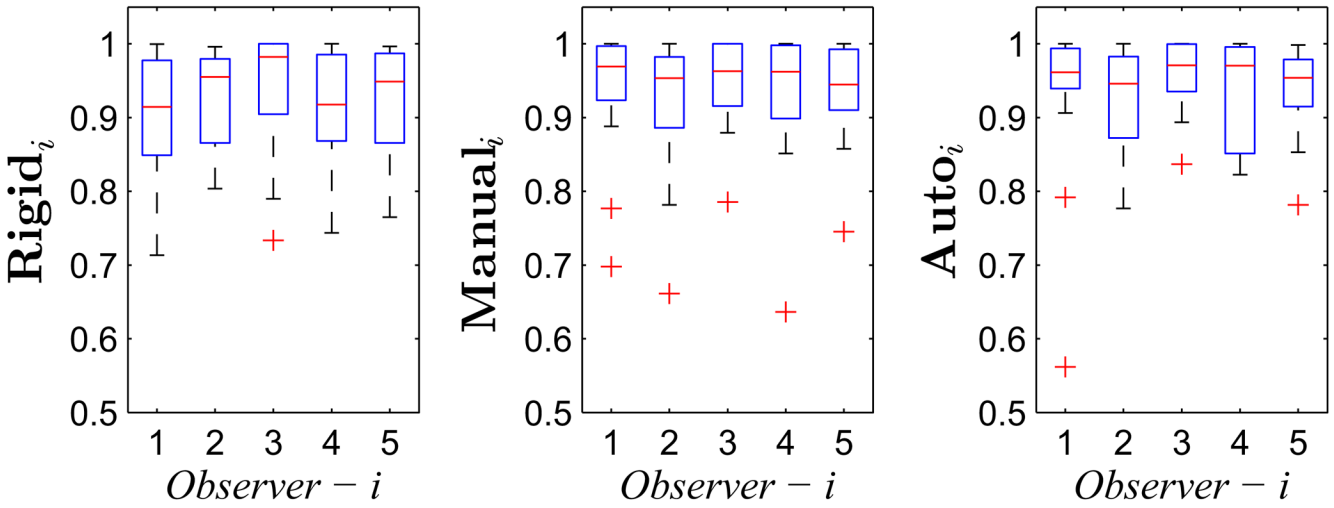


Figure 8.
Boxplots of overlap ratios for observer-methods sets from feature regular image pairs. Outlier overlap ratios from poorly registered follicles are indicated by red cross markers. Mean performance is comparable between manual nonrigid and automatic non-rigid registrations.

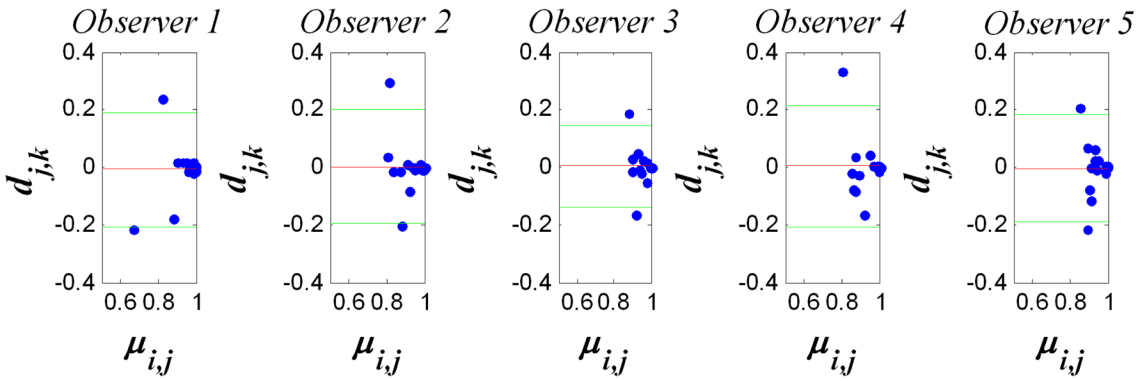


Figure 9.
Bland-Altman analysis of manual and automatic nonrigid registrations. Average difference is indicated in red. The 95% confidence limits are indicated in green.

Table 1

Summary of parameter values used in the tests and validation.

Rigid		Nonrigid	
Parameter Description	**Value**	**Parameter Description**	**Value**
Size similarity (ε_s)	0.1	Base neighborhood (R^b)	1000
Eccentricity tolerance (ε_e)	0.1	Float neighborhood (R^f)	1100
Distance tolerance ($\varepsilon_{\vec{x}}$)	0.1	Search neighborhood (S)	250
Orientation tolerance (ε_φ)	5°	Match neighborhood (δ)	30
Voting interval for $\theta(\omega_\theta)$	0.5°	$\tilde{\theta}$ angle tolerance (τ)	5°
Voting interval for $T(\omega_T)$	30	Pattern match minimum (v)	4

Table 2

Mean overlap ratios and standard deviations for observer-method sets of feature regular image pairs.

Observer *i*	Rigid_i mean ± s.d.	Manual_i mean ± s.d.	Auto_i mean ± s.d.
1	0.8943 ± 0.0930	0.9373 ± 0.0889	0.9306 ± 0.1152
2	0.9223 ± 0.0667	0.9190 ± 0.0950	0.9213 ± 0.0718
3	0.9428 ± 0.0838	0.9520 ± 0.0617	0.9562 ± 0.0477
4	0.9167 ± 0.0850	0.9278 ± 0.0969	0.9316 ± 0.0727
5	0.9247 ± 0.0732	0.9384 ± 0.0691	0.9351 ± 0.0614

Table 3

Significance values of paired t-tests for method-pair sets $\{(\mathbf{Manual}_i(j, k), \mathbf{Auto}_i(j, k))\}$ from feature regular images. The p-values indicate no statistically significant difference between the overlaps for manual and automatic nonrigid registration methods.

Observer *i*	**1**	**2**	**3**	**4**	**5**
p_i	0.7981	0.9301	0.8194	0.8901	0.8905

Table 4

Challenge image pair overlap ratios $\mathbf{Auto}_i(j, k)$, separated by follicle k, and averaged over observers i.

Follicle *k*	**1**	**2**	**3**	**4**	**5**
Image Pair $j = 4$	0.8482	0.2601	0.0505	0.8377	0.9366
Image Pair $j = 5$	0.9189	0.4540	0.9187	0.8862	0.9886

Algorithm 1

Nonrigid Feature Matching

input: Feature sets $\mathcal{B}$ and $\mathcal{F}$, neighborhood sizes R^b, R^f, S, and δ, angle tolerance τ, and vote minimum v

initialize matches $\mathcal{N} = \{\}$

apply rigid transform (θ, T) to float features and correct orientations

for each $b_i \in \mathcal{B}$

identify $\mathcal{R}_i^b = \{b_j : \left\| \vec{x}_i^b - \vec{x}_j^b \right\|_2 \leq R^b\} \backslash b_i$

identify $\mathcal{S}_i = \{f_j : \left\| \vec{x}_i^b - \vec{x}_j^b \right\|_2 \leq S\}$

initialize match candidates $\mathcal{M} = \{\}$

for each $f_j \in \mathcal{S}_i$

compare $s_i^b, s_j^f, e_i^b, e_j^f,$ and φ_i^b, φ_j^f

if (b_i, f_j) s, e, φ-consistent **then** $\mathcal{M} = \mathcal{M} \cup \{(b_i, f_j)\}$

end

for each $(b_i, f_j) \in \mathcal{M}$

identify $\mathcal{R}_i^f = \{f_k : \left\| \vec{x}_j^f - \vec{x}_k^f \right\|_2 \leq R^f\} \backslash f_j$

identify match candidates $\mathcal{X}$ between $\mathcal{R}_i^b, \mathcal{R}_j^f$

identify match pairs $\mathcal{P}$ between (b_i, f_j), $\mathcal{X}$

for each $\{(b_i, f_j), (b_k, f_l)\} \in \mathcal{P}$

compute model transformation $(\tilde{\theta}, \tilde{T}_x, \tilde{T}_y)$

if $|\tilde{\theta}| \leq \tau$ **then**

apply rigid transform $(\tilde{\theta}, \tilde{T}_x, \tilde{T}_y)$ to $\mathcal{R}_j^f$

count $b_m \in \mathcal{R}_i^b$ within δ of consistent $f_n \in (\tilde{\theta}, \tilde{T})$-transformed $\mathcal{R}_j^f$

end

$c(j)$ = max count

end

if max $c \geq v$ **AND** $\max c \geq |\mathcal{R}_i^b|/2$ **then**

$match = \arg\max_j c(j)$

$\mathcal{N} = \mathcal{N} \cup (b_i, f_{match})$

end

output: $\mathcal{N}$