Abstract—Motivated by the need for multimodal image registration in ophthalmology, this paper introduces an algorithm which is tailored to jointly align in a common reference space all the images in a complete fluorescein angiogram (FA) sequence, which contains both red-free (RF) and FA images. Our work is inspired by Generalized Dual-Bootstrap Iterative Closest Point (GDB-ICP), which rank-orders Lowe keypoint matches and refines the transformation, going from local and low-order to global and higher-order model, computed from each keypoint match in succession. Albeit GDB-ICP has been shown to be robust in registering images taken under different lighting conditions, the performance is not satisfactory for image pairs with substantial, non-linear intensity differences. Our algorithm, named Edge-Driven DB-ICP, targeting the least reliable component of GDB-ICP, modifies generation of keypoint matches for initialization by extracting the Lowe keypoints from the gradient magnitude image and enriching the keypoint descriptor with global-shape context using the edge points. Our dataset consists of 60 randomly-selected pathological sequences, each on average having up to 13 FA images. Edge-Driven DB-ICP successfully registered 92.4% of all pairs, and 81.1% multimodal pairs, whereas GDB-ICP registered 80.1% and 40.1%, respectively. For the joint registration of all images in a sequence, Edge-Driven DB-ICP succeeded in 59 sequences, which is a 23% improvement over GDB-ICP.

Index terms: Registration, iterative closest point, key-point matching, fluorescein angiogram, retinal imaging

I. INTRODUCTION

Registration of medical images is playing an increasingly important role in computer-aided diagnosis and treatment [16], [23], [27]. In ophthalmology, doctors make use of red-free (RF) images taken under natural light with a green filter and fluorescein angiographic (FA) images obtained after the injection of fluorescein dye for treatment of diseases related to retinal vasculature abnormality, where lesions are manifested as hyper-fluorescent areas in the FA images. For treatment of choroidal neovascularization (CNV), pretreatment angiographic images are used to determine the location and extent of the CNV, based on the leakage of the fluorescein, and post-treatment RF images are used to assess the extent and intensity of the treatment. By aligning pretreatment FA images and post-treatment RF images, doctors can avoid incomplete or inadequate photocoagulation [49].

To photograph fluorescein angiography [3], it is a common practice that one or two RF images are taken before intravenous injection of the sodium fluorescein dye. After the injection, photographs of the retina are taken at regular intervals during the first circulation of the dye through the retinal vessels, where fine details of the pigment epithelium and the retinal vessels can be recorded. A couple of images may be taken a few hundred seconds after the injection of the dye for examination of hyper-fluorescent areas, which will appear if pathologies are present. The circulation of fluorescein defines five successive phases of an FA sequence: arterial, arteriovenous, venous, late venous, and recirculation. An FA sequence does not always contain images of all five phases, since the speed of circulation of the dye depends very much on the cardiac output of the patient, and the caliber and status of the patient’s vessels.

The ultimate goal of our work is an automatic system tailored for joint registration of all the images in a complete fluorescein angiogram sequence, which includes the RF images and the FA images. The success of this system depends heavily on the robustness of the pairwise registration algorithm, which is the focus of this paper. Pairwise registration is carried out on all possible image pairs, including RF to RF, RF to FA, and FA to FA images.

The main challenges are caused by non-linear intensity differences between modalities and between different FA stages. In particular, vessels—the dominant structures in retinal images—are generally darker than the background tissues in the RF images, but they may be darker or brighter in the FA images, depending on the FA phase in which the image is taken. In addition, pathologies are accentuated by the leakage/blockage of the fluorescein in various phases of the FA sequence. Examples are shown in Figure 1 and 6. Two major challenges in solving the registration problem of a complete FA sequence are:

- Various parts of the vasculature are lit by the fluorescein dye during the circulation, as illustrated in Figure 1. The vessels are dark in the RF image before injection of the dye (Figure 1(a)). During the arterial phase the
arteries are bright, while some veins are showing bright boundaries as a result of the laminar flow (Figure 1(b)). In the venous and the late venous phases, vessels appear brighter than the background in general (Figure 1(c) through (e)). Figure 2 shows a zoom-in version. In addition, pathologies are often manifested as hypo- or hyper-fluorescent areas in the FA images (see Figure 6). This implies a non-linear intensity relation in both local and global scales between multimodal pairs, as well as FA image pairs of different phases.

- Images of the recirculation phase, which could be taken few hundred seconds after the late venous phase, have very low contrast as shown in Figure 1(f) and in Figure 2(f), which results in insufficient salient features for accurate alignment of images.

Our proposed system addresses the above two challenges by exploiting the structural similarity between two images in order to increase the robustness of the pairwise registration, which in turn improves the completeness of the joint registration for the image sequence. The genesis of our pairwise algorithm is Generalized Dual-Bootstrap Iterative Closest Point (GDB-ICP) by Yang et al. [53], which is a fully-automatic, feature-based method. GDB-ICP rank-orders Lowe keypoint matches [25] and refines the transformation computed from each keypoint match in succession by bootstrapping the registration region and the transformation model. When aligning images with substantial appearance differences, Yang et al. have demonstrated experimentally that keypoint matching is the least successful component for GDB-ICP, since the Lowe keypoint descriptor is designed only for invariance to linear changes in intensity, which is a limitation for multimodal registration. Focusing on improving the success of initialization for multimodal image pairs where non-linear intensity differences are expected, we propose a new keypoint matching method to replace the existing one in GDB-ICP. The new method explores the edge information by generating the Lowe keypoints from the gradient magnitude image and enhancing the keypoint descriptor with global-shape context [34] computed from edge points. We name this new pairwise registration algorithm Edge-Driven DB-ICP. The final step of the system is the joint registration [5], which generates mutually consistent transformations for all images.

The contribution of the paper is two-fold. First, we propose a new keypoint extraction and matching algorithm to robustly generate initial transformations in order to handle multimodal image pairs with substantial, non-linear intensity differences. Second, building heavily on existing techniques, we present an end-to-end automatic registration system tailored for multimodal image registration for a complete FA sequence.

II. RELATED WORK

A. Multimodal Retinal Image Registration

Image registration is a fundamental component in automatic image analysis, and it is playing an increasingly important role in the medical imaging domain [16], [23], [27], [40], where accurate alignment of intra- and inter-modality images is required for integration of information. In the area of retinal image registration, numerous techniques have been proposed for alignment of images from the same modality [1], [6], [9], [12], [17], [19], [22], [24], [26], [29], [30], [35], [38], [41], [46], [48], [54]. Due to the challenges of substantial, non-linear intensity differences between modalities and FA phases and low contrast in the late phase images, very few techniques have been developed for multimodal registration [8], [11], [28], [37], [39], among which only Choe et al. [8] and Dréo et al. [11] validated their work with images of complete FA sequences. We will focus our discussion on the related techniques which are designed either for multimodal registration, or for intra-model registration but have demonstrated success with RF-FA image pairs if one of the images in the pair is inverted first.

To achieve multimodal registration of retinal images, algorithms often extract information consistent between images and invariant to the intensity changes to drive the alignment process. Techniques in the literature can be classified as feature-based or intensity-based. For a fully-automatic feature-based approach, the focus is on extraction of interest points robust to the intensity changes. In [8], a bifurcation point (Y-shape feature) is detected by extracting 3 bright (or dark) vessels from a circular boundary of a candidate bifurcation location. Bifurcation points are matched using the local max-

![Fig. 1. Images of different phases of a complete FA sequence: (a) red-free, (b) arterial phase, (c) arteriovenous phase, (d) venous phase, (e) late venous phase, and (f) recirculation phase.](image-url)
imization of mutual information and the estimation method
is random sample consensus (RANSAC). In [39], edges are
extracted along the boundaries of vessels and opposite edges
are grouped for computation of the vessel centerlines. Several
different initial transforms are refined based on alignment of
randomly selected vessel points from the moving image with
the vessels of the reference image using Powell’s method.
Many intra-modal registration techniques have been adopted
for registration of RF with FA images, in which case one of
the images is inverted first to simulate the vessel appearance
for registration of RF with FA images, in which case one of
the refined transformations is below a threshold. Laliberte et al. [22] also detect
affine transformation is estimated using the Bayesian Hough
are matched using the angle-based invariants and the best
the vessel segments that come to meet. The bifurcation point s
segmentation, and the bifurcation points with orientation s of
Zana et al. [54] detect the vascular tree using morphological
in another modality for feature extraction [54], [22], [48].
Many intra-modal registration techniques have been adopted
for registration of RF with FA images, in which case one of
the images is inverted first to simulate the vessel appearance
for registration of RF with FA images, in which case one of
the refined transformations is below a threshold. Laliberte et al. [22] also detect
affine transformation is estimated using the Bayesian Hough
are matched using the angle-based invariants and the best
the vessel segments that come to meet. The bifurcation point s
segmentation, and the bifurcation points with orientation s of
Zana et al. [54] detect the vascular tree using morphological
in another modality for feature extraction [54], [22], [48].
Many intra-modal registration techniques have been adopted
for registration of RF with FA images, in which case one of
the images is inverted first to simulate the vessel appearance
for registration of RF with FA images, in which case one of

B. Keypoint Description

Keypoint-based methods have been widely used in registration, feature-based object recognition, and estimation of motion and structure. Keypoints are interest points enhanced with a descriptor allowing them to be matched in a similarity space. A keypoint-based method usually consists of two steps: interest point detection and descriptor computation. The former identifies the points whose neighborhoods are relatively informative in the image, and the latter provides a description of the image neighborhood and a method to compute the difference between two descriptors. The descriptor is usually crafted to be invariant to certain types of geometric deformation or intensity changes. In the past decade, there has been an impressive body of work on keypoint localization and matching. There are few review papers on quantitative evaluation of some popular methods [7], [31], [33]. One of the main challenges of our application is to match keypoints from images with substantial, non-linear intensity differences. To the best of our knowledge, most of the descriptors described in the literature are designed to be robust up to affine illumination
and bifurcation points are detected from dark vessels. Matching of the bifurcation points is performed through invariant
registration, and matches are tested one by one for initialization. The initial alignment is refined using robust Iterative-Closest-Point with dual-bootstrapping (DB-ICP) on the registration region and transformation model. The algorithm terminates as soon as the weighted average centerline alignment error of one of the refined transformations is below a threshold. Ryan et al. [42] proposed a registration technique independent of control point extraction methods. It generates possible similarity transformation coefficients for all combination of control point pairs. The cluster of coefficients is identified using Expectation-Maximization, and refined using linear regression. In general, feature-based registration methods make no assumptions on the image modality, provided that the feature-extraction algorithm can reliably extract control points for the specific modality.

Fig. 2. A zoom-in region showing the non-linear intensity differences: (a) red-free, (b) arterial phase, (c) arteriovenous phase, (d) venous phase, (e) late venous phase, and (f) recirculation phase.
changes—allowing shifting and scaling—which arise from changing the brightness and the position of the light source.

Schmid et al. [44] extract Harris corners as the keypoints, and combine the components of the local derivatives up to the third order to obtain a scale and rotation invariant descriptor for each keypoint. Also based on local derivatives, the steerable filters developed by Freeman et al. [13] steer the derivatives in a particular direction, such as the direction of the gradient for rotation invariance. Both descriptors can be made invariant to affine illumination changes if the local image patch is normalized by its mean and standard deviation [31].

Lowe [25] detects keypoints as the extrema in the Difference-of-Gaussian (DoG) scale space. For each keypoint, the intensity gradient vectors within its neighborhood are collected in histograms to form a descriptor of 128-component, named Scale Invariant Feature Transform (SIFT). The description vector is divided by the square root of the sum of squared components to provide some degree of robustness to illumination changes. In the study conducted by Mikolajczyk and Schmid on comparison of local descriptors [31], a number of descriptors are tested for invariance to scale changes, rotation, affine transformation, and affine illumination changes. SIFT outperforms the rest for its robustness to geometric deformations, whereas steerable filters perform the best for illumination changes.

Ambiguity can often arise when matching descriptors extracted from only the local neighborhood. Using the edge pixels of the entire image, Belongie et al. [2] build the relative distribution (histogram), named shape context, of the edge points with respect to a keypoint in the log-polar space to describe each keypoint. The descriptor is scale and rotation invariant. The method is robust to illumination changes, both linear and non-linear, since only edge points are involved. Mortensen et al. [34] augment the local descriptor with global context by combining SIFT and the idea of shape context. Instead of using edge points, they use maximum curvature at each point and normalize the descriptor to unit magnitude for contrast invariance. The descriptor is invariant to rotation, but not scale changes, since the diameter of the shape context neighborhood is set to be the image diagonal. Inspired by [34], our descriptor combines SIFT and shape context extracted from the gradient magnitude and intensity images, respectively, to exploit the edge information preserved under non-linear illumination changes.

A. Feature Extraction

Our image registration is driven by geometric features extracted from images. In this paper, we adopt the feature extraction technique employed by GDB-ICP [53]. A summary of the technique is provided for completeness. More detailed description and validation for the selected feature extraction method can be found in [53].

Two types of generic features are extracted: corner points and edge points. Both types of features are extracted in a Gaussian-smoothed scale space. In this scale space, the smoothing scale starts with 1.0 for the 1st level and in general, the smoothing scale for the lth level is given as \( \sigma_l = 2^{0.5(l-1)} \). The image is also sub-sampled (one for every two pixels) at each odd level, i.e., \( l = \{3, 5, 7, \ldots\} \).

At each level with the given smoothing scale \( \sigma_l \), the autocorrelation matrix \( M(x) \) is computed for each pixel location \( x \), which is a weighted sum of the outer product of intensity gradient vectors over a neighborhood:

\[
M(x) = \sum_{y \in N(x)} w(x-y) \left( \nabla I(y) / |\nabla I(y)| \right)^T,
\]

where \( I(y) \) is the image intensity at location \( y \) and \( \nabla I(y) \) is the image gradient vector. The neighborhood \( N(x) \) is a square with length of \( 6\sigma_l \) (normalized properly according the sub-sampling) centered at the location \( x \). \( w \) is a 2D zero-mean Gaussian weight function with standard deviation \( \sigma_l \). Eigen decomposition is performed on this \( 2 \times 2 \) auto-correlation matrix, \( M(x) \). The point \( x \) is treated as a potential corner if the ratio of the two eigenvalues (smaller over the larger) is above the threshold, and as an edge point otherwise. A strength value is assigned to each point, \( m(x) = \text{trace}(M(x)) \), which is used in non-maximum suppression [36] and followed by the local pruning. During the step of local pruning, potential corner or edge points are eliminated if their strength values do not meet an adaptive threshold, which is generated from the robust median and median absolute deviation (MAD) statistics computed with a sliding window technique.

The remaining corner points and edge points are sorted separately in decreasing order by strength values. A corner point is selected if no other corner points with higher strength value have been selected in its suppressing neighborhood circle \(^1\). The remaining set, denoted as \( C \), contains matchable corner points. The procedure is repeated starting from the local pruning with higher thresholds and larger exclusion neighborhood to generate a set of driving corner points, denoted as \( C' \). \( C' \) is usually a subset of \( C \). The same procedure is applied to edge points to obtain the matchable edge points \( F \) and the driving edge points \( F' \). Figure 4 provides examples of driving and matchable feature sets from the same smoothing scale.

Compared to the feature extraction techniques in the literature for registration of retinal images [4], [8], [51], [54], this technique we adopted from the GDB-ICP algorithm is mostly effective for FA sequences because the combination of local

---

\(^1\)The radius of the suppressing circle is defined as the following: 3 pixels for matchable edge points, 5 for driving edge points, \( 3\sqrt{2} \) for matchable corner points, and \( 5\sqrt{2} \) for driving corner points. The \( \sqrt{2} \) factor is to compensate the difference between normal distance for edge points and Euclidean distance for corner points.
Table 1

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>Given an image dataset, for each image $I_p$, apply feature and keypoint extraction:</td>
</tr>
<tr>
<td></td>
<td>a) Extract edge point set $F_p$ and corner point set $C_p$ in scale space. (Section III-A).</td>
</tr>
<tr>
<td></td>
<td>b) Extract DoG keypoint sets $K_p$ from gradient-magnitude images $I_p^{G}$ in scale space. (Section III-B)</td>
</tr>
<tr>
<td></td>
<td>c) Compute SIFT descriptor for each keypoint also from $I_p^{G}$. (Section III-B)</td>
</tr>
<tr>
<td></td>
<td>d) Enrich each keypoint in $K_p$ with shape context descriptor using matchable edge points from $F_p$. (Section III-B)</td>
</tr>
<tr>
<td>2)</td>
<td>For each of the possible image pairs, $I_p$ and $I_q$, apply the pairwise registration:</td>
</tr>
<tr>
<td></td>
<td>a) Initialization: Match keypoints in $K_p$ to keypoints in $K_q$ and rank-order the matches based on the similarity score combining the SIFT and shape context descriptors. (Section III-B)</td>
</tr>
<tr>
<td></td>
<td>b) Refinement: For each of the keypoint match, apply the following steps until $n$ keypoint matches have been tried:</td>
</tr>
<tr>
<td></td>
<td>i) Generate the initial (locally-accurate) similarity transformation and the initial bootstrap region. (Section III-B).</td>
</tr>
<tr>
<td></td>
<td>ii) Perform Dual-Bootstrap refinement until convergence. (Section III-C)</td>
</tr>
<tr>
<td></td>
<td>iii) Apply the decision criteria. If all pass, terminate with success. (Section III-C)</td>
</tr>
<tr>
<td></td>
<td>c) Terminate with failure.</td>
</tr>
<tr>
<td>3)</td>
<td>Treating each image in the dataset as the anchor image in turn, apply the joint registration to ensure global consistency in the anchor space (Section III-D).</td>
</tr>
</tbody>
</table>

Fig. 3. Outline of the Edge-Driven DB-ICP registration algorithm.

Fig. 4. Examples of matchable and driving features from red-free (Figure 1(a)) and FA (Figure 1(b)) images of the original resolution, 1024 x 1024. Circles indicate corners and lines indicate the normal directions of edge points. (a) Matchable features of the red-free image. (b) Matchable features of the FA image. (c) Driving features of the red-free image. (d) Driving features of the FA image. Note that these features are extracted from the level where $\sigma_t = 1$ and that other levels are not shown.

B. Initialization

It is shown in [53] that the initialization process is the least robust component of the GDB-ICP algorithm—the SIFT local descriptor is designed only for invariance to linear illumination changes [25]. It is even less robust for the image registration problem we are addressing in this paper, due to the severe non-linear intensity differences between FA phases as shown in Figure 1. To overcome this limitation, we modify the computation of keypoint generation to achieve higher success on keypoint matching for such image pairs.

Vessels, which are darker than the background in the RF images, have various levels of brightness in different stages of a FA sequence as the fluorescein circulates through the vasculature. What are preserved under the intensity changes are the edge structures of the vasculature. To align two images based on edge similarity, we make use of the gradient magnitude image in place of the original image where pixel values are needed in the computation. In doing so, we also avoid the need for intensity reversal which is a common step for registration of retinal images [54], [22], [48].

For the choice of keypoint detector, it is reported in [53] that affine-invariant multiscale Harris corners [32] and Lowe’s multiscale keypoint detector [25] perform similarly for GDB-ICP. We chose the latter since a Lowe’s keypoint match is sufficient to constrain a similarity transformation, which is a good local approximation of the actual transformation for retinal imaging\(^2\). The location of a Lowe’s keypoint is detected as the extremum in both spatial and scale dimensions in Difference-of-Gaussian (DoG) scale space, with the scale increasing in one third octave steps ($1, 2^{\frac{1}{3}}, 2^{\frac{2}{3}}, 2, \ldots$). Associated with each keypoint are the scale $\sigma_k$ at which the keypoint is detected, and the orientation $\theta_k$ determined from the distribution of local image gradients. This defines a local 2D coordinate system for a keypoint in which the local image region is described. The computation of the keypoint descriptor is performed relative to the location, scale, and orientation of the keypoint to provide

\(^2\)It is rare to observe non-proportional scaling and/or shearing present in retinal images.
invariance to similarity transformation.

In the experimental evaluation of [31], SIFT has proven to be the most effective local descriptor. However, the SIFT descriptor is designed for images with rich texture, such as outdoor images; it does not work well for images which are relatively textureless, such as retinal images, because many keypoints may have local neighborhoods of similar appearance. To increase the distinctiveness of keypoints, our descriptor is a vector consisting of a SIFT descriptor [25] and a shape context descriptor [2]: \( D = [\mathcal{L}^T \quad \mathcal{G}^T]^T \) where \( \mathcal{L} \) is the SIFT descriptor characterizing the local neighborhood, and \( \mathcal{G} \) the shape context descriptor computed from a larger neighborhood to disambiguate keypoints with similar local appearance.

To compute the SIFT descriptor of a keypoint, as presented in [25], gradient-magnitudes of pixels within a \( 16 \times 16 \) neighborhood are collected in histograms of orientations, using \( \sigma_k \) to select the level of Gaussian blur for the image. The neighborhood is partitioned into \( 4 \times 4 \) subregions, each having \( 4 \times 4 \) pixels. The gradient directions, rotated relative to \( \theta_k \), of the pixels in a subregion is summarized in a histogram of 8 orientation bins. The contribution of each gradient vector is weighted by its magnitude and a Gaussian function centered at the keypoint with \( \sigma_e = 0.75 \sigma_k \). The Gaussian weighting is to downgrade the contribution of points further away from the keypoint. Combining all the histogram entries, the descriptor is a vector of \( 4 \times 4 \times 8 = 128 \) elements. In [25], the descriptors are extracted from the intensity image. For images of a FA sequence, the non-linear, local intensity differences result in changes in the gradient directions of the neighboring pixels of a keypoint, which in turn reduce the repeatability of the descriptors. To maintain the consistency of the gradient directions of corresponding pixels between two images, we compute the SIFT descriptors from the gradient magnitude image instead.

For the computation of the shape context component, we have made several modifications to the algorithm described in [2], which is proposed for unoccluded shape matching. Centered at a keypoint location \( p_k = (x_f, y_f) \) and oriented by \( \theta_k \) (for rotation invariance), the circular neighborhood for the shape context is divided into \( 5 \times 12 \) log-polar bins [34]. The log-polar histogram records the number of matchable edge points (extracted as described in section III-A) falling into each bin. Given an edge point \( p = (x, y) \) in the circular neighborhood of \( p_k \), the angular bin index is \( a = \lfloor \frac{\pi}{8} \arctan \left( \frac{y - y_f}{x - x_f} \right) - \theta_k \rfloor \). To determine the radial bin index, the radial distance of the point to the keypoint location is first normalized by the mean distance, \( N_d \), of all points in a neighborhood with the radius equal \( \beta \times \sigma_k \) to provide scale-invariance, where \( \beta \) is empirically determined and set to 16 for all our experiments\(^7\). Only points within \( 2N_d \) are considered. 4 radial bins equally divide the logarithmic range between \( \ln 2 \) and \( \ln \frac{1}{8} \) with the radial bin size being \( s = \frac{\ln 2 - \ln(\frac{1}{8})}{4} \). The

\[ r = \begin{cases} \left\lfloor \frac{\ln d(p) - \ln(\frac{1}{8})}{s} \right\rfloor, & \text{if } d(p) > \frac{1}{8} \\ 0, & \text{otherwise} \end{cases} \]

where \( d(p) \) is the normalized distance. Figure 5 shows the bin arrangement for the shape context. The central bins are smaller than those in the periphery, making the descriptor more sensitive to edge points near the keypoint than to those further away. To reduce this effect, the contribution of each edge point \( p \) is the inverted Gaussian weight [34],

\[ w(x, y) = 1 - e^{-\frac{(x-x_f)^2 + (y-y_f)^2}{2\sigma^2}}, \]

In doing so, edge points beyond the SIFT neighborhood are more effective in the shape context descriptor. Another advantage for using an inverted Gaussian is to downgrade the contribution of the very small bins toward the center where a slight shift in the edge location can potentially result in a very different angular bin number. Finally, the value of a bin is normalized by its area to get the density value. The log-polar histogram is further normalized to unit magnitude. To avoid boundary effect as edge points shift between bins, bilinear interpolation is used to distribute the weight to adjacent histogram bins.

Keypoint matching between two images is performed by taking keypoints from one image and matching the keypoints in another image using the descriptors, each having 188 components (128 from the SIFT and 60 from the shape

\(^7\) The majority of the image pairs in our experiment succeeded with keypoint matches with the scale equal to 4, which made the radius approximately 64 pixels.
context). Same as in [34], the distance is computed differently for the two kinds of descriptors. Given the descriptors from two images, $D_p = [\mathbf{L}_p^T \mathbf{G}_p^T]^T$ and $D_q = [\mathbf{L}_q^T \mathbf{G}_q^T]^T$, the distance metric is the Euclidean distance
$$d_C = |\mathbf{L}_p - \mathbf{L}_q| = \sqrt{\Sigma_k (\mathbf{L}_{p,k} - \mathbf{L}_{q,k})^2}$$
for the SIFT descriptor, whereas it is the $\chi^2$ statistic
$$d_G = \frac{1}{2} \Sigma_k (\mathbf{G}_{p,k} - \mathbf{G}_{q,k})^2$$
for the shape context descriptor as proposed in [2]. Both distances have equal contributions to the final distance, which is simply the sum of $d_C$ and $d_G$, and the match with the shortest distance is found for each keypoint. Keypoint matches are sorted with increasing distance, and the top 50 are tested in order. The locations, scales, and orientations (from the Lowe keypoint) of a keypoint match are sufficient to constrain an initial similarity transformation in a local region (the initial bootstrap region) in each image. The region is a square centered at each keypoint location, with half width $30 + 3\sigma_q$. The initial transformation and the initial region are the input to the Dual-Bootstrap refinement.

C. Dual-Bootstrap Refinement

The refinement process is driven by iterative alignment of the corner points and the edge points. This work is adopted from GDB-ICP. In [53], it was shown that Dual-Bootstrap’s refinement process correctly refined 80% of the initial transformations computed from correct keypoint matches and that it has a 99% success rate given three correct keypoint matches. We’ll provide a brief summary of the work for completeness, and readers are referred to [53] for details.

Given an initial similarity transformation from a keypoint match and an initial bootstrap region surrounding the keypoint location in each image, the algorithm iterates four steps: (1) bi-directional closest point matching, (2) applying a robust M-estimator to estimate the parameters of the candidate transformation models, (3) applying model selection to determine if a more appropriate transformation model may be used, and (4) expanding the bootstrap region based on the uncertainty of the mapping, which is computed from the covariance matrix of the current transformation estimate. The final transformation is then tested for correctness when the bootstrap region covers the entire overlap region between the two images. The four steps are illustrated in details as follows.

A match is generated by mapping a driving feature in $I_p$ to $I_q$ and finding the closest point in $I_q$. Recall the set of driving features is a sparse subset of the matchable features. It is more efficient to map and match using the driving features. Meanwhile, the dense matchable features ensure features are always available to be matched. GDB-ICP uses bi-directional matching to produce more constraints; matches are also generated by mapping driving features in $I_q$ and matching to the matchable features in $I_p$. This bi-directional matching has been proven to improve the numerical stability of estimates, especially for small bootstrap regions. As a result of bi-directional matching, both forward and backward transformations are estimated.

The model hierarchy appropriate for our application consists ofable to improve the numerical stability of the matchable features in

Joint registration serves two purposes in our application. Firstly, if viewing the images as nodes and successful pairwise registration between two images as the edge connecting the nodes in a graph setting, joint registration turns a connected

D. Joint Registration

Joint registration serves two purposes in our application. Firstly, if viewing the images as nodes and successful pairwise registration between two images as the edge connecting the nodes in a graph setting, joint registration turns a connected
graph into a fully connected graph. In other words, image pairs, including those that overlap but fail pairwise registration and those that do not overlap at all, can obtain transformations from “traversing” the constraints of other successful pairwise registration results. Secondly, joint registration aligns a set of images in a manner that is globally consistent when simultaneously transformed into one common coordinate system. This coordinate system is established the same as a chosen image in the sequence, which is named the anchor image. Without global consistency, two overlapping, non-anchor images, each independently registered with the anchor image, are not constrained in the region outside the anchor image and may appear to have misalignments when transformed onto the anchor image. Both properties are important for registration of a complete FA sequence, since our ultimate goal is to be able to register every image to any image as the reference image (the anchor image) with global consistency in an FA sequence.

Any algorithm which satisfies the aforementioned two properties can be used for our system. A common technique is to generate the transformation between any image to the chosen anchor image by cascading the transformation parameters [8], [14], [21]. This approach is mathematically infeasible for our problem since our transformation model is not closed under composition [6]. To construct globally consistent mosaics, we adopt the algorithm by Can et al. [5], which is similar in many respects to the important work of Sawhney et al. [43] and of Shum and Szeliski [47]. The advantage of the chosen joint algorithm over other similar algorithms is its incremental and graph-based nature by efficiently determining which images to match pairwise without exploring all $O(N^2)$ pairs where $N$ is the number of images in the sequence. The algorithm also makes no assumption on the ordering imposed by a video sequence. However, since one major goal of the paper is to study the effect of our novel initialization technique for pairwise registration, we initialize the joint algorithm with pre-computed pairwise transformations from our experiments.

Given a designated anchor image, $I_a$, which defines the reference space, the joint registration process generates a set of transformations $\Theta_{p,a}$, each transforming image $I_p$ to $I_a$, using all edge point matches resulting from successful pairwise alignment. A match $(p,q)$ from the pairwise alignment of two images $I_p$ and $I_q$ constrains transformations onto $I_a$ in one of two forms. If neither $I_p$ nor $I_q$ is the anchor, the transformations $\Theta_{p,a}$ and $\Theta_{q,a}$ are constrained by the residual term $(X(p) - X(q) - \Theta_{p,a}q)\eta_q$, where $\eta_q$ is the transformed normal direction at $q$. If one image, say $I_q$, is the anchor, then the mapping of $I_p$ onto $I_a$ is constrained by the residual term $(X(p) - \Theta_{p,a}q)\eta_q$, as in the case for pairwise registration. The objective function here is the sum of all the constraints from the successful pairwise alignment. Minimizing the objective function, we obtain a set of transformations which jointly align the images on $I_a$. This is a standard least-squares problem, where the unknowns are the set of transformation parameters and the coefficients are computed from the edge point locations and normals. This process is repeated for each choice of $I_a$, resulting in globally consistent transformations for all images.

It is important to point out the dependency between pairwise and joint registration. The joint registration provides the missing transformation between two images if there exists a path between the two image nodes in the alignment graph defined above. However, the ultimate success of multi-image registration depends heavily on the robustness of pairwise registration—if an image fails to register with any other image in the set, it cannot be part of the joint registration.

### IV. Experimental Evaluation

Since there is not yet a public FA database that provides the common ground for validation of algorithms, we chose to evaluate our algorithm with the same dataset on which DB-ICP was tested on [50]. The dataset consists of 60 clinical angiogram sequences of a mixture of pathologies, including age-related macular degeneration, central retinal vein occlusion, and retinitis pigmentosa. These images were taken with a Topcon digital fundus camera of resolution $1024 \times 1024$. It is a common practice that one or two RF images are taken before the injection of the fluorescein. The number of images in each sequence (including the RF) ranges from 10 to 16. Counting all pairs within a sequence, there is a total of 4919 image pairs, out of which 1303 are RF-FA pairs, 3560 are FA-FA pairs, and 56 are RF-RF pairs.

We compare our algorithm with DB-ICP and GDB-ICP in performance on both pairwise and joint registration. The DB-ICP has been validated with a large suite of pathological FA sequences [50], and GDB-ICP has been made available to the public and tested with a variety of challenging images with high robustness. Please note, when running DB-ICP and GDB-ICP on inter-modal pairs, we inverted the FA images for higher success rates.

To assess the correctness of a 2D transformation, there are several common approaches in the literature [52]. For qualitative assessment, the simplest is fusion of the registered images—on which one image transformed to the space of another—displayed either in different colors or in a checkerboard pattern [22], [41], [54]. For quantitative assessment, the common approach is to ask specialists to provide a sufficient number of pairs of corresponding points in both images to constrain the transformation, which is treated as the ground-truth [8], [29], [28], [42]. This approach may allow objective comparison between algorithms if there exists a public dataset with known ground-truth transformations. Cooper et al. [10] proposed a validation technique using optical flow which works well only for images without much intensity change.

For this study we performed the clinical evaluation only on the joint registration for two reasons. First, for pairwise registration, we apply exactly the same correctness test as in GDB-ICP. Its ability to reject incorrect transformations has been demonstrated in [53] with a 99.8% success rate using 1722 image pairs (with 83,550 incorrect initializations). In fact, this test is so effective that we did not encounter any incorrect transformation during the development of our pairwise algorithm. Second, jointly registered sequences are

---

4This chosen image can be any image in the sequence. There is no advantage of choosing one over another.
the final products of our system for the physicians, but the clinical evaluation for the joint algorithm is not found in the original paper [5].

To clinically evaluate the joint results, a retina specialist was asked to pick one of the four categories after reviewing a registered sequence: (1) no movement visible, (2) localized vessel movements visible, but still acceptable for clinical purpose, (3) regional vessel movements visible, and might affect treatment, and (4) global distortion, rendering the sequence useless. 41 sequences fall in category 1 and 19 in category 2 with extremely localized misalignments in areas where edemas or scarrings are found. Such local misalignments are expected and they result from the modeling error—a surface with edemas or scarrin is no longer a quadratic surface. Although we did not perform the clinical evaluation on the pairwise registration, we still asked one graduate student (not one of the authors) to manually go over all the successful pairs to look for structural mis-alignment—if one side of a vessel of at least a few pixels in width is aligned with the other side of the corresponding vessel or anything beyond this scope. It is confirmed that all the successful pairs have correct transformations without any structural mis-alignment.

A. Component Analysis on Initialization

To isolate the effectiveness of each major component of our keypoint matching algorithm for initialization, we conducted the performance analysis on our method and six variations, three on the intensity image and three on the image of gradient magnitude, using the sequence containing images in Figure 1. Edge-Driven DB-ICP differs from the 6 variations in the way that both intensity and gradient magnitude images are involved for initialization—the Lowe’s keypoints and SIFT descriptors are computed from the gradient magnitude image and the edge points used in the shape context descriptor computation are computed from the intensity image. The complete sequence consists of 2 RF images and 10 FA images, with a total of 66 image pairs. For all variations, the multiscale Lowe’s keypoints are computed using the same image from which the descriptors are extracted. The three variations on keypoint description are:

1) Using only the SIFT descriptor.
2) Using only the shape context descriptor computed from edge points without inverse Gaussian weighting.
3) Combining the SIFT and shape context descriptors.

We compare the pairwise success rates and the accuracy of the six variations against our method. The success rate is defined as the number of image pairs with a transformation passing the decision criteria, stated in section III-C, out of total number of pairs. We also compare the success rate if using the first match only. The results are shown in Table I.

We would like to point out three observations here. Firstly, for both sets of experiments using the original intensity and gradient magnitude images, respectively, a descriptor combining SIFT and shape context outperforms the individual descriptor acting alone in terms of the overall success rate and the likelihood of succeeding on the first match. Secondly, comparing the choice of the intensity image versus gradient magnitude image, the improvement is 31.5% (69.7/53) overall, and 66.7% (75/45) for RF-FA pairs if using SIFT alone on the gradient magnitude image. Similar results are obtained with combined descriptors. Thirdly, it is not a surprise to see that the third experiment with gradient magnitude image has similar performance as the Edge-Driven DB-ICP. This is because edge structures are preserved between modalities and FA phases for the computation of shape context descriptors whether the edge points are extracted from the intensity or the gradient magnitude image. In this paper, we chose to use the intensity image to compute the shape context descriptor for the purpose of computation efficiency. Otherwise, edge-point extraction has to be carried out twice on both the intensity and the gradient magnitude images5. The higher alignment errors for the two experiments can be explained by the fact that more challenging pairs with substantial appearance differences (resulting in higher misalignment errors) were registered.

In addition to the testing of major components, we also performed a retrospective sensitivity analysis on the multiplier, \( \beta \), for \( N_d \), which defines the neighborhood for computation of the shape context descriptor. We studied values of \( \beta \) in the range of 4 to 64 using the same sequence as for component-wise testing. There was no change in terms of overall success rate, but \( \beta = 16 \) (the value in all our experiments) has the highest success rate on the first match, which can degrade more than 50% at the low end.

B. Performance Analysis on Pairwise Registration

First, let us define the notion of successful pairwise alignment. For Edge-Driven DB-ICP and GDB-ICP, the success is defined using the 3 decision criteria as defined in section III-C. For DB-ICP, a successful pairwise alignment has a weighted average error below 1.5 pixels [50]. The results are shown in Table II6. We are aware that for both Edge-Driven DB-ICP and GDB-ICP, the value of the threshold on the final alignment error doubles the value set for DB-ICP. However, this does not result in an unfair comparison with DB-ICP. First, errors are computed with different features—DB-ICP uses only vessel centerline points whereas the former two use edge points everywhere in the image. As a result, given the same transformation with inaccurate alignment in areas with sparse vessel distribution, DB-ICP returns a lower alignment error comparing to the other two algorithms. This justifies a lower threshold value for DB-ICP. Second, setting the threshold value to 3.0 pixels for the former two algorithms does not result in incorrect alignments, since a transformation with an error between 1.5 and 3 pixels has to meet the desired stability and consistency to be considered correct, whereas DB-ICP is safe-guarded solely by the accuracy measure, and if the threshold is lifted to 3 as for the other two algorithms, the transformations with errors between 1.5 and 3 are likely to be incorrect, as demonstrated in [6].

In terms of the pairwise success rate, Edge-Driven DB-ICP outperforms other two algorithms—our algorithm has the

---

5GDB-ICP refinement is always carried out with the intensity image as it was described in [53].
6Because both our algorithm and GDB-ICP perform bi-directional matching, the ordering of the two images of a pair becomes irrelevant. Thus we only test one ordering for any image pair.
highest success rate of the three (see Table II). For FA-FA pairs, DB-ICP is limited by its feature extraction capability, which identifies salient vessels. Such structures could be obscured by lesions or are less prominent in images of late phases due to low contrast. Meanwhile, Edge-Driven DB-ICP doubles the success rate of GDB-ICP in registration of RF-FA pairs (81.1% vs. 40.1%), because the latter is limited by initialization using Lowe keypoints, which was not designed for images with substantial non-linear intensity differences. 

Figure 6 shows image pairs for which our algorithm succeeded but either or both of the other two failed. Figure 7 demonstrates the accurate alignment for image pairs shown in Figure 6. Using checkerboard mosaics, it helps to visualize the appearance differences between images. In terms of accuracy, the slightly higher numbers for our algorithm can be explained by the fact that it succeeded on more challenging image pairs. However, since the statistics are affected by the thresholds set on the alignment errors, they are less significant for our study.

The effect of our more robust initialization technique can also be realized by the number of pairs which succeed with the first keypoint match. Table III shows the comparison with GDB-ICP of the percentages of successful pairs using the top $n$ keypoint matches. Our algorithm has higher success on the first keypoint match alone than the cumulative success rate for GDB-ICP on the top 5 keypoint matches. This also helps to reduce the execution time of the joint registration substantially on a complete FA sequence. Running on a computer with AMD Athlon 64 Processor and 1G RAM, the average execution time for an image pair that succeeds on the first initial transformation is about 35 seconds—6 seconds for keypoint extraction, 2 seconds for extraction of multi-scale features, 2 seconds for keypoint matching, and 25 seconds for refinement initialized from one successful keypoint match.

C. Performance Analysis on Joint Registration

The goal of our work is to be able to register all images in a complete FA sequence. To define a measure for performance, we first asked a physician to label FA images of 5 FA stages: arterial, arteriovenous, venous, late-venous, and recirculation. The stage label of an image allows us to identify the earliest stage in which the algorithm fails for the specific sequence. We define the success rate for a specific phase as the percentage of sequences for which joint registration successfully aligned all images in the sequence up to and including the phase.

Table IV is the summary of the performance on complete FA sequences. Edge-Driven DB-ICP only failed to register one of the 60 pathological sequences. For both Edge-Driven DB-ICP.
and GDB-ICP, if RF images are excluded from the sequences, the success rates are 100% for all phases. Figure 8 and Figure 9 show the mosaics resulting from joint registration of the FA sequences from which images in Figure 1 and Figure 6 are extracted, respectively.

For the one failed sequence, our algorithm failed to register any of the multimodal image pairs. The sequence turned out to be a case of retinitis pigmentosa. To analyze the cause of failure, we examined all keypoint matches in multimodal image pairs in search of one keypoint match that could lead to a good initialization. We found one good keypoint match as shown in Figure 10. The keypoint match was ranked the 15-th. A zoomed-in view of the match is shown in Figure 5. Figure 10(e) is the checkerboard mosaic of the initial alignment, and no significant misalignment is detected. Using this initial alignment, the refinement process failed to converge, and the cause of failure was lack of sufficient consistent features to drive the correct alignment. As shown in Figure 10(d), except in the region of the optic disk, vessels are ill-represented by edge points in the FA image. The two images also highlight different pathologies in the background due to the disease.

We have also estimated the accuracy of the joint registration. In order to estimate the average alignment error using only reliable feature correspondences, we only include correspondences generated from image pairs with successful pairwise registration. The accuracy summary is also shown in Table IV. The higher average alignment errors for joint registration (compared to pairwise registration) can partially be attributed to the modeling error—the quadratic transformation is an approximation for the complex imaging system—and partially to the fact the lower pairwise alignment error for an image pair is compromised to ensure global consistency with other images that overlap with them. The latter can be explained by inconsistency in feature extraction which can be affected by the flow of the fluorescein. Again, Edge-Driven DB-ICP has slightly higher alignment error compared to other two algorithms because it succeeded on more complete sequences with more challenging image pairs.

V. DISCUSSION AND CONCLUSION

We have presented a robust registration algorithm for alignment of images in a complete FA sequence containing a mixture of RF and FA images of various phases through the dye circulation. Different from many existing techniques in the literature, our algorithm demands no prior knowledge of the image modality or the FA phase. This enables a fully-automatic registration system for a complete FA sequence.

Difficulties for registration of a complete FA sequence include substantial, non-linear intensity differences and low contrast. Extending the work of Generalized Dual-Bootstrap ICP (GDB-ICP), our algorithm further exploited the structural similarity between images for initialization—Lowe keypoints and their descriptors are extracted from the gradient magnitude image, and the descriptors are further enriched with global-shape context using the edge points to increase the distinctiveness of a keypoint. As in GDB-ICP, initialization is done by keypoint matching, and only uses individual matches to generate transformation estimates accurate in small image regions for dual-bootstrapping refinement.

The effectiveness of the algorithm is demonstrated by the results on a suite of 60 randomly selected pathological FA sequences including diseases of age-related macular degeneration, central retinal vein occlusion, and diabetic retinopathy. It achieved an overall success rate of 92.4% for pairwise registration with an average alignment error of 0.94 pixel. 59 out of 60 sequences are jointly aligned.

The only sequence which our algorithm failed to jointly align has substantial structural changes between modalities due to the nature of the disease. This results in a high percentage of outliers in matching, which drives the alignment away from the correct estimate even if it started with a good initialization. Using vessel centerlines only, DB-ICP succeeded with many multimodal pairs of this sequence if the FA images were inverted. This suggests a possible direction of improvement.
for our algorithm by considering a weighting scheme on the edge features using a measure of connectivity in the features in a local neighborhood [15].

GDB-ICP has demonstrated high success with a wide variety of image pairs. However, its weakness lies in initialization. Our proposed algorithm improved the initialization step of GDB-ICP for image pairs with substantial, non-linear intensity difference with an overall improvement of 15.4% and 22.9% for pairwise and joint registration, respectively. Like GDB-ICP, using generic features for registration, our algorithm is widely applicable to applications of multi-modal image registration outside the retinal imaging domain.

VI. ACKNOWLEDGMENTS

The authors would like to acknowledge the generous help from the Vision Lab in the Department of Computer Science of Rensselaer Polytechnic Institute. They provided us the clinical data for testing and made the GDB-ICP executable available to the public. We are grateful to Dr. Tse-Hsiang Yeh from Taipei Veterans General Hospital for his most generous help on validation.

REFERENCES


Fig. 7. Checkerboard mosaics of image pairs in Figure 6. Notice the accurate alignment and also the significant amount of difference in appearance. Alignment errors are: 2.1, 2.2, 1.8, and 2.2 pixels, respectively.
Fig. 8. A joint mosaic of the FA sequence containing images in Figure 1.

Fig. 9. Joint mosaics of FA sequences from which the second and the forth image pairs, respectively, in Figure 6 are extracted.

Fig. 10. Failure analysis. The image pair, image (a) and (b), is from the only sequence in our test set which can not be registered completely, since all multimodal pairs failed. The red crosses indicating the corresponding keypoint locations of a keypoint match generated by our algorithm. A zoomed-in view of the keypoint match can be found in Figure 5. Image (c) and (d) show the edge points extracted in the neighborhoods of the keypoint locations for Dual-Bootstrap refinement which failed to converge from initial alignment shown in Image (e).