Published in final edited form as:
IEEE Trans Biomed Eng. 2010 July ; 57(7): 1707–1718. doi:10.1109/TBME.2010.2042169.

A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration

Jian Chen
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA. He is now with IBM China Research Laboratory, Beijing 100027, China (jianchen@cn.ibm.com)

Jie Tian*[Fellow, IEEE]
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Noah Lee
Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (nl2168@columbia.edu)

Jian Zheng
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (zhengjian@fingerpass.net.cn)

R. Theodore Smith
Retinal Image Analysis Laboratory, Edward S. Harkness Eye Institute and the Department of Ophthalmology, Columbia University, New York, NY 10027 USA (rts1@columbia.edu)

Andrew F. Laine[Fellow, IEEE]
Heffner Biomedical Imaging Laboratory, Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (laine@columbia.edu)

Abstract

Detection of vascular bifurcations is a challenging task in multimodal retinal image registration. Existing algorithms based on bifurcations usually fail in correctly aligning poor quality retinal image pairs. To solve this problem, we propose a novel highly distinctive local feature descriptor named partial intensity invariant feature descriptor (PIIFD) and describe a robust automatic retinal image registration framework named Harris-PIIFD. PIIFD is invariant to image rotation, partially invariant to image intensity, affine transformation, and viewpoint/perspective change. Our Harris-PIIFD framework consists of four steps. First, corner points are used as control point candidates instead of bifurcations since corner points are sufficient and uniformly distributed across the image domain. Second, PIIFDs are extracted for all corner points, and a bilateral matching technique is applied to identify corresponding PIIFDs matches between image pairs. Third, incorrect matches are removed and inaccurate matches are refined. Finally, an adaptive transformation is used to register the image pairs. PIIFD is so distinctive that it can be correctly identified even in nonvascular areas. When tested on 168 pairs of multimodal retinal images, the Harris-PIIFD far outperforms existing algorithms in terms of robustness, accuracy, and computational efficiency.

* tian@ieee.org .

Keywords

Harris detector; local feature; multimodal registration; partial intensity invariance; retinal images

I. Introduction

THE PURPOSE of retinal image registration is to spatially align two or more retinal images for clinical review of disease progression. These images come from different screening events and are usually taken at different times or different fields of view. An accurate registration is helpful to diagnose various kinds of retinal diseases such as glaucoma, diabetes, and age-related macular degeneration [1]–[4], [54]. However, automatic accurate registration becomes a problem when registering poor quality multimodal retinal images (severely affected by noise or pathology). For example, it is difficult to register an image pair taken years apart, which were acquired with different sensors due to possible differences in the field of view and modality characteristics [5]–[8]. Retinopathy may cause severe changes in the appearance of the whole retina such as obscure vasculature patterns (see Fig. 1). Registration algorithms that rely on vascular information may fail to correctly align such image pairs.

Thus, in this paper, we propose a novel distinctive partial intensity invariant feature descriptor (PIIFD) and describe a fully automatic algorithm to register poor quality multimodal retinal image pairs. In the following, we will first briefly introduce prior work regarding existing retinal registration algorithms and then propose our Harris-PIIFD framework.

A. Prior Work

Existing registration algorithms can be classified as area-based and feature-based approaches [9]–[11]. The area-based approaches [12]–[24] compare and match the intensity differences of an image pair under a similarity metric such as mutual information [19]–[22] and cross correlation [12], [15], and then apply an optimization technique [23], [24] to maximize the similarity metric by searching in the transformation space. The similarity metric is expected to reach its optimum when two images are properly registered. However, in the case of low overlapping area registration, the area-based approaches usually fail [55]. In other words, the similarity metric is usually misled by nonoverlapping areas. To overcome this problem, a widely used solution is to assign a region of interest (ROI) within one or both images for computing the similarity metric [24]. The area-based approaches are also sensitive to illumination changes and significant initial-misalignment, suggesting that area-based approaches may be susceptible to occlusion, background changes caused by pathologies, and pose changes of the camera [35].

Compared with area-based registration, feature-based approaches [25]–[41] are more appropriate for retinal image registration. Feature-based approaches typically involve extracting features and searching for a transformation, which optimizes the correspondence between these features. The bifurcations of retinal vasculature, optic disc, and fovea [27], [28] are examples of such widely used feature cues, respectively. The main advantage of feature-based approaches is the robustness against illumination changes. However, extraction of such features in poor quality images is difficult. Feature-based approaches for retinal image registration usually distinguish themselves through minor differences and rely on the assumption that vasculature network is able to be extracted. For instance, the use of different structures of the retina as landmark points [27], a focus on improving the performance of landmark extraction algorithm [36], a narrowing down of the search space by manually or automatically assigning “matched” points [30], and a more complicated

mapping strategy to estimate the most plausible transformation from a pool of possible landmark matches [26] have been described and all of them rely on the extraction of retinal vasculature.

A hybrid approach that effectively combines both area-based and feature-based approaches has also been proposed [55]; however, it still relies on retinal vasculature.

General feature-based approaches that do not rely on vasculature are also discussed. Scale invariant feature transform (SIFT), an algorithm for extracting distinctive invariant features has been proposed [42]–[46]. The SIFT features proposed in this algorithm are invariant to image scale and rotation and provide robust matching across a substantial range of affine distortion, change in 3-D viewpoint/perspective, addition of noise, and changes in illumination. These features are highly distinctive in a sense that a single feature can be correctly matched with high probability against a large database of features from many images. However, SIFT is designed for monomodal image registration, and its scale invariance strategy usually cannot provide sufficient control points for high order transformations. Another local feature named speeded up robust features (SURF) [57] has also been proposed, which is several times faster and more robust against different image transformations than SIFT claimed by its authors. SURF is based on Haar wavelet, and its good performance is achieved by building on the strengths of SIFT and simplifying SIFT to the essential [57]. Soon after, a SURF-based retinal image registration method, which does not depend on vasculature has been proposed [59]; however, it is still only applicable for monomodal image registration.

General dual bootstrap iterative closest point algorithm (GDB-ICP) [35], [60] which uses “corner” points and “face” points as correspondence cues is more efficient than other existing algorithms. To our knowledge, the GDB-ICP algorithm is the best algorithm reported for poor quality retinal image registration. There are two versions of this approach. The first version uses Lowe's multiscale keypoint detector and the SIFT descriptor [42]–[46] to provide initial matches. In comparison, the second version uses the central line extraction algorithm [36] to extract the bifurcations of the vasculature to provide initial matches. Then GDB-ICP algorithm is applied to iteratively expand the area around initial matches by mapping the “corner” or “face” points. The authors declare that only one correct initial match is enough for subsequent iterative registering process. However, in some extreme cases no correct match can be detected by their two initial matching methods. Further, for very poor quality images, even if there are some correct initial matches, the GDB-ICP algorithm may still fail because the distribution of “corner” and “face” points are severely affected by noise.

B. Problem Statement and Proposed Method

As mentioned earlier, the existing algorithms cannot register poor quality multimodal image pairs in which the vasculature is severely affected by noise or artifacts. The retinal image registration can be broken down to two situations: multimodal image registration and poor quality image registration. The existing algorithms can achieve good performance when these two situations are not combined together. On one hand, vasculature-based registration methods can correctly align good-quality multimodal retinal image pairs. On the other hand, some robust local features such as SIFT and SURF can achieve satisfactory results for poor quality monomodal registration. However, it is hard to register poor quality multimodal retinal images. An illustration of retinal image registration combined these two situations is shown in Fig. 2, in which two images are of poor quality and different modalities.

A robust local feature descriptor may bring to success the registration of poor quality multimodal retinal images, as long as it solves the following two problems: 1) the gradient

orientations at corresponding locations in multimodal images may point to opposite directions and the gradient magnitudes usually change. Thus, how can a local feature achieve intensity invariance or at least partial intensity invariance? and 2) the main orientations of corresponding control points in multimodal images usually point to the opposite directions supposing that two images are properly registered. How can a local feature achieve rotation invariance?

In this paper, we propose a novel highly distinctive local feature descriptor named PIIFD [58] and describe a robust automatic retinal image registration framework named Harris-PIIFD to solve the aforementioned registration problem. PIIFD is invariant to image rotation, partially invariant to image intensity, affine transformation, and viewpoint/perspective change. Note that PIIFD is a hybrid area-feature descriptor since the area-based structural outline is transformed to a feature-vector.

The remainder of this paper is organized as mentioned in the following. Section II is devoted to the proposed Harris-PIIFD framework including the novel PIIFD feature descriptor. Section III describes the experimental settings and reports the experimental results. Discussion and conclusion are given in Section IV.

II. Proposed Registration Framework

Our suggested Harris-PIIFD framework comprises the following seven distinct steps.

1) Detect corner points by a Harris detector [47].

2) Assign a main orientation for each corner point.

3) Extract the PIIFD surrounding each corner point.

4) Match the PIIFDs with bilateral matching.

5) Remove any incorrect matches.

6) Refine the locations of each match.

7) Select the transformation mode.

The flowchart of the Harris-PIIFD framework is shown in Fig. 3. First, corner points are used as control point candidates instead of bifurcations (step 1) since corner points are sufficient and uniformly distributed across the image domain. We assume that there are two subsets of control point candidates, which could be identically matched across two images. Second, PIIFDs are extracted relative to the main orientations of control point candidates therefore achieve invariance to image rotation, and a bilateral matching technique is applied to identify corresponding PIIFDs matches between image pairs (steps 2–4). Third, incorrect matches are removed and inaccurate matches are refined (steps 5–6). Finally, an adaptive transformation is applied to register the image pairs based on these matched control point candidates (step 7).

Three preprocessing operations are applied before detecting control point candidates: 1) convert the input image format to grayscale; 2) scale the intensities of the input image to the full 8-bit intensity range [0, 255]; and 3) zoom out or in the image to a fixed size (about 1000 × 1000 pixels, in this paper). The third operation is not necessary but has twofold advantages: 1) some image-size-sensitive parameters can be hold fixed and 2) the scale difference can be reduced in some cases.

A. Detect Corner Points by Harris Detector

The lack of control points is likely to result in an unsuccessful registration for a feature-based algorithm. In retinal image registration, bifurcations are usually regarded as control point candidates. However, it is hard to extract the bifurcations in some cases, especially in poor quality retinal images. Take the image in Fig. 1(b), for example, only four bifurcations are detected by a central line extraction algorithm [see Fig. 4(a)] [36]. On the contrary, a large number of Harris corners are detected and uniformly distributed across the image domain [see Fig. 4(b)]. Therefore, we introduce Harris detector [47] to generate control point candidates in our registration framework. The basic concept of the Harris detector is to measure the changes in all directions when convoluted with a Gaussian window, and the changes can be represented by image gradients. For an image I, assume the traditional image gradients are given as follows:

$$\begin{bmatrix} G_{xt} \\ G_{yt} \end{bmatrix} = \begin{bmatrix} \partial I/\partial x \\ \partial I/\partial y \end{bmatrix}. \tag{1}$$

Thus, the Harris detector can be mathematically expressed as

$$M = \begin{bmatrix} G_{xt}^2 & G_{xt}G_{yt} \\ G_{yt}G_{xt} & G_{yt}^2 \end{bmatrix} * h \tag{2}$$

$$R = \det(M) - k\,\mathrm{tr}^2(M) \tag{3}$$

where h is a Gaussian window, k is a constant (usually $k = 0.04 \sim 0.06$ [47]), and det and tr are the determinant and trace of the matrix, respectively. Given a point $p(x, y)$, it is considered as a corner point if and only if $R(p) > 0$. For more details about Harris detector please refer to [47].

Extracting the PIIFDs is the most time-consuming stage of the proposed Harris-PIIFD framework, and its runtime is directly proportional to the number of corner points (control point candidates). It has been confirmed that 200 Harris corner points are sufficient for subsequent processing, thus, in our experiments about 200 Harris corner points are detected by automatically tuning the sigma of Gaussian window.

The corner points in our framework are not directly used as features for the registration algorithm. Instead, they just provide the locations for calculating PIIFDs. Thus, the proposed method can still work if these corner points are disturbed in the neighborhood or even be replaced by a set of randomly distributed points. The only difference may be a change in accuracy.

B. Assign Main Orientation to Each Corner Point

A main orientation that is relative to the local gradient is assigned to each control point candidate before extracting the PIIFD. Thus, the PIIFD can be represented relative to this orientation and therefore achieve invariance to image rotation. In the present study, we introduce a continuous method, average squared gradients [48], [49], to assign the main orientation. This method uses the averaged perpendicular direction of gradient which is limited within $[0,\pi)$ to represent a control point candidate's main orientation. For image I, the new gradient $[G_x G_y]^T$ is expressed as follows:

$$\begin{bmatrix} G_x \\ G_y \end{bmatrix} = \operatorname{sgn}(G_{yt}) \begin{bmatrix} G_{xt} \\ G_{yt} \end{bmatrix} \tag{4}$$

where G_{xt} and G_{yt} are the traditional gradients defined in (1). In this equation, the second element of the gradient vector is always positive for the reason that opposite directions of gradients indicate equivalent main orientations. To compute the main orientation, the image gradients should be averaged or accumulated within an image window. Opposite gradients will cancel each other if they are directly averaged or accumulated, but they are supposed to reinforce each other because they indicate the same main orientation. A solution to this problem is to square the gradient vector in complex domain before averaging. The squared gradient vector $[G_{s,x}G_{s,y}]^T$ is given by

$$\begin{bmatrix} G_{s,x} \\ G_{s,y} \end{bmatrix} = \begin{bmatrix} G_x^2 - G_y^2 \\ 2G_xG_y \end{bmatrix}. \tag{5}$$

Next, the average squared gradient $\left[\bar{G}_{s,x} \quad \bar{G}_{s,y}\right]^T$ is calculated within a Gaussian-weighted circular window

$$\begin{bmatrix} \bar{G}_{s,x} \\ \bar{G}_{s,y} \end{bmatrix} = \begin{bmatrix} G_{s,x} * h_\sigma \\ G_{s,y} * h_\sigma \end{bmatrix} \tag{6}$$

where h_σ is the Gaussian-weighted kernel, and the operator * means convolution. The σ of the Gaussian window can neither be too small nor too big, for the reason that the average orientation computed in a small window is sensitive to noise and in a large window cannot represent the local orientation. In this study, the σ of Gaussian window is set to five pixels empirically.

The main orientation ϕ of each neighborhood with $0 \le \phi < \pi$ is given by

$$\phi = \frac{1}{2} \begin{cases} \tan^{-1}\left(\bar{G}_{s,y} / \bar{G}_{s,x}\right) + \pi, & \bar{G}_{s,x} \ge 0 \\ \tan^{-1}\left(\bar{G}_{s,y} / \bar{G}_{s,x}\right) + 2\pi, & \bar{G}_{s,x} < 0 \cap \bar{G}_{s,y} \ge 0 \\ \tan^{-1}\left(\bar{G}_{s,y} / \bar{G}_{s,x}\right), & \bar{G}_{s,x} < 0 \cap \bar{G}_{s,y} < 0 \end{cases} . \tag{7}$$

Thus, for each control point candidate $p(x, y)$,its main orientation is assigned to $\phi(x, y)$.

The SIFT algorithm uses an *orientation histogram* to calculate the main orientation [42]. However, the main orientations in multimodal images calculated by *orientation histogram* may direct to unrelated directions. This may result in many incorrect matches. In addition, the *orientation histogram* is discrete, suggesting that their directional resolution is related to the number of histogram bins. Compared with *orientation histogram*, our *averaging squared gradients* is continuous, more accurate and computational efficient. As long as the structural outlines are the same, the main orientations calculated by our method remain the same. Therefore, our method for calculating main orientation is suitable for multimodal image registration.

C. Extract PIIFD Surrounding Each Corner Point

Given the main orientation of each control point candidate (corner point extracted by Harris Detector), we can extract the local feature in a manner invariant to image rotation [42] and partially invariant to image intensity. As shown in Fig. 5(a), supposing the centered point is a control point candidate, and the big square which consists of 4 ×4 small squares is the local neighborhood surrounding this control point candidate. Note that the main orientation of this control point candidate is illustrated by the arrow. The size of neighborhood is a tradeoff between distinctiveness and computational efficiency. In Lowe's SIFT algorithm, the size of neighborhood is automatically decided by the scale of control point. By carefully investigating the retinal images, we empirically set the size to fixed 40 × 40 pixels in our experiments for the reason that the scale difference is slight.

To extract the PIIFD, the image gradient magnitudes and orientations are sampled in this local neighborhood. In order to achieve orientation invariance, the gradient orientations are rotated relative to the main orientation. For a given small square in this neighborhood [e.g., the highlighted small square shown in Fig. 5(a)], an orientation histogram, which evenly covers 0°–360° with 16 bins (0°, 22.5°, 45°, …, 337.5°) is formed. The gradient magnitude of each pixel that falls into this small square is accumulated to the corresponding histogram entry. It is important to avoid the boundary affects in which the descriptor abruptly changes as a sample shifts smoothly from being within one histogram to another or from one orientation to another. Therefore, bilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. The processes between extracting PIIFD and SIFT are almost the same, therefore, PIIFD and SIFT have some common characteristics. For example, both PIIFD and SIFT are partially invariant to affine transformation [42]–[46].

In an image, an outline is a line marking the multiple contours or boundaries of an object or a figure. The basic idea of achieving partial intensity invariance involves extracting the descriptor from the image outlines. This is based on the assumption that regions of similar anatomical structure in one image would correspond to regions in the other image that also consist of similar outlines (although probably different values to those of the first image). In this study, image outline extraction is simplified to extract the constrained image gradients. The gradient orientations at corresponding locations in multimodal images may possibly point to opposite directions and the gradient magnitudes usually change. In order to achieve partial intensity invariance, two operations are applied on the image gradients. First, we normalize the gradient magnitudes piecewise to reduce the influence of change of gradient magnitude. In a neighborhood surrounding each control point candidate, we normalize the first 20% strongest gradient magnitudes to 1, second 20% to 0.75, and by parity of reasoning the last 20% to 0. Second, we convert the orientation histogram with 16 bins to a degraded orientation histogram with only 8 bins (0°, 22.5°, 45°, …,157.5°) by calculating the sum of the opposite directions [see Fig. 5(b)]. If the intensities of this local neighborhood change between two image modalities (for instance, some dark vessels become bright), then the gradients in this area will also change. However, the outlines of this area will almost remain unchanged. The degraded orientation histogram constrains the gradient orientation from 0 to π, and then the histogram achieves invariance when the gradient orientation rotates by 180°. Consequently, the descriptor achieves partial invariance to the aforementioned intensity change. The second operation is based on the assumption that the gradient orientations at corresponding locations in multimodal images point to the same direction or opposite directions. It is difficult to mathematically prove this assumption as “multimodal image” is not a well-defined notation, although for intensity inverse images (an ideal situation), this assumption is absolutely sustainable. Actually, the degraded orientation histogram is not as distinctive as the original one, but this degradation at the cost of distinctiveness is acceptable for achieving partial invariance to image intensity. For the case shown in Fig. 5, there are in

total 4×4 = 16 orientation histograms (one for each small square). All these histograms can be denoted by

$$H=\begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \\ H_{41} & H_{42} & H_{43} & H_{44} \end{bmatrix} \tag{8}$$

where H_{ij} denotes an orientation histogram with eight bins.

The main orientations of corresponding control points may point to the opposite directions in multimodal image pair. This situation will still occur even we have already constrained the gradient orientations to the range [0°, 180°], and break the rotation invariance. For example, the main orientations of corresponding control points extracted from an image and its rotated version by 180° always point to the opposite directions. In this paper, we propose a linear combination of two subdescriptors to solve this problem. One subdescriptor is the matrix H computed by (8). The other subdescriptor is a rotated version of H: $H : Q = \mathrm{rot}(H, 180°)$. The combined descriptor, PIIFD, can be calculated as follows:

$$\mathrm{des}=\begin{bmatrix} (H_1+Q_1) \\ (H_2+Q_2) \\ c|H_3-Q_3| \\ c|H_4-Q_4| \end{bmatrix} \tag{9}$$

$$H_i=[H_{i1} \quad H_{i2} \quad H_{i3} \quad H_{i4}] \tag{10}$$

$$Q_i=[Q_{i1} \quad Q_{i2} \quad Q_{i3} \quad Q_{i4}] \tag{11}$$

where c is a parameter to tune the proportion of magnitude in this local descriptor. The absolute value of descriptor is normalized in the next step. In our algorithm, c is adaptively determined by making the maximum of two parts the same. The goal of the linear combination is to make the final descriptor invariant to two opposite directions. This linear combination is reversible, so it will not reduce the distinctiveness of the descriptor. It is obvious that PIIFD is a 4×4×8 matrix. For the convenience of matching, it is quantized to a vector with 128 elements. Finally, the PIIFD is normalized to a unit length.

D. Match PIIFDs by Bilateral Matching Method

We use the best-bin-first (BBF) algorithm [50] to match the correspondences between two images. This algorithm identifies the approximate closest neighbors of points in high dimensional spaces. This is approximate in the sense that it returns the closest neighbor with the highest probability. Suppose that the set of all PIIFDs of image I_1 is F_1, and the set of I_2 is F_2, then for a given PIIFD $f_{1i} \in F_1$, a set of distances from f_{1i} to F_2 is defined as follows:

$$D(f_{1i}, F_2)=\left\{f_{1i} \bullet f_{2i} | \forall f_{2j} \in F_2\right\} \tag{12}$$

where ● is the dot product of vectors. It is obvious that this set comprises all the distances between f_{1i} and descriptors in I_2. Let f_{2j}' and f_{2j}'' be the biggest and second-biggest elements

of $D(f_1, F_2)$, which correspond to $f'_{1i}s$ closest and second-closest neighbors, respectively. If the closest neighbor is significantly closer than the second-closest neighbor, $f_{2j}''/f_{2j}' < t$, then consider it a unilateral match (or a pair of corresponding points) from f_{1i} to F_2. In this study, the threshold t for the nearest-neighbor criterion is set to 0.9 empirically. Otherwise, the PIIFD f_{1i} is discarded.

The BBF algorithm mentioned above is unilateral, which cannot exclude the following mismatch: two descriptors in I_1 are matched to the same descriptor in I_2. The bilateral BBF algorithm is as simple as the unilateral one. Suppose the above unilateral matches are denoted as $M(I_1, I_2)$, and other unilateral matches $M(I_2, I_1)$ are also applied, then the same matches between these two set of matches are the bilateral matches.

E. Remove any Incorrect Matches

Even the bilateral BBF algorithm cannot guarantee that all matches are correct. Fortunately, it is easy to exclude the incorrect matches using the control point candidates' main orientations and the geometrical distribution of matched control point candidates. Suppose there are K bilateral matches in total, $m(p_{11}, p_{21}), \ldots, m(p_{1k}, p_{2k})$, where p_{1i} denotes the control point candidate for extracting feature descriptor f_{1i} in image I_1, and p_{2i} denotes the corresponding control point candidate in I_2. It is obvious that all differences between any $p'_{1i}s$ and $p'_{2i}s$ orientations are almost the same. If one of these differences of orientations is much bigger or smaller than the others, then this match is definitely incorrect. Most incorrect matches are actually excluded according to this criterion.

Next, we calculate the geometrical distribution of matched control point candidates. The ratio of distances of two matches is defined as $r_{ij} = d_e(p_{1i}, p_{1j})/d_e(p_{2i}, p_{2j})$, where $d_e(p_{1i}, p_{1j})$ means the Euclidian distance of the two control point candidates. For retinal image registration, there is no significant affine transformation, thus, all ratios between any two correct matches will also remain the same. We can remove the remaining incorrect matches according to this criterion.

These two criterions used to remove incorrect matches are based on the same assumption that there are more correct matches than incorrect matches. In our framework, the assumption is satisfied when the threshold for the nearest neighbor criterion in bilateral matching is set to equal to or less than 0.9.

F. Refine Locations of Matches

For each match between two images, the two control points may not be accurately the same corresponding to each other. For instance, assume p_i and p_j are two control points of a match. p'_i, a neighbor of p_i, which is not a control point, may be better than p_i when matched by p_j. This inaccuracy is caused by the Harris detector for the reason that two images for registration are not exactly the same due to multimodality, noise, and some other factors, thus, the corner points detected by Harris detector may not be very accurate. In this paper, we develop an algorithm to refine the inaccurate matches. The basic idea is to find the best match of PIIFDs when searching corresponding control points in a small neighborhood (radius < ε). Given that the two control points have already been coarsely matched, a small radius for searching is enough. Thus, ε is set to 2.5 pixels empirically. Following are the details of implementation.

Assume $m(p_i, p_j)$ is a match and p'_i is a neighbor of p_i. The coordinates of p_i and p'_i are (x_i, y_i) and (x'_i, y'_i), respectively. P', a set of neighbors of p_i, is defined as follows:

$$p' = \left\{ p_i' \middle| \left(x_i' - x_i\right)^2 + \left(y_i' - y_i\right)^2 \le \varepsilon^2 \right\}. \tag{13}$$

For convenience, we replace the circular neighborhood with a square neighborhood. In that way, P' is a 5 × 5 pixels square centered at p_i. For refining this match, 25 PIIFDs extracted around all pixels in P' are used to compare with the PIIFD of p_j. The pixel, p_{i_new}, that corresponding to the best match are recognized as the refined location of p_i. Thus, the match $m(p_i, p_j)$ is refined as $m(p_{i\_new}, p_j)$.

An example of bilateral matching, removing incorrect matches and refining inaccurate matches is shown on Fig. 6. This example shows that some control points situated in the middle of the homogeneous region are identified as correct matches. This phenomenon is reasonable for the reason that the Harris detector does not remove those points automatically. Thus, the PIIFDs will be matched, and some of them will be identified as correct matches because the local area in which PIIFD extracted may contain some inhomogeneous region.

G. Select Transformation Mode

This subsection is necessary for the integrality of our registration framework, but it is not an emphasis of the present study. Transformation mode is adaptively selected by the number of matches. *Linear conformal* [13], [17], *affine* [14], [29], and *second-order polynomial* [27], [47], [51] transformations are used in our framework. The simplest transformation (*linear conformal*) requires at least two pairs of control points. Therefore, no transformation is applied on a floating image if there is no match or only one match. We always apply a higher order transformation as long as the number of matches is sufficient. In other words, if there are only two matches then *linear conformal* transformation is applied, if the number of matches is bigger than or equal to three but less than six then *affine transformation* is applied, and if the number is bigger than or equal to six then *second-order polynomial transformation* is applied.

When a transformation mode has been applied on the floating image, we simply superpose the transformed image on the fixed image to produce a mosaic image.

III. Experiments and Results

In this section, we evaluate the proposed Harris-PIIFD and compare it with GDB-ICP [52] and SIFT [42] on 168 pairs of multimodal retinal images. All comparative experiments are implemented on an IBM Core 2 duo 2.4-GHz desktop computer with 2 GB of RAM.

The Harris-PIIFD algorithm is implemented in MATLAB. All parameter settings of our algorithm are explained in the methods section. SIFT algorithm is a classical feature extraction algorithm. The publicly available version of SIFT is an executable program implemented in C++ [56]. This program can extract SIFT features and identify the SIFT matches between two images. For the convenience of comparison, we replaced the original unilateral matching method in SIFT with our bilateral matching method and apply a transformation algorithm on the floating image based on the matched SIFT keypoints to provide the mosaic image. GDB-ICP algorithm is so far the best registration algorithm for processing poor quality retinal images. This algorithm can be downloaded as a binary executable program written in C++ at [53]. All parameter settings of SIFT and GDB-ICP are suggested by their authors/ implementers. For GDB-ICP, a switch ([–complete]) is used to register difficult image pairs—those having substantial illumination difference or physical

changes. This switch is applied if and only if the normal mode has failed [53]. The thresholds for the nearest neighbor criterion in SIFT and Harris-PIIFD are of the same value.

These three preprocessing operations discussed in Section II are also applied before SIFT and GDB-ICP algorithms to ensure that the comparative experiments are fair.

A. Data

The overall performance and comparative results are demonstrated on two retinal image datasets. One contains 150 pairs of multimodal images provided by the Department of Ophthalmology, Columbia University. These images are of three modalities: auto fluorescence, infrared, and red-free (which is ordinary fundus photography with a green filter). Some of those pairs were taken at the same time, while others were taken months or even years apart. The images range in size from 230 × 230 to 2400 × 2000 pixels. The least image pairs overlap is 20%, the biggest rotation angle is 30°, and they differed in scale (measured in one dimension) by a factor as high as 1.8. All the images are more or less affected by retinopathies. The second dataset contains 18 pairs of images collected from Internet. All images in the second dataset have the resolution of approximately 500 × 500 pixels.

B. Evaluation Method

A reliable and fair evaluation method is very important for measuring the performance since there is no public retinal registration dataset. We have considered several evaluation methods, including automatic pseudo ground-truth [35] and superposition of the vasculature tree [36], [55]. However, we have to take account into the maneuverability, because it is extremely difficult to automatically or manually extract the accurate vasculature tree in some images of our test datasets. Thus, we could not adopt these evaluation methods.

In our experiments, we manually select six pairs of corresponding landmarks, i.e., bifurcations and corner points, from the resized original images to generate the ground-truth. This evaluation method is boring and time consuming, and may be less accurate than automatic registration in some cases. But it has two advantages. It can: 1) handle very poor quality images and 2) provide a relatively reliable and fair measurement over all images. We manually select the landmarks satisfying the following constraints: 1) landmarks should be salient in both images and 2) landmarks should be uniformly distributed in the overlapping area. We have to select these most salient landmarks because there is no chance to accurately locate the control point if image quality is extremely poor. The corresponding landmarks could not be close to each other; otherwise, small location errors may result in big scale and rotation estimation errors. Some cases in our datasets require special care, on which we usually select the landmarks over and over again. This task takes approximately 10 h, and afterward we develop a program to estimate the transformation parameters and overlapping percentage.

Centerline error measure (CEM) [26], [35] is a popular measurement in retinal image registration, which measures the median error of the centerline of vasculature. However, it is extremely difficult to extract the vasculature from some images in our datasets. In the present study, we introduce the median error (MEE) and the maximal error (MAE) over all correspondences (six pairs used in generating ground truth) to evaluate the performance of these algorithms. MEE is similar to CEM but more suitable for poor quality images in which the vasculature cannot be extracted. For some cases, only a small fraction of image is mismatched, which result in a small MEE, but a large MAE. This is why we apply MAE to identify this kind of incorrect registration. In our experiments, all results are manually validated to guarantee that the entire image is matched without significant error (MAE > 10

pixels). These alignments with significant errors are treated as "*incorrect*" registration, while these alignments without significant errors are treated as "*effective*" registration, including "*acceptable*" registration (MEE ≤ 1.5 pixels and MAE ≤ 10 pixels) and "*inaccurate*" registration (MEE > 1.5 pixels and MAE ≤ 10 pixels). The acceptance threshold of 1.5 pixels on the MEE is established empirically by [26]. The "*incorrect*" threshold of 10 pixels on MAE is also established empirically by postviewing the results, and is reliable for the reason that the "*incorrect*" registration and "*effective*" registration decided by this threshold are desirable.

C. Special Examination

In our test datasets, the largest scaling factor is 1.8, and the largest rotation angle is 30°. These values are not large enough to test the scale insensitivity and rotation invariance of the proposed algorithm. Thus, we randomly select 20 pairs from our datasets, zoom and rotate to test the scale insensitivity and rotation invariance. Moreover, we also test the performance on small overlapping area registration and on multimodal images.

1) Rotation Invariance Test—We automatically rotate the floating images in the selected pairs from 0° to 180° with a 10° Step. It should be noted that the reference image is held fixed. We apply our Harris-PIIFD algorithm on the reference image and the rotated floating image. The results of this test show that the Harris-PIIFD algorithm successfully registers all image pairs regardless of the rotation angle, demonstrating that our Harris-PIIFD algorithm is rotation invariant. Two samples of this test are shown in Fig. 7.

2) Scale Change Test—PIIFD provides robust matching across a substantial range of scale change. In this section, we automatically zoom in the floating image to reach a scale factor from 1 to 3 with a step of 0.1 and apply the Harris-PIIFD algorithm on all these images. The registration success rate to scale difference can be seen in Fig. 8. The results of this experiment indicate that our proposed Harris-PIIFD can provide robust matching when the scale factor is below 1.8. The reason lies behind is that PIIFD is computed in a scale-insensitive manner. However, if the scale factor is above 1.8, our Harris-PIIFD usually fails. Fortunately, most retinal image pairs in clinic are of very small-scale difference (usually less than 1.5). Thus, our proposed Harris-PIIFD framework is still suitable for retinal image registration.

3) Overlapping Percentage Test—Ten pairs of images with large non-overlapping area are selected to test the performance on small overlapping area image registration. The images are cropped to reach the overlapping percentage from 5% to 50% with an increasing step of 5%. All phantom images in this test are generated manually with Photoshop (Adobe Systems Incorporated, San Jose, CA), thus, the overlapping percentages are not very precise. The registration success rate relative to overlapping percentage can be seen in Fig. 9. Results show that our method registers all pairs effectively when the overlapping percentage is above 30%. The success rate drops to 60% when the overlapping area decreases to 20%. Our method fails in some cases, as there are insufficient minutiae in the overlapping area. This may result in insufficient control point candidates in the small overlapping area. In these cases, we expect that other algorithms will also fail.

4) Multimodal Image Test—In this test, 400 pairs of corresponding control points and 400 pairs of noncorresponding control points are chosen from 20 pairs of different-modal retinal images. For each control point, PIIFD and SIFT are extracted. The similarity between two PIIFDs at corresponding control points and noncorresponding control points are measured. Similarly, the similarities between two SIFTs are also measured. The mean and standard deviation of these similarities can be seen in Fig. 10. Results indicate that PIIFDs

extracted around corresponding locations are much more similar than SIFTs, and PIIFDs extracted around noncorresponding locations are of almost the same similarity as SIFTs. These data suggest that PIIFDs can be easily identified at corresponding control points, while SIFTs cannot be identified.

D. Overall Test and Comparative Results on Real Retinal Images

1) Overall Test—It takes approximately 41.3 min to register all 168 pairs of retinal images using our Harris-PIIFD algorithm (14.75 s per pair, standard deviation of 4.65 s). This experiment is implemented automatically by a "for" loop in MATLAB. The time for loading and saving images (*I*/*O*) is not included.

One-hundred fifty-one out of 168 pairs of images (14 from the second dataset) are registered very well ("*acceptable*" registration with MEE ≤ 1.5 pixels and MAE ≤ 10 pixels, respectively). Sixteen pairs of the remaining (three from the second dataset) are "*inaccurate*" registration with MEE > 1.5 pixels and MAE ≤ 10 pixels, and only one "*incorrect*" registration (MAE > 10 pixels) is reported (see Fig. 11). The percentage of "*acceptable*" registration is 89.9%. The percentage of "*effective*" registration (including "*acceptable*" and "*inaccurate*" registrations) is 99.4%. Selected cases with "*acceptable*" registration can be seen in Figs. 12 and 13, respectively.

In some extreme cases, the number of control point candidates is only about 100, since most minutiae in these images are lost. Fortunately, 100 control point candidates are enough to guarantee the Harris-PIIFD to work as expected. But for robustness, we usually select approximately 200–300 control point candidates in each image. In our experiments, the average number of control point candidates is 231, the average number of initial matches (including incorrect matches) is 64.6, and the average number of final matches (after removing incorrect matches) is 43.2. One-hundred fifty-nine of 168 pairs have more than six final matches, which is the least number for the *second-order polynomial transformation*. About 40% matches are refined by our subsequent procedure after removing the incorrect matches. Most of these refined control points have shifted less than two pixels, eight control points have shifted about three pixels, and no control point have shifted more than four pixels.

The final registration is insensitive to the σ of the Gaussian window in Section II-B. The algorithm with σ from four to ten pixels produces almost the same final results. Thus, in this study, we set this parameter to a fixed five pixels. The threshold of the ratio parameter (see Section II-D) is also very important. The present value of this threshold is 0.9 in our experiments. If this value increases to 0.95, the number of initial matches increases by about 30%; however, almost all the increased matches are incorrect. On the other hand, if the threshold decreases to 0.85, the number of final matches drops by approximately 20%, which may result in some failed registration.

In our experiments, the size of PIIFD is always 4 × 4 × 8. This is inherited from Lowe's SIFT algorithm, and has been proven to be the best size for such a local feature descriptor.

2) Comparative Results—Herein comparative experiments between the Harris-PIIFD, SIFT, and GDB-ICP are implemented. These algorithms are applied to all 168 pairs of real retinal images. The mean and standard deviation of runtime can be seen in Table I. The Harris-PIIFD is much faster than GDB-ICP since no iterative and optimization schemes are used. The standard deviation runtime of Harris-PIIFD is also smaller than that of GDB-ICP, as the iterative times of the GDB-ICP algorithm on poor quality images are much more than that on good quality images. The SIFT algorithm is more computationally efficient than

Harris-PIIFD for the reason that SIFT is implemented in "C++" but Harris-PIIFD is in MATLAB.

In this comparison, there are 30 pairs of images, which have no outputs for the GDB-ICP algorithm, meaning that the GDB-ICP algorithm decides that these 30 pairs are not registrable. This number for SIFT is 19. Compared with those two algorithms, this number for the proposed Harris-PIIFD algorithm is 0. For all outputs of these three algorithms, we divide them into three classes: "*incorrect*" registration, "*inaccurate*" registration, and "*acceptable*" registration (see Table I). The criterion for this classification is discussed in Section III-B. The comparative accuracy (means and standard deviations of median errors) of all outputs can be seen in Table II.

This comparison demonstrates that Harris-PIIFD algorithm far outperforms the SIFT and GDB-ICP algorithms in multimodal retinal image registration.

IV. Discussions and Conclusion

The Harris-PIIFD algorithm is computationally efficient and totally automatic. It works on very poor quality images in which the vasculature is hard or even impossible to extract. Experimental results show that the Harris-PIIFD algorithm far outperforms the other algorithms both in accuracy and computational efficiency when testing on our 168 pairs of multimodal retinal images. This method can deal with a large initial misalignment registration such as perspective distortion, change of scale, and arbitrary rotation. It can also deal with registration of field images with large nonoverlapping areas.

PIIFD can be used as the features in other feature-based algorithms. For example, the PIIFD algorithm can provide the initial matches and robust low-contrast features for the GDB-ICP algorithm. PIIFD is designed for retinal image registration, but it can also be applied to register other multimodal images like cerebral MR T1-T2 images. In future, we will mainly concentrate on applying PIIFD on other image registration fields.

Acknowledgments

The authors would like to thank the owners of the retinal images that are used in this paper.

This work was supported in part by the National Eye Institute under Grant R01 EY015520-01, by the NYC Community Trust (RTS), by the unrestricted funds from Research to prevent blindness, by the Project for the National Basic Research Program of China (973) under Grant 2006CB705700, by Changjiang Scholars and Innovative Research Team in University (PCSIRT) under Grant IRT0645, by CAS Hundred Talents Program, by CAS Scientific Research Equipment Development Program under Grant YZ200766, by the Knowledge Innovation Project of the Chinese Academy of Sciences under Grant KGCX2-YW-129 and Grant KSCX2-YW-R-262, by the National Natural Science Foundation of China under Grant 30672690, Grant 30600151, Grant 60532050, Grant 60621001, Grant 30873462, Grant 60910006, Grant 30970769, and Grant 30970771, by Beijing Natural Science Fund under Grant 4071003, and by the Science and Technology Key Project of Beijing Municipal Education Commission under Grant KZ200910005005.

Biographies

Jian Chen received the B.S. degree in electronic engineering from Beijing Normal University, Beijing, China, in 2003, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2009.

From 2007 to 2008, he was a Graduate Research Assistant in the Department of Biomedical Engineering, Columbia University, New York, NY. He is currently an Associate Researcher at IBM China Research Laboratory, Beijing. His research interests include medical image analysis, computer vision, pattern recognition, and machine learning.

Jie Tian (M'02–SM'06–F'10) received the Ph.D. degree (with honors) in artificial intelligence from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1992.

From 1995 to 1996, he was a Postdoctoral Fellow at the Medical Image Processing Group, University of Pennsylvania, Philadelphia. Since 1997, he has been a Professor at the Institute of Automation, Chinese Academy of Sciences, where he has been involved in the research in Medical Image Processing Group. His research interests include the medical image process and analysis and pattern recognition. He is the author or coauthor of more than 70 research papers published in international journals and conference proceedings.

Dr. Tian is the Beijing Chapter Chair of The Engineering in Medicine and Biology Society of the IEEE.

Noah Lee received the B.S. and M.S. degrees from the Computer Science Department, Beuth University of Applied Sciences for Technology, Berlin, Germany, in 2004, and the M.S. and M.Phil. degrees in biomedical engineering from Columbia University, New York, NY, in 2006.

He is currently a Graduate Research Assistant in the Department of Biomedical Engineering, Columbia University, where he has been involved in the research on exploring image segmentation algorithms for cardiac, ophthalmic, and oncologic imaging applications. His research interests include medical image analysis, computer-aided diagnosis, and machine learning.

Jian Zheng received the B.E. degree in automation from the University of Science and Technology of China, Beijing, China, in 2005. He is currently working toward the Ph.D. degree as a Team Member in the Medical Imaging Tool Kit (MITK), Medical Image Processing Group at the Institute of Automation, Chinese Academy of Sciences, Beijing.

Since 2005, he has been engaged in the development of MITK and 3-D medical image processing and analyzing system (3DMed). His current research interests include medical image registration.

R. Theodore Smith received the B.A. degree from Rice University, Houston TX, in 1969 and the Ph.D. degree from Warwick University, Coventry, U.K., in 1972, both in mathematics, and the M.D. degree from Albert Einstein College of Medicine, Bronx, NY, in 1981.

He is currently the Director of the National Institute of Health (NIH) supported Retinal Image Analysis Laboratory, Edward S. Harkness Eye Institute, Columbia University, New York, NY, and a Professor of clinical ophthalmology and biomedical engineering in the Department of Ophthalmology. He has used patented image analysis techniques over the past ten years in collaboration with the Department Biomedical Engineering, Columbia University, for quantitative study of retinal degenerations, particularly Stargardt macular dystrophy, retinitis pigmentosa, and age-related macular degeneration (AMD). As a Practicing Retinal Specialist, he has also been the Clinical Principal Investigator of the Columbia Macular Genetics Study, since 1999, which with 2977 subjects and controls is the largest study of its kind in the world, and which discovered the complement factor H risk gene for AMD. His current NIH funded research projects include: developing a specially modified ophthalmoscope to measure absolute autofluorescence, thus, quantifying the fluorescence associated with lipofuscin, which is an important biomarker in these diseases, the study of reticular macular disease, which is strongly associated with advanced AMD and may indicated an infectious trigger for the inflammation underlying advanced AMD in genetically susceptible individuals, and hyperspectral images of drusen, the hallmark lesions of AMD, which vary biochemically and are key to understanding the disease, with a snapshot hyperspectral imaging system for noninvasive spectral biopsy of the retina. His current research projects have the potential to increase the scientific understanding of retinal disease well beyond the present level, ultimately providing a firmer basis for monitoring patient progress and response to new therapies, including gene therapy.

Andrew F. Laine (M'83–SM'07–F'10) received the D.Sc. degree in computer science from the School of Engineering and Applied Science, Washington University, St. Louis, MO, in 1989, and the B.S. degree from Cornell University, Ithaca, NY.

He is currently the Director of the Heffner Biomedical Imaging Laboratory, Department of Biomedical Engineering, Columbia University, New York, NY, and a Professor of biomedical engineering and radiology (physics). His research interests include quantitative image analysis; cardiac functional imaging: ultrasound and magnetic resonance imaging, retinal imaging, intravascular imaging, and biosignal processing.

Prof. Laine is a Fellow of the American Institute for Medical and Biological Engineering. He has served as the Chair of the Technical Committee on Biomedical Imaging and Image Processing for the Engineering in Medicine and Biology Society (EMBS) during 2006–2009 and of the IEEE International Symposium on Biomedical Imaging (ISBI) Steering Committee, since 2006, the Program Chair for the IEEE EMBS Annual Conference, New York, in 2006, the Program Co-Chair for IEEE ISBI, Paris, France, in 2008; and the Vice President of Publications for IEEE EMBS, since 2008.

References

[1]. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy-ETDRS report number 12. Ophthalmology 1991;98:823–833. [PubMed: 2062515]

[2]. Saine, PJ.; Tyler, ME. Retinal Photography, Angiography, and Electronic Imaging. Butterworth-Heinemann; Boston, MA: 1997.

[3]. Sanchez-Galeana C, Bowd C, Blumenthal EZ, Gokhale PA, Zangwill LM, Weinreb RN. Using optical imaging summary data to detect glaucoma. Ophthalmology 2001;108(no. 10):1812–1818. [PubMed: 11581054]

[4]. Zhou L, Rzeszotarski MS, Singerman LJ, Chokreff JM. The detection and quantification of retinopathy using digital angiograms. IEEE Trans. Med. Imag Dec;1994 13(no. 4):619–626.

[5]. Zitova B, Flusser J. Image registration methods: A survey. Image Vis. Comput Oct;2003 21(no. 11):977–1000.

[6]. Crum WR, Hartkens T, Hill DLG. Non-rigid image registration: theory and practice. Br. J. Radiol 2004;77:140–153.

[7]. Hill DLG, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. Phys. Med. Biol 2001;46:R1–R45. [PubMed: 11277237]

[8]. Hajnal, JV.; Hill, DLG.; Hawkes, DJ. Medical Image Registration. CRC Press; Boca Raton, FL: 2001.

[9]. Maintz JBA, Viergever MA. A survey of medical image registration. Med. Image Anal Mar;1998 2(no. 1):1–36. [PubMed: 10638851]

[10]. Brown LG. A survey of image registration techniques. ACM Comput. Surv Dec;1992 24(no. 4): 325–376.

[11]. Lester H, Arridge S. A survey of hierarchical nonlinear medical image registration. Pattern Recognit 1999;32(no. 1):129–149.

[12]. Cideciyan V. Registration of ocular fundus images. IEEE Eng. Med. Biol. Mag Jan;1995 14(no. 1):52–58.

[13]. Matsopoulos GK, Mouravliansky NA, Delibasis KK, Nikita KS. Automatic retinal image registration scheme using global optimization techniques. IEEE Trans. Inf. Technol. Biomed Mar;1999 3(no. 1):47–60. [PubMed: 10719503]

[14]. Mourvliansky N, Matsopoulos G, Delibasis K, Nikita K. Automatic retinal registration using global optimization techniques. Proc. Int. Conf. Eng. Med. Biol. Soc Nov;1998 2:567–570.

[15]. Peli E, Augliere RA, Timberlake GT. Feature-based registration of retinal images. IEEE Trans. Med. Imag Sep;1987 MI-6(no. 2):272–278.

[16]. Ritter N, Owens R, Cooper J, Eikelboom R, van Saarloos P. Registration of stereo and temporal images of the retina. IEEE Trans. Med. Imag May;1999 18(no. 5):404–418.

[17]. Lloret D, Serrat J, Lopez AM, Soler A, Villaneuva JJ. Retinal image registration using creases as anatomical landmarks. Proc. IEEE Proc. Int. Conf. Pattern Recognit Sep;2000 3:203–206.

[18]. Ballerini L. Temporal matched filters for integration of ocular fundus images. Proc. IEEE Conf. Digital Signal Process Jul;1997 2:1161–1164.

[19]. Skokan M, Skoupy A, Jan J. Registration of multimodal images of retina. Proc. IEEE Conf. Eng. Med. Biol Oct;2002 2:1094–1096.

[20]. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag Apr;1997 16(no. 2):187–198.

[21]. Thevenaz P, Unser M. Optimization of mutual information for multiresolution image registration. IEEE Trans. Imag. Proc Dec;2000 9(no. 19):2083–2099.

[22]. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imag Aug;2003 22(no. 8):986–1004.

[23]. Wachowiak MP, Smolikova R, Zheng YF, Zurada JM, Elmaghraby AS. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Trans. Evol. Comput Jun;2004 8(no. 3):289–301.

[24]. Orchard J. Efficient least squares multimodal registration with a globally exhaustive alignment search. IEEE Trans. Imag. Proc Oct;2007 16(no. 10):2526–2534.

[25]. Becker DE, Can A, Tanenbaum HL, Turner JN, Roysam B. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. IEEE Trans. Biomed. Eng Jan;1998 45(no. 1):105–118. [PubMed: 9444845]

[26]. Can A, Stewart CV, Roysam B, Tanenbaum HL. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans. Pattern Anal. Mach Mar; 2002 24(no.3):347–364.

[27]. Ege, B.; Dahl, T.; Sndergaard, T.; Larsen, O.; Bek, T.; Hejlesen, O. Automatic registration of ocular fundus images. Workshop Computer Assist. Fundus Image Anal.; Copenhagen, Denmark. 2000.

[28]. Hart W, Goldbaum M. Registering retinal images using automatically selected control point pairs. Proc. IEEE Int. Conf. Image Process 1994;3:576–581.

[29]. Jagoe R, Blauth C, Smith P, Arnold J, Taylor K, Wootton R. Automatic geometrical registration of fluorescein retinal angiograms. Comput. Biomed. Res Oct;1990 23(no. 5):403–409. [PubMed: 2225786]

[30]. Mahurkar AA, Vivino MA, Trus BL, Kuehl EM, Datiles MB III, Kaiser-Kupfer MI. Constructing retinal fundus photomontages. A new computer-based method. Investigat. Ophthalmol. Visual Sci Jul;1996 37(no. 8):1675–1683.

[31]. Mendonca A, Campilho J, Nunes A. A new similarity criterion for retinal image registration. Proc. IEEE Int. Conf. Image Process 1994:696–700.

[32]. Pinz A, Bernogger S, Datlinger P, Kruger A. Mapping the human retina. IEEE Trans. Med. Imag Aug;1998 17(no. 4):606–620.

[33]. Shen H, Stewart C, Roysam B, Lin G, Tanenbaum H. Framerate spatial referencing based on invariant indexing and alignment with application to laser retinal surgery. IEEE Trans. Pattern Anal. Mach Mar;2003 25(no. 3):379–384.

[34]. Yu J, Hung C, Liou BN. Fast algorithm for digital retinal image alignment. Proc. 11th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc 1989:0734–0735.

[35]. Stewart CV, Tsai CL, Roysam B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans. Med. Imag Nov;2003 22(no. 11):1379–1394.

[36]. Laliberte F, Gagnon L, Sheng Y. Registration and fusion of retinal images—An evaluation study. IEEE Trans. Med. Imag May;2003 22(no. 5):661–673.

[37]. Tsai CL, Stewart CV, Roysam B, Tanenbaum HL. Covariance driven retinal image registration initialized from small sets of landmark correspondences. Proc. IEEE Int. Symp. Biomed. Imag Jul;2002 :333–336.

[38]. Zhang EH, Zhang Y, Zhang TX. Automatic retinal image registration based on blood vessels feature point. Proc. IEEE Int. Conf. Mach. Learning Cybern Nov;2002 4:2010–2015.

[39]. Heneghan C, Maguire P, Ryan N, de Chazal P. Retinal image registration using control points. Proc. IEEE Int. Symp. Biomed. Imag Jul;2002 :349–352.

[40]. Park J, Keller JM, Gader PD, Schuchard RA. Hough-based registration of retinal images. Proc. IEEE Int. Conf. Syst., Man, Cybern Oct;1998 5:4550–4555.

[41]. Jasiobedzki P. Registration of retinal images using adaptive adjacency graphs. Proc. IEEE Symp. Comput. Based Med. Syst Jun;1993 :40–45.

[42]. Lowe DG. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis 2004;60(no. 2):91–110.

[43]. Lowe DG. Object recognition from local scale-invariant features. Proc. Int. Conf. Comput. Vis 1999:1150–1157.

[44]. Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach Oct;2005 10(no. 27):1615–1630.

[45]. Se S, Lowe DG, Little J. Vision-based mobile robot localization and mapping using scale-invariant features. Proc. IEEE Int. Conf. Robot. Autom. (ICRA) 2001:2051–2058.

[46]. Lindeberg T. Feature detection with automatic scale selection. Int. J. Comput. Vis 1998;30(no. 2):77–116.

[47]. Harris C, Stephens MJ. A combined corner and edge detector. Proc. Alvey Vis. Conf 1988:147–152.

[48]. Bazen AM, Gerez SH. Systematic methods for the computation of the directional fields and singular points of fingerprints. IEEE Trans. Pattern Anal. Mach. Intell Jul;2002 24(no. 7):905–919.

[49]. Hong L, Wan Y, Jain A. Fingerprint image enhancement: Algorithm and performance evaluation. IEEE Trans. Pattern Anal. Mach. Intell Aug;1998 20(no. 8):777–789.

[50]. Beis J, Lowe DG. Shape indexing using approximate nearest neighbor search in high dimensional spaces. Proc. Conf. Comput. Vis. Pattern Recognit., Puerto Rico 1997:1000–1006.

[51]. Ryan N, Heneghan C, de Chazal P. Registration of digital retinal images using landmark correspondence by expectation maximization. Image Vis. Comput 2004;22:883–898.

[52]. Yang G, Stewart CV, Sofka M, Tsai C. Registration of challenging image pairs: Initialization, estimation, and decision. IEEE Trans. Pattern Anal. Mach. Intell Nov;2007 29(no. 11):1973–1989. [PubMed: 17848778]

[53]. Stewart, CV.; Yang, G. The generalized dual bootstrap-ICP executable. Jan. 2008 [Online]. Available: http://www.vision.cs.rpi.edu/download.html

[54]. Sabaté-Cequier A, Dumskyj M, Usher D, Himaga M, Williamson T, Boyce, Nussey S. Accurate registration of paired macular and nasal digital retinal images: A potential aid to diabetic retinopathy screening. Tech. Rep. 2000

[55]. Chanwimaluang T, Fan G, Fransen SR. Hybrid retinal image registration. IEEE Trans. Info. Tech. Biomed Jan;2006 10(no. 1):129–142.

[56]. Lowe, D. The SIFT executable. Jul. 2005 [Online]. Available: http://www.cs.ubc.ca/~lowe/keypoints/

[57]. Bay H, Tuytelaars T, Gool LV. Surf: Speeded up robust features. Proc. 9th Eur. Conf. Comput. Vis 2006:404–417.

[58]. Chen J, Smith RT, Tian J, Laine AF. A novel registration method for retinal images based on local features. Proc. IEEE EMBS, Vancouver, Canada Aug;2008 :2242–2245.

[59]. Cattin PC, Bay H, Gool LV, Szekely G. Retina mosaicing using local features. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI) Sep;2006 4191:185–192.

[60]. Tsai C-L, Li C-Y, Yang G. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. IEEE Trans. Med. Imag Mar;2009 29(no. 3)

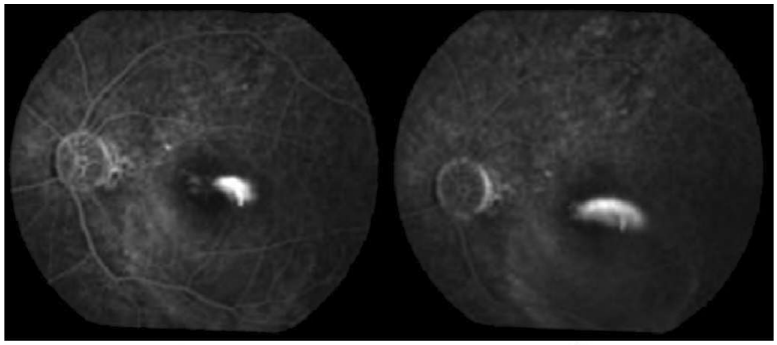

Fig. 1.
(a) and (b) Poor quality retinal images taken at different stages. Traditional feature-based approaches usually fail to register this image pair since it is hard to detect the vasculatures in (b).

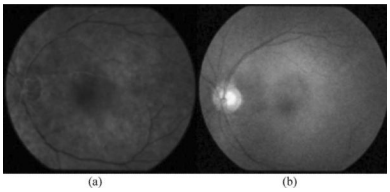

Fig. 2.
Pair of poor quality multimodal retinal images. These two images were taken from the same eye.

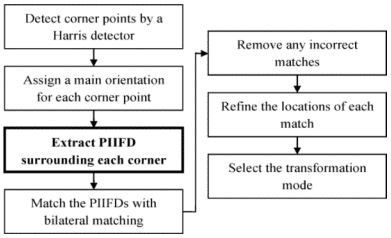

Fig. 3.
Flowchart of our registration framework. The key contribution of this study (see Section II-C) is highlighted in bold.

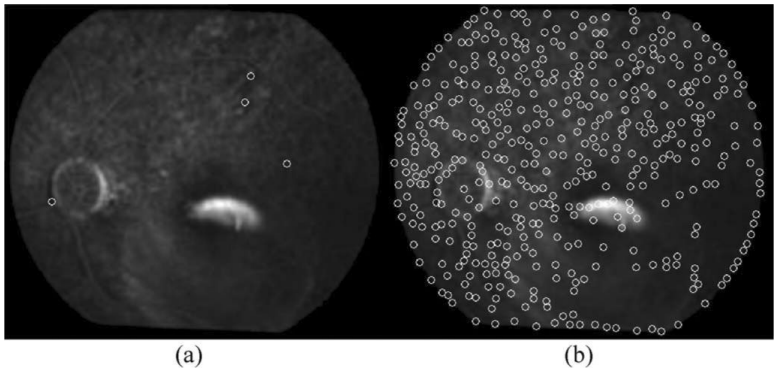

Fig. 4.

Spatial distributions of the control point candidates represented by (a) bifurcations of vasculature detected by an automatic central line extraction method and (b) corner points detected by a Harris detector.

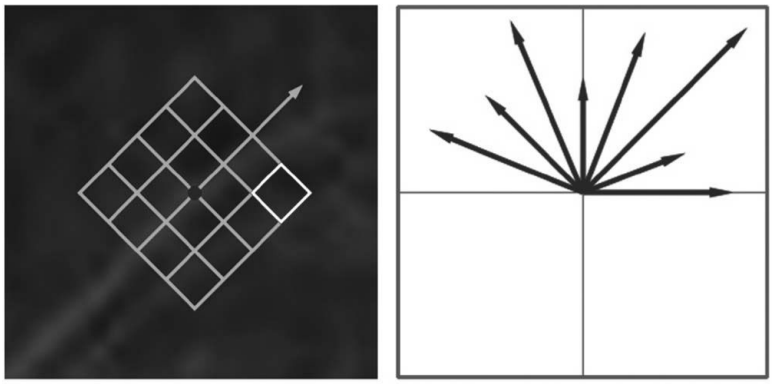

Fig. 5.
Extracting PIIFD relative to main orientation of control point candidate. (a) Neighborhood surrounding the control point candidate (centered point) is decided relative to the main orientation. (b) Orientation histogram extracted from the highlighted small square in (a).

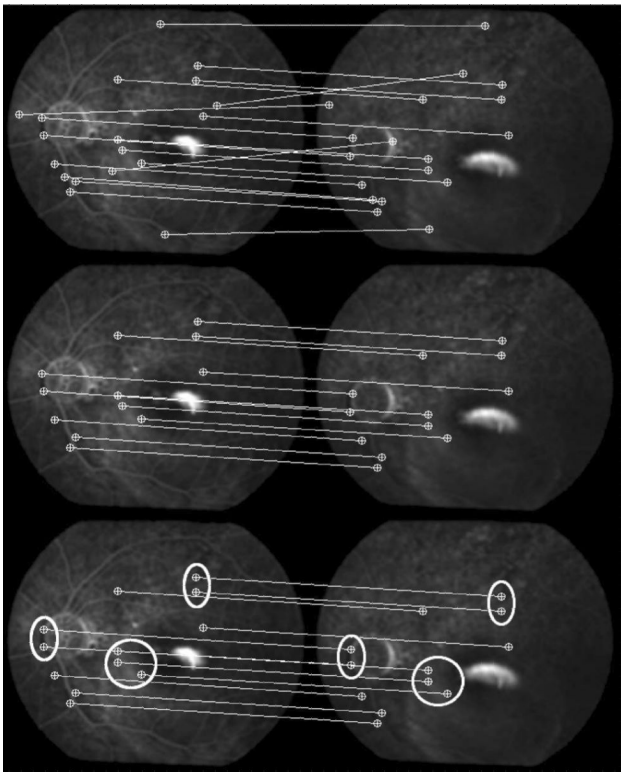

Fig. 6.
Matches that are identified by our method between the two images in Fig. 1. (First row) initial matching results. (Second row) results after removing the incorrect matches. (Third row) results after refining the inaccuracy matches.

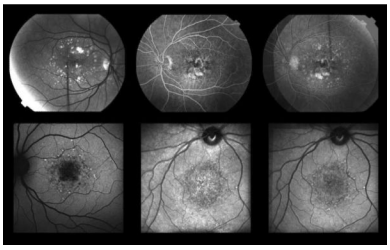

Fig. 7.
Two pairs in rotation-invariance test. The rotation angle of the first image pair (the first two images in the top row) is approximately 180°. The rotation angle of the second pair (the first two images in the bottom row) is approximately 90°. The mosaic images of registration are shown in the third column.

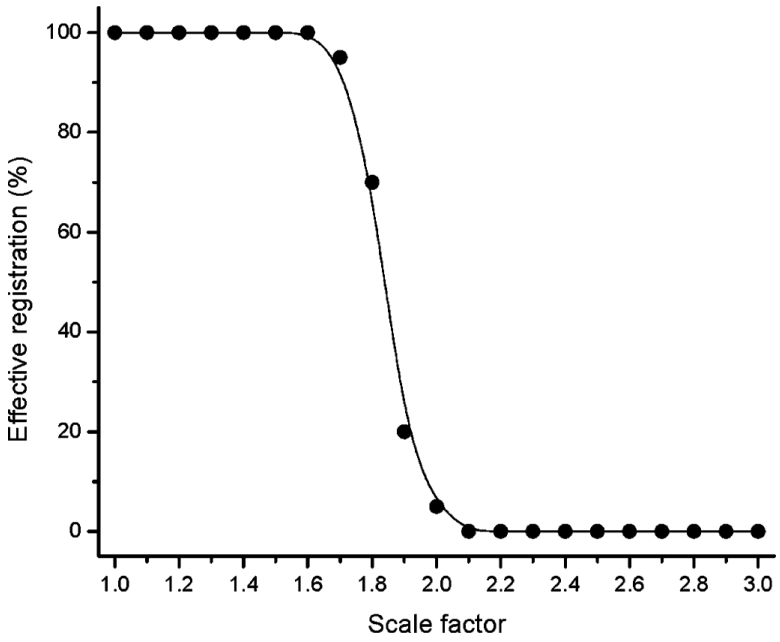

Fig. 8.

Percentage of effective registration relative to scale factor. The Harris-PIIFD framework can register retinal images with nearly flawless performance when the scale factor is below 1.8.

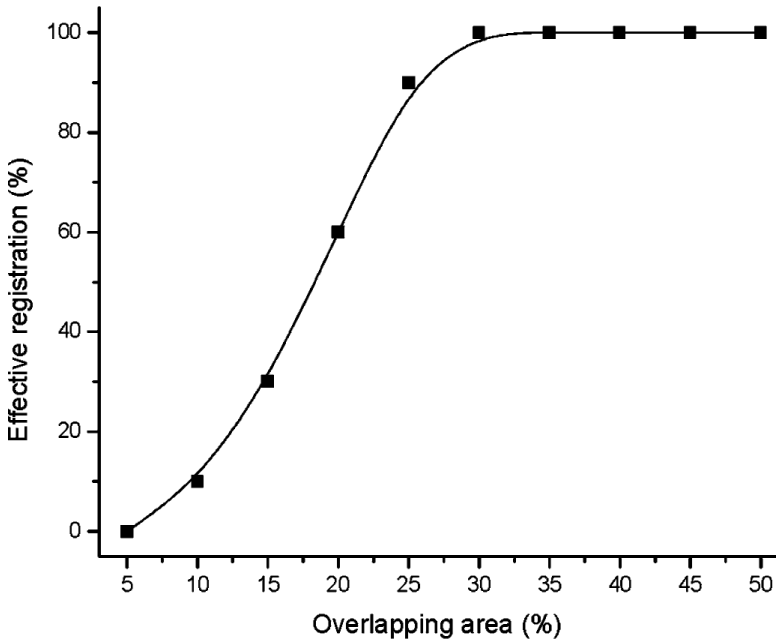

Fig. 9.

Percentage of effective registration by Harris-PIIFD relative to overlapping area. In this phantom image test, our method registers all pairs effectively when the overlapping percentage is above 30%. The success rate drops to 60% when the overlapping area is 20%.

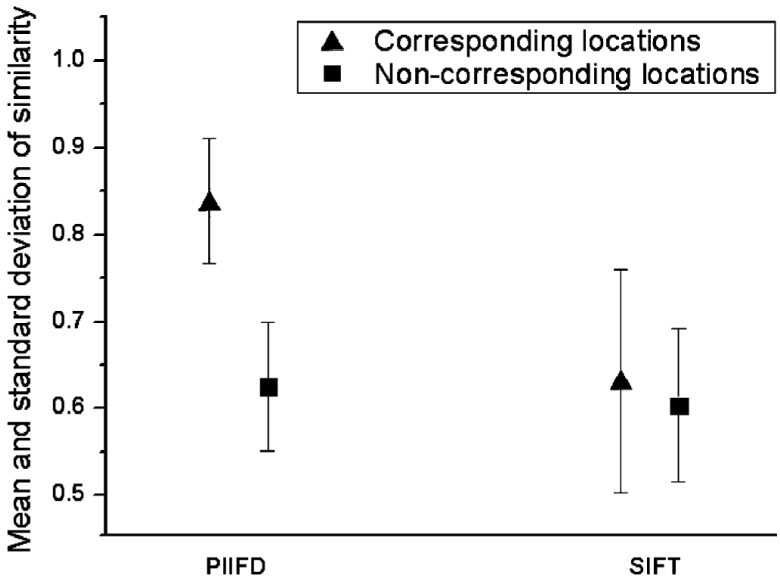

Fig. 10.

Mean and standard deviation of the similarities measured at corresponding locations and noncorresponding locations. PIIFDs are much more similar than SIFTs at corresponding locations, but at noncorresponding locations the PIIFDs are of almost the same similarity as SIFTs.

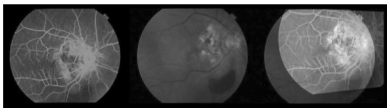

Fig. 11.
Only pair in our test datasets, which has not been registered by our method. The result of Harris-PIIFD is shown in the right.

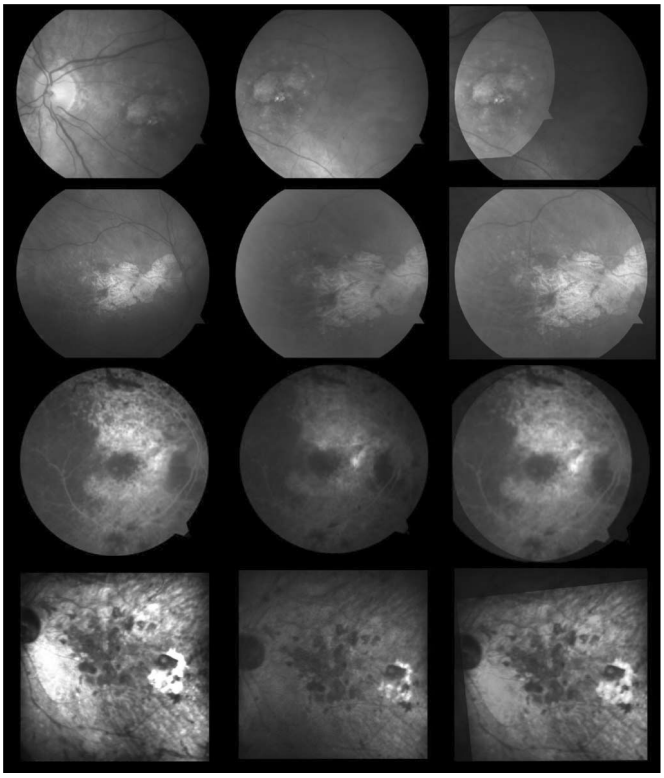

Fig. 12.
Four pairs selected from our test datasets (mono modality). The overlapping area of the first pair (first row) is approximately 40%. The factor of scale change of the second pair is approximately 1.5. The third and the fourth pairs are severely affected by retinopathies. The results of the Harris-PIIFD algorithm are shown in the third column.

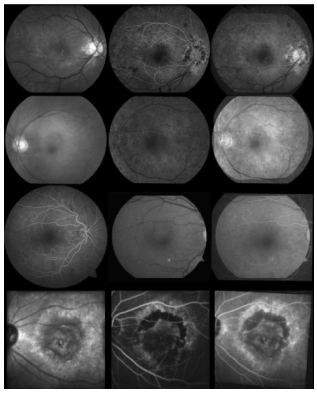

Fig. 13.
Four pairs selected from our test datasets (different modalities). The results of the Harris-PIIFD algorithm are shown in the third column.

Table I

Runtime and Success Rates of SIFT, GDB-ICP, and Harris-PIIFD

		SIFT	GDB-ICP	Harris-PIIFD
Runtime: mean±sd (s)		3.5±1.8	53.1±28.9	14.7±3.4
Failed (%)		11.3	17.9	0
Non-Failed (%)	incorrect	49.4	16.1	0.6
	inaccurate	12.5	4.2	9.5
	acceptable	**26.8**	**61.9**	**89.9**

Table II

Means and Standard Deviations of Median Errors (in Pixels) for all Outputs of SIFT, GDB-ICP, and Harris-PIIFD

	SIFT	GDB-ICP	Harris-PIIFD
Incorrect	19.9±8.1	18.3±7.7	16.3±0
Inaccurate	3.5±1 4	3.3±1.3	3.7±1.5
Acceptable	0.8±0.4	0.7±0.3	0.8±0.4
Total	**11.8±10.9**	**4.4±7.6**	**1.1±1.2**