Retinal Image Registration from 2D to 3D

Yuping Lin and Gérard Medioni
Computer Science Department, University of Southern California
3737 Watt Way, PHE 101 Los Angeles, CA, 90089
{yupingli, medioni}@usc.edu

Abstract

We propose a 2D registration method for multi-modal image sequences of the retinal fundus, and a 3D metric reconstruction of near planar surface from multiple views. There are two major contributions in our paper. For 2D registration, our method produces high registration rates while accounting for large modality differences. Compared with the state of the art method [5], our approach has higher registration rate (97.2% vs. 82.31%) while the computation time is much less. This is achieved by extracting features from the edge maps of the contrast enhanced images, and performing pairwise registration by matching the features in an iterative manner, maximizing the number of matches and estimating homographies accurately. The pairwise registration result is further globally optimized by an indirect registration process. For 3D registration part, images are registered to the reference frame by transforming points via a reconstructed 3D surface. The challenge is the reconstruction of a near planar surface, in which the shallow depth makes it a quasi-degenerate case for estimating the geometry from images. Our contribution is the proposed 4-pass bundle adjustment method that gives optimal estimation of all camera poses. With accurate camera poses, the 3D surface can be reconstructed using the images associated with the cameras with the largest baseline. Compared with state of the art 3D retinal image registration methods, our approach produces better results in all image sets.

1. Introduction

Multi-modality image registration is a fundamental task in medical image applications. Images captured from different sensors or across time frames often have different modalities and the alignment of these images is critical for diagnosis. Retinal image registration is one of the applications. The intensity of the the angiograms vary substantially while the sodium fluorescein dye in the retinal vessels circulates. Figure 1 shows an example of a retinal image sequence. In such a problem domain, mutual information [16] is widely used. It measures the statistical dependency between the image intensities, which is maximum when the images are geometrically aligned. However, it is time-consuming, thus impractical. Several attempts have been made in finding invariant features in retinal images. [5] uses high level Y features extracted at vessel junctions that are invariant to intensity variance. However, such features are too sparse and not well distributed for a robust and accurate registration. Stewart proposed the dual bootstrap iterative closest point algorithm [15] that has almost perfect registration rate. It starts from a local region where two landmarks are matched, and expand the region by aligning the detected blood vessel centerlines using ICP algorithm. However, the approach is limited to 2D registration since the matches produced by aligning points on vessel centerlines are not accurate enough for estimating 3D geometry. Another weakness of the two approaches is that the extraction high level features is time-consuming.

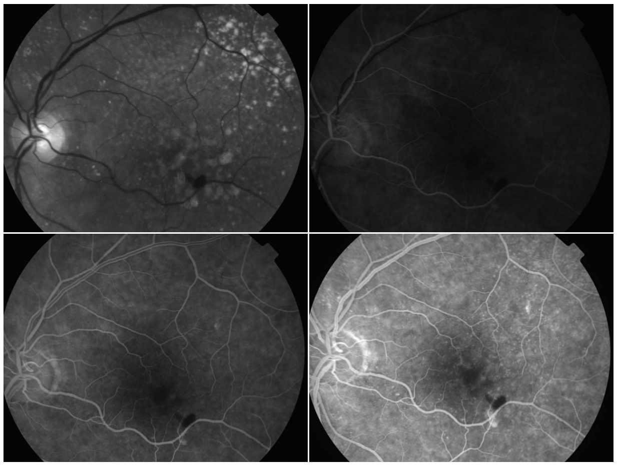


Figure 1. Retinal images in different modalities

We propose here a feature based method for 2D registration that registers multi-modality images with a high success rate. The left part of Figure 2 shows the flow chart of the process. We extract SIFT [12] features from contrast enhanced edge response images. A similar approach was pro-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

posed in [4], in which SURF features are used. However, it didn't address the issue of registering images with high variation in intensities. For each image pair, we match the features and derive the *homography* relationship between images in an iterative manner. Although Can et al. [3] showed that the curved nature of retina can best be taken into account by a quadratic model, the use of a homography model can be justified in the following reasons: First, we only register a local portion of the retinal surface, which can be considered as near planar. Second, a homography model has fewer parameters and is easier to estimate, especially when the matches are noisy. Third, we are interested in depth variations due to anatomical features, which cannot be accounted for by the homography, nor the quadratic model. Instead, we need a 3D surface model to provide good 3D registration. The accuracy of the 3D reconstruction relies on the accuracy of the fundamental matrix estimation, where we use plane+parallax method, and the homography can be plugged right in.

At each iteration, the homography estimated from the previous iteration is used to improve the quality of feature matching. This method registers most image pairs in the set. The result is further improved by chained registration, in which the homography of a poorly registered image pair is estimated by chaining a sequence of related accurate homographies. Our method has several advantages:

1. High registration rate. Our method has over 97% registration rate over 40 image sets with 574 image pairs.

2. It makes no assumptions about the content of the images. Compared with [15] [5] which extracts features from vessel junctions, our method extracts SIFT features that can be used on other multi-modality image registration problems that do not have similar vascular structures.

3. It is fast. SIFT feature extraction is much faster than high level feature extraction. The most time-consuming process is the iterative feature matching that can be limit by using fewer iterations. Compared with [5], our approach is 8 times faster.

4. It can be extended to 3D registration. The correspondences produced from SIFT feature matching is sub-pixel level accurate and their number large enough, which is suitable for 3D registration.

The output from our 2D registration method can be used in 3D reconstruction and 3D registration. The reconstruction of 3D surface has several advantages. By inspecting the 3D shape of a retinal surface, blisters which result from lesions can be easily identified. The 3D registration process also needs accurate 3D surface to infer point transformation between images. The reconstruction of retinal images belongs to the category of near-planar surface reconstruction, which is carefully studied in [6]. It is a difficult problem due to the lack of depth information, which is a quasi-degenerate case for the estimation of the 3D structure [8].

The 3D surface reconstruction and registration processes are illustrated in the right part of Figure 2. First the fundamental matrix for each image with respect to the reference frame is estimated. The corresponding fundamental matrix inliers are then input to a 4-pass bundle adjustment to estimate all camera poses. To reconstruct the 3D surface, we use the image associated with the camera that has the widest baseline with respect to the reference camera for dense stereo matching. The resulting dense correspondences are used to triangulate the 3D points on the surface. Finally, images are registered to the reference frame by back projection. The novelty of our approach is the 4-pass bundle adjustment in which the objective is to estimate the poses of all cameras. In [6], the camera selection strategy does not take the baseline into account, and produces poor results when two cameras are close.

The rest of the paper is organized as follows: Section 2 describes the 2D registration approach to multi-modality image sequences. We describe the image pre-processing, the iterative nearest neighbor matching and the chained registration, followed by experimental results. Section 3 describes the 3D registration approach. We describe the 4-pass bundle adjustment method and evaluate its performance in terms of accuracy. The last section concludes the paper and outlines future work.

2. 2D Registration of Multi-Modality Images

Since our approach uses some of the techniques in [1], we start by reviewing their method. Let Γ_i, Γ_j denotes the SIFT features extracted from I_i, I_j respectively. To match features, the nearest neighbor matching (NN-Match) is used. Each SIFT feature in Γ_i is matched to its nearest neighbor in Γ_j by computing the Euclidean distance in the feature space. Moreover, to prevent false matching, the distance of the nearest neighbor has to be less than the second-nearest neighbor by a ratio (we use 0.8). Note that with a larger search area, there will be more matching candidates and less probability for a feature to be matched. Let $M_{i,j}$ denotes the set of matches produced from NN-Match, $H_{i,j}$ denotes the homography that warps image I_i to image I_j, i.e., $I_j = H_{i,j}(I_i)$. To estimate $H_{i,j}$, RANSAC [7] is employed to perform a robust estimation. It performs several iterations on $M_{i,j}$. At each iteration, a homography model is built and its corresponding inliers correspondences are determined. The best homography estimate is the one with the largest number of inliers.

Our registration consist of three major steps. First the images are preprocessed to enhance the contrast and further transformed into edge map. Then we extract SIFT features

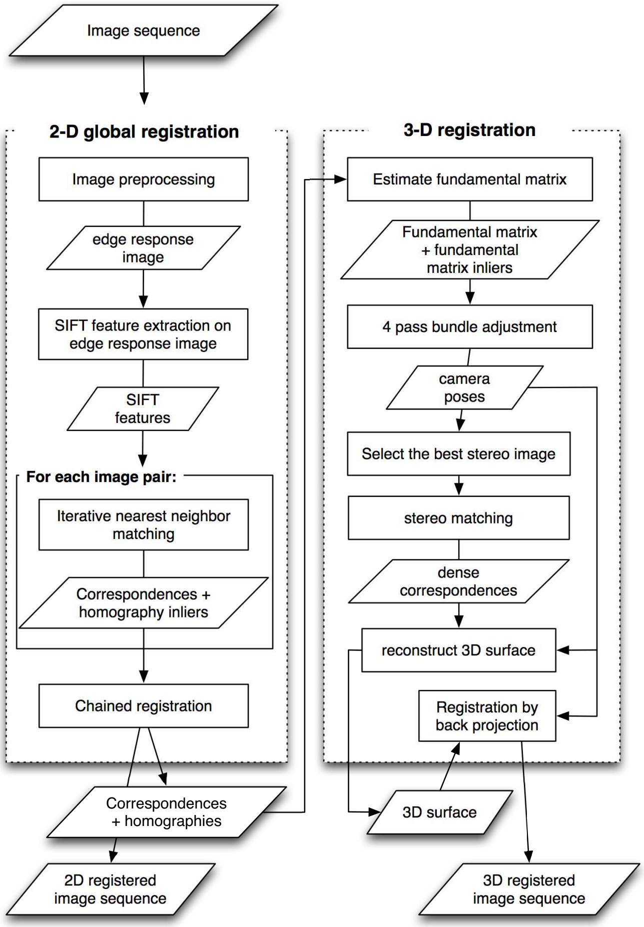


Image sequence
2-D global registration
Image preprocessing
edge response image
SIFT feature extraction on edge response image
SIFT features
For each image pair:
Iterative nearest neighbor matching
Correspondences + homography inliers
Chained registration
Correspondences + homographies
2D registered image sequence
3-D registration
Estimate fundamental matrix
Fundamental matrix + fundamental matrix inliers
4 pass bundle adjustment
camera poses
Select the best stereo image
stereo matching
dense correspondences
reconstruct 3D surface
Registration by back projection
3D surface
3D registered image sequence

Figure 2. Flow chart

from the edge maps and perform pairwise image registration using iterative nearest neighbor matching. Finally, the chained registration method is used to further improve the registration result from previous stage. These three steps are described in the following sections.

2.1. SIFT on Edge Response Images

SIFT feature extraction can produce large amount of descriptive features that increases the number of reliable matches. Edge information can be retrieved easily, and is widely spread across the image. More importantly, edge response preserves the gradient magnitude of the surface and ignores its gradient direction, which is invariant in multi-modality imagery. [10] use a similar idea, gradient mirroring, which associates opposite gradient directions in the SIFT descriptor. Although this approach is invariant to contrast reversals, the descriptor is less descriptive and would degrade the performance of matching.

Figure 3 shows two retinal images and their corresponding edge response images. Though the two images are quite different in intensity structure, their edge responses are similar. It is exptected that features extracted from edge response images are more likely to be matched.

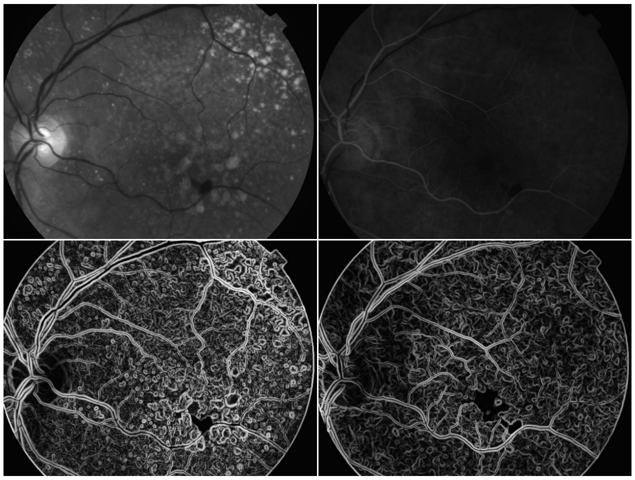


Figure 3. Retinal images and their corresponding edge responses

2.1.1 Image Pre-processing

To extract a sufficient large number of SIFT features from the edge response images, we preprocess the images by enhancing their contrast and removing noise. First, the intensity histogram is equalized to a Gaussian distribution with $\mu = 128, \sigma = 48$. In images that have very low contrast, such an operation generates much noise. We denoise such images with Non-local Mean Filter [2]. Then we compute the edge response on the contrast enhanced images using a Sobel filter. Finally, we use contrast limited adaptive histogram equalization (CLAHE)[13] to enhance the contrast of the edge. In our experiments, such a preprocessing yields the greatest number of true matches.

2.1.2 Experimental Results

We compare the quality of matching using SIFT features extracted from edge response images with those from original images. We select 10 pairs of images that have very different intensity profiles. In each image pair, the SIFT features on the original images and edge response images are extracted and matched respectively. We also label some control points so that we can compute the ground truth homographies of the image pairs and the number of true matches available. The true matches are the ones that comply with the ground truth homography. Figure 4 shows that matching features extracted from edge response images produce more true matches than original images.

Figure 4 also plots the number of true matches and total matches in the edge response images. It show that the number of true matches are too few too for RANSAC to estimate a good homography model. The issue now becomes how to estimate an accurate homography from the noisy matches.

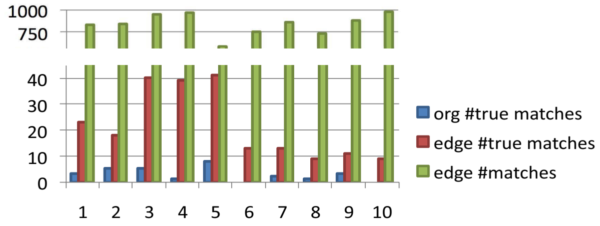


Figure 4. Number of matches and true matches in the original images and edge response images

2.2. Feature Matching in Two Images

In our data sets, the displacement between images could be large. To account for large displacement, the size of the neighborhood should be large. However, as described earlier, this results in fewer matches, which may decrease the accuracy of the homography estimation. Hence, there is a trade off between the displacement we can account for and the number of matches we can produce.

2.2.1 Iterative Nearest Neighbor Matching

We propose Iterative Nearest Neighbor Matching algorithm to solve the problem. If we have a rough estimate of the $H_{i,j}$, a feature at x in I_i can have a higher chance to be matched by performing NN-Match in a smaller neighborhood which centers at $H_{i,j}(x)$. With more matches, a more accurate homography can be estimated. Our approach is illustrated in algorithm 1. It starts with a large search area with radius $r = r_{max}$ and assume the homography $H_{i,j}$ as identity. At each iteration, every feature $x \in \Gamma_i$ is matched to its nearest neighbor in Γ_j where the matching candidates are in the search area centers at $H_{i,j}(x)$ with radius r. Then RANSAC is employed on the matches to estimate a more accurate homography for the next iteration. It iterates until the smallest search area $r = r_{min}$ is reached. Although this approach looks similar to ICP algorithm, it is unique as the estimated homography in the previous iteration can be used in the next to narrow down the search region and produce more reliable matches.

```
input  : Features $\Gamma_i$ and $\Gamma_j$ with respect to image
         $I_i$ and $I_j$
output : Homography $H_{i,j}$ and matches $M_{i,j}$
initialize $H_{i,j} \leftarrow I, r \leftarrow r_{max}$;
while $r > r_{min}$ do
    $M_{i,j} \leftarrow$ NN-Match ( $\Gamma_i, \Gamma_j, H_{i,j}, r$ );
    $H_{i,j} \leftarrow$ RANSAC ($M_{i,j}$);
    $r \leftarrow r/2$;
end
```

Algorithm 1: Iterative Nearest Neighbor Matching

2.2.2 RANSAC Inlier Score

We also made modifications to RANSAC. Let H and HI denotes a homography and its corresponding inliers respectively. In standard RANSAC, the best homography is the one with the largest number of inliers. In other words, the best homography is the one with the highest score which is defined as

$$Score(H) = |HI| \tag{1}$$

However, as shown in Figure 4, there are some cases where the true matches are few. In such cases, a good homography estimate does not have significantly more inliers than a poor one. Worse, two homographies have same amount of inliers and using eq(1) do not update the homography estimate from one to a better one. We solve this issue by further assigning each inlier a weight. A higher weight is given when the orientations of the two matched SIFT features are consistent with respect to the homography estimate:

$$G(m) = CI(H_{i,j}(Ort(x_i)), Ort(x_j)) \tag{2}$$

where $Ort(\cdot)$ is the orientation of a SIFT feature, and $m = (x_i, x_j)$ is a pair of match of features in Γ_i, Γ_j repectively. CI is the consistency indicator function. We use a Gaussian function with $\mu = 0, \sigma = 0.5$ for CI. Then the new score for a homography is defined as:

$$Score(H) = \sum_{m^k \in HI} G(m^k) \tag{3}$$

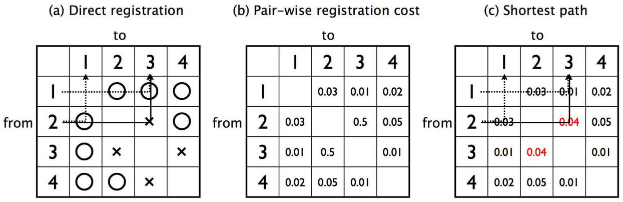
where m^k is the k-th homography inlier in HI.

We also want to bias inliers that spread across the entire image since with the same amount of inlier correspondences, those that are more uniformly distributed contribute less error to the homography estimate. Therefore, matches that are densely congregated should be given lower weights. Let $D(\cdot)$ be the function that measures the inlier density around an inlier, we have each inlier m_k weighted by $1/D(m_k)$. From 3, we have the third edition for $Score(H)$:

$$Score(H) = \sum_{m^k \in HI} G(m^k)/D(m^k) \tag{4}$$

2.3. Chained Registration

Chained registration is a global optimization approach after all the images are pairwise registered. With the pairwise registration result, images can be registered to one another tindirectly. This is illustrated in Figure 5(a). Let a circle and a cross in row i column j denotes a successful and a failed registration from image I_i to I_j respectively. In the example, I_2 cannot be registered to I_3 directly. However, since I_2 can be registered to I_1 directly, and I_1 can be registered to I_3 directly, I_1 becomes the bridge to register I_2 to I_3. We can perform another run of registration, where at algorithm 1 line 1, we initialize $H_{2,3}$ as $H_{1,3}H_{2,1}$ instead of I, and $r = r_{min}$ instead of r_{max}. In other words, we perform a NN-Match directly in a small area that roughly centers at the matching point. In such a way, we can overcome failure pairs that do not have enough true matches for Iter-NNM to converge to a correct homography.



(a) Direct registration

from / to

	1	2	3	4
1		○	○	○
2	○		×	○
3	○	×		×
4	○	○	×	

(b) Pair-wise registration cost

from / to

	1	2	3	4
1		0.03	0.01	0.02
2	0.03		0.5	0.05
3	0.01	0.5		0.01
4	0.02	0.05	0.01	

(c) Shortest path

from / to

	1	2	3	4
1		0.03	0.01	0.02
2	0.03		0.04	0.05
3	0.01	0.04		0.01
4	0.02	0.05	0.01	

Figure 5. Chained registration

2.3.1 Cost Function for Chained Registration

To select the best images for chained registration, we use the shortest path algorithm, in which each node represents an image. The internal nodes along the shortest path from i to j are the images used for chained registration. The edge cost $C_{i,j}$ is defined as the inverse of $Score(H_{i,j})$:

$$C_{i,j} = 1/Score(H_{i,j}) \tag{5}$$

Figure 5(b,c) gives a simple example. (b) is the cost table that results from direct pair-wise registration, in which the registration from 2 to 3 has a lower score (higher cost). To improve it, we run an additional direct registration in which we initialize the homography as $H_{1,3}H_{2,1}$ since I_2, I_1, I_3 is the shortest path from I_2 to I_3.

We run such a registration over every image pair (I_i, I_j). Every time $Score(H_{i,j})$ increases, all image pairs that have their shortest paths going through edge $\overline{I_iI_j}$ are re-registered again since the initialization may improve. The process continues until no image pairs get higher homography scores. Note that the final $Score(H_{i,j})$ could still be low so that the corresponding shortest path goes through other nodes. In such a case, we use the chained homographies along the shortest path as the best homography estimate.

Finally, we use all-pair shortest path algorithm on the set of images [6]. The image with the least total shortest path cost to all the other images is selected as the reference frame.

$$ref = \arg\min_i \sum_j Cost_{ShortestPath}(I_i, I_j) \tag{6}$$

2.3.2 Experimental Results

We compare three methods, the NN-Match, the iterative NN-Match and the iterative NN-Match + chained registration over 10 data sets. Figure 6 plots the successful registration rate. It shows that by using chained registration, most of the data sets' registration rate can be improved to 100%.

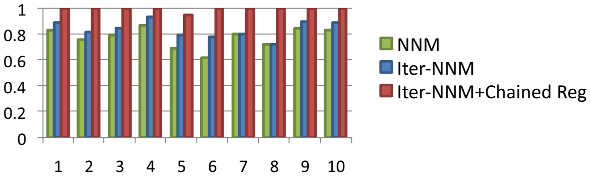


Figure 6. Registration rate comparison of NN-Match, Iterative NN-Match and Iterative NN-Match + Chained registration

We perform another thorough test over 40 data sets. Among all 574 image pairs which contain large image variance in terms of intensity, scale, and displacement, only 16 images are not registered (97.2% registration rate). The average time required for an image set with 20 images is 40 minutes. On the machine with the same computational power, the Y feature based registration method takes 300 minutes. Figure 7 shows one of the difficult cases for which our method can still produce accurate registration. (a) and (b) are two images significantly different from each other. (c) is the registration result shown in a checkerboard view.

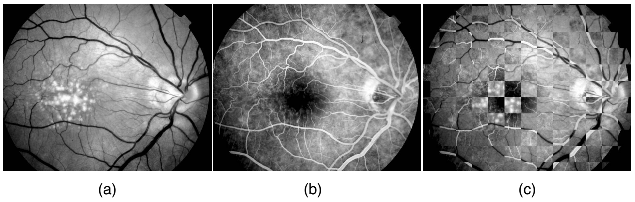


Figure 7. Registration of 2 images with large modality difference

Figure 8 shows one of the failure cases. (a) is the intensity histogram of the original image. We can see the entire image information is squeezed within 10 discrete levels. (b) and (c) are the results of the original image thresholded at 9 and 11 respectively.

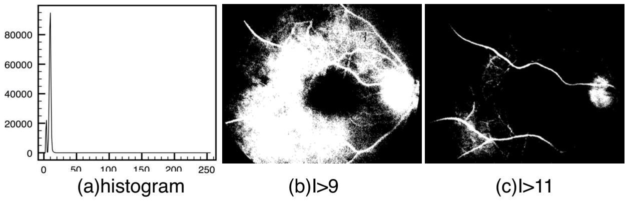


Figure 8. Poor dynamic range

3. 3D Registration and Reconstruction of Near-Planar Surfaces

Other than registration rate, the performance of image registration methods can be measured by the residual error. However, this is only true when the points in the image are co-planar. In our case, the retinal surface is only near planar and it contributes residual error in a 2D registration. We need to register the image sequence in 3D, where a lower residual error represents a better registration.

3.1. Camera Pose Estimation Using 4-Pass Bundle Adjustment

Our approach to the 3D registration is as follows. After 2D registration, the reference frame I_{ref} is determined by the all pair shortest path algorithm. For each image I_i in the sequence, the fundamental matrix F_i with respect to I_{ref} is estimated using the plane+parallax method [11]. The corresponding fundamental matrix inliers, denoted as FI_i, are input to a 4-pass bundle adjustment (4P-BA) to accurately estimate the camera poses as illustrated in Figure 9.

In the first-pass bundle adjustment, we assume the structure X_i on the plane frontal parallel to the reference camera, and estimate the camera pose by minimizing the re-projection error:

$$E(P_i) = \sum_{m^k \in FI_i} \|x_i^k - P_i X_i^k\|^2 \tag{7}$$

where $m^k = (x_i^k, x_{ref}^k)$ is the k-th fundamental matrix inlier consist of features in Γ_i, Γ_j respectively. $P_i = K[R_i|T_i]$, where K is the internal camera parameter.

In the second-pass bundle adjustment, we fix the camera pose and estimate the 3D structure X_i by triangulation.

In the third-pass bundle adjustment, we refine both the camera pose and the 3D structure again by minimizing the re-projection error:

$$E(P_i, X_i) = \sum_{m^k \in FI_i} \|x_i^k - P_i X_i^k\|^2 \tag{8}$$

where P_i and X_i are initialized using the output from the first-pass and second-pass bundle adjustment respectively.

Let I_{best} denotes the best image with respect to I_{ref} for estimating the geometry. After we have estimated the poses of all cameras, we select the camera that has the widest baseline to the reference camera (illustrated as the green camera in Fig. 9) and use the associated image as I_{best} to estimate the sparse structure. We fix this structure and refine all other camera poses in the fourth-pass bundle adjustment.

To estimate the 3D surface, we compute the dense correspondences of image I_{ref} and I_{best}. First the images are rectified using the algorithm that minimize re-sampling effects [9]. Then we compute the disparity map using window based stereo matching [14], in which mutual information [16] is used as the similarity measurement of two windows. The surface can be triangulated the same way as we reconstruct the sparse structure in the fourth bundle adjustment.

Finally, with all camera poses and the 3D surface, images are registered to the reference by back projection. Let x_{ref} be the projection of X, the back projection function is

$$X = bp(x_{ref}) \tag{9}$$

and image I_i is registered to I_{ref} by

$$x_i = P_i X = P_i bp(x_{ref}) \tag{10}$$

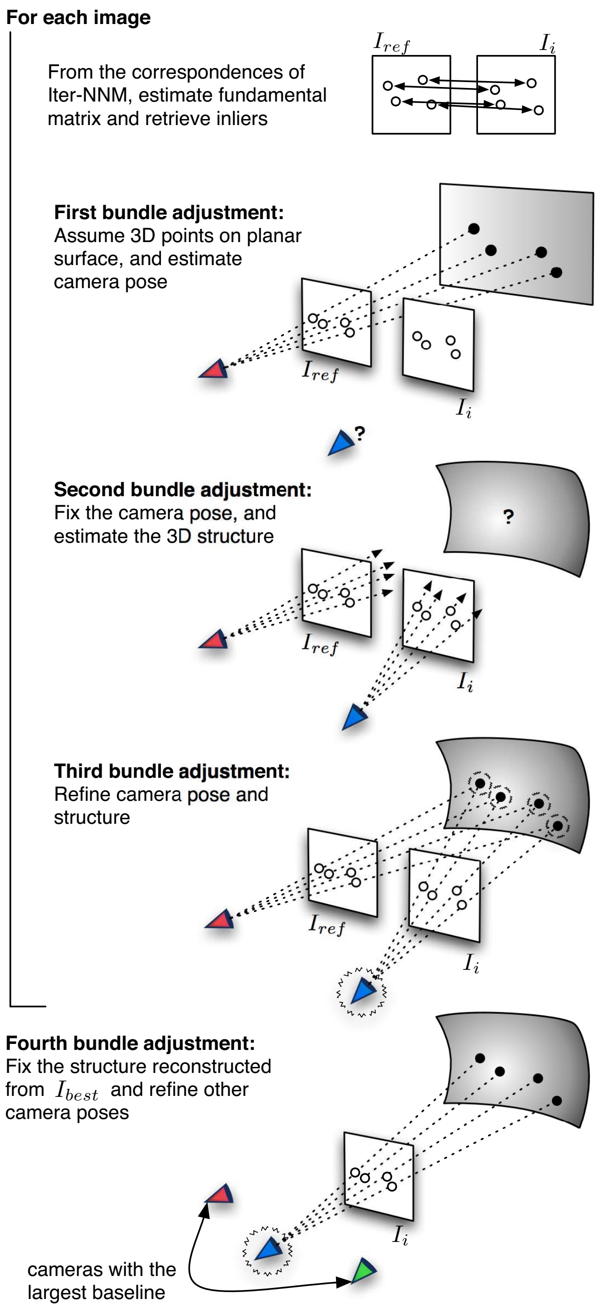

For each image
From the correspondences of Iter-NNM, estimate fundamental matrix and retrieve inliers
I_{ref}
I_i
First bundle adjustment: Assume 3D points on planar surface, and estimate camera pose
I_{ref}
I_i
?
Second bundle adjustment: Fix the camera pose, and estimate the 3D structure
?
I_{ref}
I_i
Third bundle adjustment: Refine camera pose and structure
I_{ref}
I_i
Fourth bundle adjustment: Fix the structure reconstructed from I_{best} and refine other camera poses
I_i
cameras with the largest baseline

Figure 9. Flow chart of 4 pass bundle adjustment

Although our proposed approach looks similar to [6], they are different in the following aspects and ours produces significantly better results:

1. Accurate fundamental matrix estimates. Our SIFT correspondences are significantly more numerous than Y correspondences, and produce more accurate fundamental matrix estimates.

2. Accurate camera pose and 3D surface estimates. The

objective of the 4-pass bundle adjustment is to estimate accurate poses of all cameras. Using the image associated with the camera of the widest baseline, we can reconstruct the sparse structure and the 3D surface with minimum error. Choe's method uses the image with the best 2D registration score with respect to I_{ref}. It is obvious that when two images are identical, the registration error is zero, but carries no 3D information at all, thus unsuitable for 3D reconstruction.

3. Accurate registration. In the fourth bundle adjustment, the cameras are refined using the accurate sparse structure. Together with the accurate 3D surface, images can be registered with minimum error. In Choe's method, both the 3D surface and the camera poses are not optimal, and the 3D registration can be inaccurate.

We will see in the experiment that using the camera with the largest baseline minimizes the error in terms of camera pose, 3D structure and the 3D registration.

3.2. Experimental Results

The first experiment evaluates the accuracy of the recovered camera poses. As depicted in Figure 10, we synthesize a near-spherical surface that mimics the shape of the retina, and create a reference camera with 0 translation and 0 rotation that faces the z-direction where the surface has a depth of 1000. We then create 4 sets of images that are translated 10, 20, 40 and 80 away from the reference camera respectively that cover the range in the real image data. Each set has 4 images that are translated in different directions on the X-Y plane. The cameras are rotated respectively so that a large area in the two images can overlap.

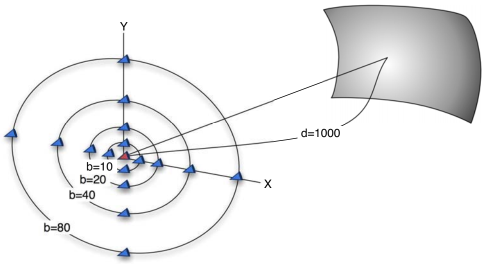


Figure 10. The camera configurations of the experiment. b and d are the length of the baseline and depth respectively

To understand the camera pose error with respect to the baseline, we use each camera as the best camera and employ 4P-BA to compute the corresponding average error of all cameras. Given the true camera translation and rotation, the relative error of the estimated translation and rotation are determined by $E_{trans} = \|t_{true} - t\|/\|t\|$ and $E_{rot} = \|r_{true} - r\|/\|r\|$ respectively. Figure 11(a) and (b) plots the average relative error of the 4 sets of images. It shows that the camera pose recovered by 3P-BA has error rate less than 7.5% when the translation of the camera is 80. It also shows that cameras with the largest baseline ($b = 80$) have the lowest error.

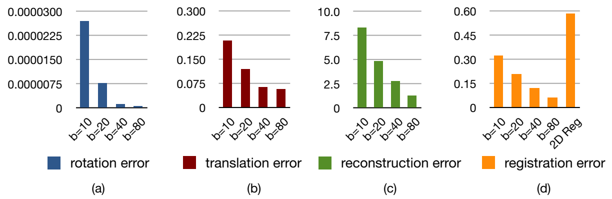


Figure 11. Errors with respect to different camera baselines

The second experiment evaluates the reconstruction error with respect to different camera poses using the same 16 images. For each image, we recover the corresponding camera pose and estimate the corresponding 3D surface. We then compute the average error of the 3D surface with respect to the baseline. Given the true position of the 3D structure, the average error of the estimated structure is determined by $E_{struct} = \frac{1}{n}\sum_{j=1}^{n}(\|X^j_{true} - X^j\|)$ where n is the number of points. The result is shown in Figure 11(c). The error decreases significantly when the baseline increases.

The third experiment compares the registration errors of 3D registration and 2D registration. We use the same data set. For each image, we use it as I_{best} to estimate all camera poses P^i and the corresponding 3D surface. Then we project the ground truth 3D points and the estimated 3D points using P^i_{true} and P^i respectively and compute the projection error in pixels. We also compute the average 2D registration error. The result is shown in Figure 11(d). The back projection error drops significantly when baseline increases.

Finally, we compare our method with Choe's on the reconstructed surfaces of real image sets. Without ground truth surface data, we measure the average local depth variance of the surface since a smooth shallow surface should have a relatively low variance in depth. The local depth variance is computed at every dense point with a neighborhood of size 100. The result is shown in Table 1. In all 9 image sets, we produce more near planar surfaces. Particularly in case 1 and 9, the surfaces reconstructed using Choe's method are totally wrong and result in extremely high variance in depth. There is only one case where Choe's method selects the same image for 3D reconstruction. Figure 12 shows some of the cases. In each row, the left and right surfaces are the reconstructed results from the same view point using our approach and Choe's respectively. The surfaces on the right side are poorly reconstructed. Our surfaces, on the other hand, are smooth and slightly curved just like the way retinal surfaces should be. Our ability to recover the true 3D shape of an object is very important to retinal diagnosis and other operations where 3D volumes reveal information that images do not.

	1	2	3	4	5	6	7	8	9
3P-BA	28.32	10.78	4.13	15.43	6.59	5.56	3.48	8.14	636.18
4P-BA	4.89	6.68	4.03	3.93	6.59	3.05	2.45	5.01	4.68

Table 1. Average local depth variance comparison of the reconstructed surfaces

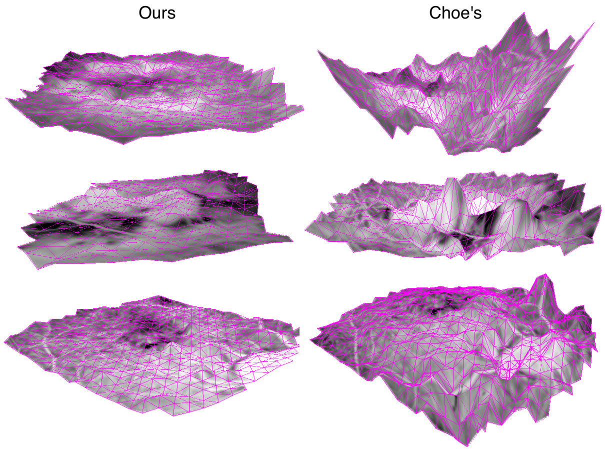


Figure 12. Comparison of the 3D retina surface reconstruction

4. Conclusions

We have presented a 2D registration method that register multi-modality images accurately with high registration rate and a 3D surface reconstruction method that recovers accurate camera poses and 3D structure for 3D registration.

In 2D registration, SIFT features are extracted from edge maps and iteratively matched to produce a larger set of reliable matches from which an accurate homography can be estimated. The chained registration further optimized the pairwise registration result globally by registration in an indirect way. Compared with state of the art 2D retinal image registration methods, ours achieves a near perfect registration rate over a massive real data set, while the time required is much less.

In 3D registration, we use 4P-BA to recover accurate camera poses, which gives us the clue to select the best image associated with the camera with the largest baseline. Using the best image, both the 3D surface and the camera poses can be accurately recovered, and the 3D registration by back projection is therefore accurate. Compared with state of the art near planar surface reconstruction method, we produce more accurate surfaces in all data sets.

Our future work includes validating the approach on a larger data set, increasing the 3D registration rate, and 3D surface reconstruction from multiple views.

Acknowledgement

This research was supported by the National Institutes of Health (NIH) under grant No. R21 EY015914-03. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. We would like to thank the Doheny Eye Institute for providing data sets and clinical validation of the results.

References

[1] M. Brown and D. G. Lowe. Recognising panoramas. *ICCV '03*, vol. 2, pp. 1218–1225. 2

[2] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. *CVPR '05*, vol. 2, pp. 60–65. 3

[3] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(3):347–364, 2002. 2

[4] P. Cattin, H. Bay, L. V. Gool, and G. Szekely. Retina mosaicing using local features. *MICCAI*, 4191:185–192, 2006. 2

[5] T. E. Choe and I. Cohen. Registration of multimodal fluorescein images sequence of the retina. *ICCV '05*, vol. 1, pp. 106–113. 1, 2

[6] T. E. Choe and G. Medioni. 3-d metric reconstruction and registration of images of near-planar surfaces. *ICCV '07*, vol. 1. 2, 5, 6

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981. 2

[8] J.-M. Frahm and M. Pollefeys. Ransac for (quasi-)degenerate data (qdegsac). *CVPR '06*, vol. 1, pp. 453–460. 2

[9] J. Gluckman and S. Nayar. Rectifying transformations that minimize resampling effects. *CVPR '01*, vol. 1, pp. 111–117. 6

[10] A. Kelman, M. Sofka, and C. V. Stewart. Keypoint descriptors for matching across multiple image modalities and nonlinear intensity variations. *CVPR '07*. 3

[11] R. Kumar, P. Anandan, and K. Hanna. shape recovery from multiple views: A parallax based approach. *DARPA IU Workshop, Monterey, Calif.*, Nov. 1994. 6

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. 1

[13] A. M. Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst.*, 38(1):35–44, 2004. 3

[14] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *SMBV '01*, pp. 131–140. 6

[15] C. V. Stewart, C.-L. Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans on Medical Imaging*, 22(11):1379–1394, Nov. 2003. 1, 2

[16] P. Viola and I. William M. Wells. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, 1997. 1, 6