Automatic 3D Change Detection for Glaucoma Diagnosis

Lu Wang, Vinutha Kallem, Mayank Bansal, Jayan Eledath, Harpreet Sawhney
SRI International
Princeton, NJ, USA
jayan.eledath@sri.com

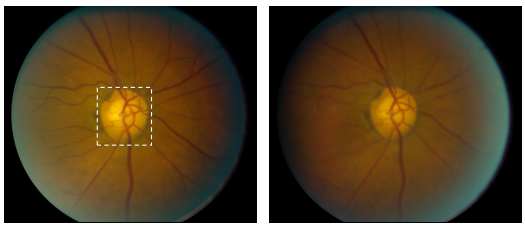
Denise J. Pearson, Richard A. Stone
Scheie Eye Institute, University of Pennsylvania School of Medicine
Philadelphia, PA, USA
stone@mail.med.upenn.edu *

Abstract

Important diagnostic criteria for glaucoma are changes in the 3D structure of the optic disc due to optic nerve damage. We propose an automatic approach for detecting these changes in 3D models reconstructed from fundus images of the same patient taken at different times. For each time session, only two uncalibated fundus images are required. The approach applies a 6-point algorithm to estimate relative camera pose assuming a constant camera focal length. To deal with the instability of 3D reconstruction associated with fundus images, our approach keeps multiple candidate reconstruction solutions for each image pair. The best 3D reconstruction is found by optimizing the 3D registration of all images after an iterative bundle adjustment that tolerates possible structure changes. The 3D structure changes are detected by evaluating the reprojection errors of feature points in image space. We validate the approach by comparing the diagnosis results with manual grading by human experts on a fundus image dataset.

1. Introduction

Glaucoma, the second leading cause of blindness in the United States and world-wide, results in vision loss because of a characteristic form of atrophy of the optic nerve. Stability or change in the 3D structure of the optic disc is an important criterion in the evaluation and treatment of glaucoma. Traditionally, a clinician obtains a 3D perception of the optic disc surface by looking through a stereoscope at two fundus images taken from slightly different viewpoints. Fig. 1 shows a typical pair of stereo fundus images where the optic disc is the bright circular region in the image center, highlighted in the left image. By alternating between two stereo images taken at different times, the clinician can make a judgment if the optic disc area does or does not have interval structural changes, including 3D changes. However, this judgment is subjective and difficult because of variations in stereo baseline, alterations in camera pose, etc.



(a) (b)

Figure 1. A typical pair of stereo fundus images. The bright circular region in each view and highlighted in the left image is the optic disc.

In recent years, emerging technologies such as OCT are available for *in vivo* 3D ocular imaging with high resolution, but OCT instruments are expensive, continuously evolving with instrument-specific image formats, require special capture sessions, and are still being investigated for improved glaucoma diagnosis. Up to now, fundus photography remains the "golden standard" for glaucoma studies and diagnosis, despite the limits of photography. To overcome the subjective nature in traditional photograph-based diagnosis and to develop a convenient, diagnostically accurate, format-independent method to assess the optic disc in glaucoma, we propose a quantitative approach to evaluate changes in 3D structure based on computer vision. The approach requires only two stereoscopic fundus images captured at two different times in an individual subject. The

*This work is supported by NIH Grant R01-EY017299, the Paul and Evanina Bell Mackall Foundation Trust (RAS), and Research to Prevent Blindness (RAS).

fundus images can be uncalibrated with unknown intrinsic and extrinsic parameters, the typical clinical scenario.

3D retinal reconstruction from stereoscopic fundus images is very challenging due to several reasons. First, traditional stereo matching algorithms have great difficulties in matching fundus images because of low-texture, low-contrast, image blur, non-Lambertian reflectance, and noise from the illumination conditions [10]. Second, most of the retinal surface in a fundus image is nearly planar except the optic disc area [1, 4]. The scene of a flat plane is a degenerate case for estimating epipolar geometry. Third, fundus images are usually uncalibrated, and only a quite limited number of archived images may exist in the clinical record of a patient. Besides these challenges in 3D reconstruction, detecting the subtle changes related to glaucoma progression from the reconstructed 3D models poses additional difficulties.

1.1. Previous work

Most prior works on image-based computer-assisted glaucoma diagnosis focused on computing 2D image features that are indicative of the disease, such as the ratio between the area of optic disc and the area of optic cup [15, 14]. However, accurate segmentation of optic disc and optic cup from 2D images can be very difficult depending on the illumination condition and the camera viewpoint. In addition, some important information such as the changes in the 3D depth and slope of the optic cup and the 3D position shift of blood vessels due to optic nerve damage cannot be retrieved with these 2D approaches.

3D reconstruction from stereo fundus images has been studied in [10, 6] in which the images come from fixed-base stereo fundus cameras with known epipolar geometry. However, simultaneous stereo fundus cameras are not available in most clinics. The most cost-efficient way of obtaining stereo fundus images is by taking images sequentially with a monocular fundus camera from different viewpoints, in which camera focal length and the epipolar geometry between images are unknown.

In [1], Choe and Medioni proposed a 3D metric reconstruction approach for uncalibrated fundus images. Due to the near-planar shape of retinal surface, the approach exploits a plane+parallax strategy [3] for epipolar geometry estimation. The 3D model of the retinal surface, the camera focal length and camera pose of each image are estimated through a 3-pass bundle adjustment. The initial 3D model of the retinal surface input to the bundle adjustment is assumed to be a flat plane parallel to the image plane of the reference image and with the same size of the image. However, depending on the actual camera focal length and the view angle of the reference image, this approach may yield a bad initial solution and cause the bundle adjustment to snag in a local minimum. In addition, [1] did not address the problem of detecting changes in the reconstructed 3D models for disease diagnosis.

Kuthirummal et. al. [4] proposed a change detection algorithm for glaucoma diagnosis by aligning 3D models reconstructed from uncalibrated fundus images. Similar to [1], the plane+parallax strategy is used for epipolar geometry estimation. However, instead of achieving a metric up-to-scale 3D reconstruction, the approach only obtains a projective reconstruction in which the 3D model differs from the ground truth by an unknown projective transformation [2]. To align and compare reconstructed 3D models from different time sessions, a 3D homography is estimated between the two projective 3D models. By computing the distance from a 3D point in one 3D model to its closest point in the other model after alignment, a changed 3D point is identified if the distance is larger than a threshold.

There are two significant problems with the above approach in [4]. First, due to the inherent difficulty in 3D reconstruction from two fundus images, the projective 3D model reconstructed separately for each time session can have significant errors, often leading to significant alignment error between 3D models. Therefore, the distance between corresponding 3D points can be caused by inaccuracy in the reconstructed 3D models and alignment error between the models instead of actual structure changes in the optic disc. Bundle adjustment has been widely applied in multi-view camera geometry [2] that does non-linear optimization to the initial 3D reconstruction and all the camera parameters jointly, but it is not applied in [4] because the 3D reconstruction is not metric. Another problem is the uncertainty in selecting an optimum threshold on the 3D distance for change detection since the distance is defined in an unknown projective space. Therefore, instead of automatically generating a diagnosis decision, the system only provides a tool for visualizing possible changes by manually varying the threshold.

1.2. Overview of our approach

To overcome the difficulty of epipolar geometry estimation for a near-planar surface, we exploit an additional constraint based on the fact that camera focal length is usually kept constant during each time session, even if different sessions can have different focal length. A RANSAC process based on the 6-point algorithm in [9] is applied for estimating the camera focal length and the relative pose between two fundus images from a set of feature correspondences. After triangulation, metric 3D reconstruction can be obtained.

Because multiple possible solutions exist based on the 6-point algorithm [9] and also for making the approach less sensitive to noise from feature matching, instead of keeping a single best solution as in the traditional RANSAC process, we construct a list of top candidate reconstruction so-

lutions for a pair of fundus images. To select the top candidates, not only the number of point match inliers but also the shape of the 3D model is considered since we have a prior knowledge of the possible shape of a retinal surface. For each possible 3D reconstruction, the camera poses of all the images from different time sessions are estimated relative to the 3D model. Bundle adjustment is then applied to optimize the 3D model, camera focal length, and camera pose of each image jointly. To handle possible 3D structure changes between different time sessions, in the bundle adjustment we distinguish feature matches across all views and feature matches for each separate time session.

In the end, the optimal 3D reconstruction is selected as the one with the best 3D registration across all views. Since the reconstruction is up to an unknown scale, unlike the approach in [4] that sets a threshold on the distance between corresponding points in 3D space, we identify changed feature points by examining their reprojection errors in each image. To evaluate the performance of the approach, the automatic diagnosis results are compared with the manual gradings by glaucoma experts on a fundus image dataset consisting of 7 eyes with interval glaucoma changes and 14 eyes without changes, in which the precision and recall rates of our approach are 75% and 85.7% respectively. To the best of our knowledge, ours is the first work based on 3D analysis that reported statistical results of automatic glaucoma diagnosis.

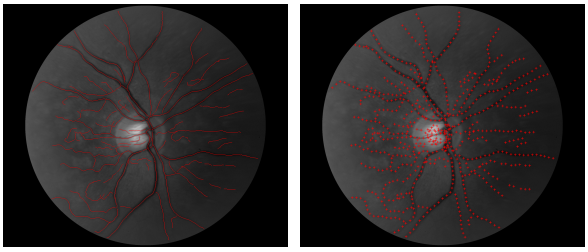
The remainder of the paper is organized as follows: in Section 2 we describe the feature matching between fundus images. Section 3 gives details on the metric 3D reconstruction from two fundus images with RANSAC based on the 6-point algorithm. Section 4 introduces the bundle adjustment process for aligning all the images in 3D with possible structure changes. Section 5 presents the change detection criteria. In Section 6, some experimental results and discussion are given, and the paper is concluded in Section 7.

2. Image matching

Detecting feature correspondences between images is usually the first step for image-based 3D reconstruction. Image matching for fundus images is challenging because of low-texture, non-Lambertian reflectance, large illumination variations, and image blur due to eye movement and camera defocus. Several techniques exist for matching fundus images. In [1], 'Y' features and mutual information based template matching are used. A multi-scale stereo matching approach is presented in [10] based on a modified normalized cross-correlation. While any of the matching techniques for fundus images can be incorporated into our 3D reconstruction and change detection pipeline, in the present work we use a two-step approach similar to that in [4] for matching two fundus images.

The approach is based on the observation that in a fundus image the retinal surface is nearly planar outside of optic disc. Two images of a 3D plane are related by a homography transformation [2]. Therefore, the first step of the approach is to estimate the homography from a set of SIFT features [5] detected outside of the optic disc area. To aid this, the approach in [12] is used for optic disc segmentation. Based on the estimated homography, one of the images is warped to the other. The remaining image parallax is due to 3D structures off the plane.

As noted in [4], SIFT matches are typically not dense and accurate enough for 3D reconstruction from low-textured fundus images. Therefore after correcting the global image transformation with the homography warping, Normalized Cross-Correlation (NCC) is used for more accurate feature matching. In [4], Harris corners are used as feature points. However, due to low-textureness, Harris corners are usually not well distributed in a fundus image. Since blood vessels provide the major image texture, we uniformly sample image points (a point for every 15 pixels for images with 2000x1712 pixels in our experiments. Note that all parameters of our approach are chosen empirically.) along the central lines of vessels as feature points. Vessel fragments are extracted automatically with an approach similar to [11]. In addition, Harris corners are also detected inside the optic disc area that capture features from non-vessel structures. Fig. 2 shows an example of the blood vessel extraction and the feature points.



(a) (b)

Figure 2. Left: automatically extracted blood vessel fragments. Right: feature points that are on vessel fragments or that are strong Harris corners inside the optic disc area.

For a feature point in one of the images, its corresponding point in the other image is the pixel with the highest NCC matching score in its neighborhood. Since only the optic disc area has significant parallax after the image warping, the required search neighborhood is smaller for the feature points outside the optic disc than that required for the inside points (10x10 and 25x25 pixels respectively in our experiments).

3. Metric 3D reconstruction from two images

The near-planarity of the retinal surface makes accurate epipolar geometry estimation difficult since a planar scene

is a degenerate case for estimating epipolar geometry. The plane+parallax strategy is used in [1, 4] to deal with this problem. However, it is still sensitive to the noise in feature matching since most parallax is from the optic disc area that is usually a small part of the whole fundus image . Therefore, we exploit additional constraints on camera intrinsic parameters to make the estimation more robust and accurate. Another important advantage of our approach is that an estimation of camera focal length and a metric 3D reconstruction can also be obtained.

Similar to [1] and many structure-from-motion approaches [8], we assume that a fundus camera has square pixels, no skew and with the principle point at the image center. In other words, the only unknown camera intrinsic parameter is the camera focal length. In addition, although camera focal length can vary between different time sessions, we assume it is fixed during each session. This is true in most cases since fundus stereo images are typically captured sequentially without varying the zoom setting with the same session. With these assumptions, from 6 point correspondences between two images, the camera focal length and the fundamental matrix can be calculated with the approach in [9]. However, there can be up to 15 possible solutions. From each solution, the relative camera pose between the two images can be computed [2], and further a 3D point cloud can be reconstructed from the point matches after triangulation [2]. This reconstruction is metric with only an unknown scale.

Given a set of detected point matches containing possible outliers, RANSAC [2] based on the 6-point algorithm is applied to estimate the epipolar geometry. To avoid the degenerate case in which all 6 points are on a plane, each random sample of 6 point matches has at least one point match from the optic disc area.

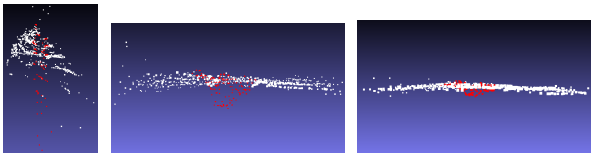
Since the solution of the 6-point algorithm [9] is not uniqueor, and also for making the approach less sensitive to the noise and errors in point matches, unlike the regular RANSAC process in which only a single best solution is selected, we generate a list of top candidates. Each candidate solution is then used as a seed to initiate the bundle adjustment described in Section 4 that does 3D alignment across all time sessions. In the end, the optimal solution is selected as the one that achieves the best 3D alignment.

3.1. Shape analysis for selecting candidate 3D reconstruction solutions

To select the top candidate solutions during the RANSAC process, not only the number of inliers among the point matches is a selection criterion, we also check if the reconstructed 3D model looks like a retinal surface, a nearly planar surface with an appropriate "cup shaped" optic disc in the middle. This model represents the most typical retinal structure near the optic disc area, and is also used in prior works by other researchers such as [1].

First, Cheirality check is applied to remove any solutions in which not all of the 3D points are in front of both cameras. A 3D plane is fit to the 3D points outside of the optic disc area. The thickness of the plane is the average distance of the 3D points to the plane. After projecting the 3D points onto the plane, the width and height of the plane are defined to the length of the major and minor principle axes of the 2D projections respectively. A reasonable plane of the retinal surface should be thin and have an aspect ratio close to that of the fundus image. In our experiments, if $thickness/height > 0.1$ or the aspect ratio $height/width < 0.4$, the 3D plane is regarded as out of shape and the solution will be discarded. Note that the parameters used for the model fitting in our approach allow considerable deviation from the ideal model so that the prior model constraint is relatively relaxed.

Since the optic cup is a depression in the center of the optic disc and projects away from the camera center in each fundus image, to see if the 3D points inside the optic disc form the general shape of a cup, we check if most of them are off the plane and are on the different side of the plane relative to the camera centers. A 3D point is off the plane if its distance to the plane is larger than the plane thickness. In our experiments, if over 30% 3D points inside the optic disc area are off the plane and on the different side than the camera centers, and less than 10% points are on the same side as the camera centers, we say the optic disc forms a cup; otherwise, the reconstruction solution will be discarded.



(a) (b) (c)

Figure 3. Three possible 3D reconstruction solutions. Left: a reconstruction solution with an unreasonable shape. Middle: a reconstruction solution with focal length = 2666 pixels. Right: a reconstruction solution with focal length = 5257 pixels. In each image, the red and white points are points inside and outside the optic disc area respectively, and the viewing direction is parallel to the plane of the retinal surface.

The process for selecting the best reconstruction solutions can be summarized as follows. During the RANSAC, we keep the top 100 solutions that have the most number of point match inliers that satisfy the estimated epipolar geometry. To avoid keeping very close solutions, we require the candidate solutions that have sufficiently different focal length to each other (10% difference in our experiments). Out of the 100 solutions, the ones that do not pass the shape verification will be discarded. On average, about 20 candidate solutions will be kept for each time session.

Fig. 3 illustrates an example of possible reconstructions from two stereo fundus images. The left model is out of shape and will be discarded even if this reconstruction solution has the most number of point inliers. The middle and the right reconstruction solutions have reasonable shape and almost the same number of inliers. However, they have very different camera focal length (2666 pixels and 5257 pixels respectively), therefore it is necessary to keep both of them as candidates.

4. 3D alignment with bundle adjustment

In order to detect if there are structure changes between two time sessions, we need to align the 3D reconstructions from different sessions in the same space. This is based on the assumption that a significant part of the retinal surface especially outside of the optic disc area is unchanged and can be used for alignment. Since each reconstructed 3D model may have significant errors, directly aligning two 3D models as in [4] is difficult. Instead, we select the reconstructed 3D model of one time session as the reference, and estimate the camera pose of each image in the other session relative to it. The 3D model, camera focal length and camera pose of all images are then jointly optimized in an iterative bundle adjustment process.

Given a 3D model of the first time session and to estimate the camera pose of an image in the second session, we need to have a set of 2D-3D correspondences, which can be obtained by detecting the feature correspondences between the image and a reference image in the first session. In addition, we need to know the camera focal length of the image. The possible values of this focal length are given by the candidate 3D reconstructions of the second time session. For each possible focal length, the camera pose of the image can be estimated with the 3-point algorithm [7].

Since bundle adjustment is computationally intensive, to reduce computation we do not run bundle adjustment for each combination of a possible 3D model of the first session and a possible focal length of the images in the second session. Instead, we rank all the possible combinations based on the total number of point inliers during the camera pose estimation. Only the top 50 combinations with the most number of inliers are kept.

Bundle adjustment does a nonlinear optimization on the initial 3D point cloud, camera focal length and camera pose by minimizing the reprojection errors of the 3D points in the images. Since there can be structure changes between the two time sessions, some 3D points may only appear in one of the sessions. Therefore, in our bundle adjustment there are 3 groups of 3D points: the points that are shared between both time sessions; those, only in the first session; and those, only in the second session respectively. The shared 3D points constrain the camera pose of all images jointly while the other 3D points only constrain the camera pose of the images in their corresponding session.

Each shared 3D point corresponds to a feature track across all images. To detect the feature tracks, an arbitrary images in the first session is selected as the reference image. The feature points of the feature tracks in the other image of the first session and in the first image of the second session are obtained by matching these images with the reference image. However, the corresponding feature points in the second image of the second session are obtained by matching it with the first image of the second session instead of the reference image because matching between images of the same session is more reliable than that between images of different sessions.

The 3D point associated with a feature track may be an outlier because it may correspond to a structure change between different sessions. Therefore, for each feature track, we also create a 3D point that has corresponding feature points only in the images of the first session, and similarly a 3D point only for the second session. These 3D points place constraints on the camera pose of images in each session separately. They share the same 2D feature points with the feature track in the corresponding images but they are independent 3D points in the bundle adjustment. The initial position of the shared 3D point and the 3D point in the first session are calculated from the feature points in the images of the first session with triangulation while the initial position of the 3D point in the second session is computed from the images of the second session.

In addition, outliers of the 3D points should not be included in the bundle adjustment. A 3D point is an outlier if its reprojection error in one of its associated images is larger than a threshold (5 pixels in our experiments). Initially, the large reprojection error of a 3D point can be caused by inaccuracy of its initial 3D position and camera poses rather than wrong matches. After the bundle adjustment, the 3D points and camera pose are refined and some of the initial outliers can become inliers. Therefore, instead of a single run of bundle adjustment, we do several iterations. At each iteration, we recompute the inliers of the 3D points and rerun the bundle adjustment on the inliers. The iterations stop when the number of inliers cannot be further increased. Each run of the bundle adjustment minimizes the following error:

$$E = \sum_{k=1}^{N}\sum_{i=1}^{4} |P_i X_k^1 - u_{ki}|^2 + \sum_{k=1}^{N}\sum_{i=1}^{2} |P_i X_k^2 - u_{ki}|^2 + \sum_{k=1}^{N}\sum_{i=3}^{4} |P_i X_k^3 - u_{ki}|^2, \quad (1)$$

where the 3 items are the sum of the reprojection errors for the 3 groups of 3D points: shared 3D points, 3D points in the first session and 3D points in the second session respectively. The images are numbered from 1 to 4 with the first

two images from the first session and the last two images from the second session. $P_i = K_i[R_i T_i]$ is the projection matrix of image i. x_k^j is the k-th 3D point in group j, and u_{ki} is its 2D measurement in image i. N is the number of 3D points in each group.

For efficiency, the multi-core implementation of bundle adjustment based on GPU [13] is used in our experiments. After running the iterative bundle adjustment for each of the top 50 combinations of a candidate 3D model in the first session and a candidate focal length of the images in the second session, the best 3D reconstruction is selected as the one with the most number of point inliers.

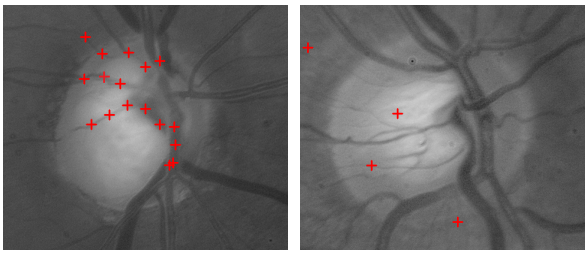
5. Change detection and glaucoma diagnosis

Once we obtain the camera pose of all images in the same coordinate system, one way to detect changes between different time sessions is to do a dense 3D reconstruction for each session and detect the 3D points in the first 3D model whose distance to their closest points in the second model is larger than a threshold, which is similar to the approach in [4]. However, there are 3 problems with this strategy. First, since the 3D reconstruction is up to an unknown scale, it is hard to set a good threshold on the distance. Second, a large portion of a fundus image lacks texture while 3D reconstruction for low-texture image area is not accurate. Third, 3D reconstruction from images with small baseline is sensitive to the noise in feature matching, therefore directly comparing two 3D points in two independently reconstructed 3D models without joint bundle adjustment may not be reliable.

Our approach for change detection is based on the bundle adjustment result obtained in Section 4. Instead of setting a threshold on a 3D distance, we analyze the reprojection errors of the 3D points in 2D images. The outliers among the 3D points shared by both time sessions whose 2D reprojection errors in one of the images are larger than a threshold (3 pixels in our experiments) are good indicators for structure changes. In addition, these feature points are either on vessel fragments or are strong Harris corners, both of which have relatively rich texture.

However, the outliers of the shared 3D points may also be caused by wrong feature matches. To remove these outliers, we first remove the feature tracks that do not satisfy the epipolar geometry between the images of the first session or that between the images of the second session. These outliers are obviously due to wrong matches between the images in the same session. For the rest of the outliers, we find that those due to structure changes usually form clusters in an image since the changed areas usually have a certain size, whereas those due to wrong matches are usually isolated from each other since they are caused by random noise. Fig.4 gives an example. In Fig.4a, the point outliers are dense and form clusters reflecting the structure changes in the optic disc. In Fig.4b, the isolated outliers are due to wrong feature matches instead of structure changes.

Therefore, we cluster the outliers in the reference image based on their image distance. Two outliers are in the same cluster if their distance is less than a threshold (15 pixels in our experiments). The clusters with less than 3 points are removed. In the end, if there are still outliers in the optic disc area in the reference image, we conclude that the patient's optic disc has undergone interval change between the image sets. In addition, the locations of the outliers indicate where the changes are.



(a) (b)

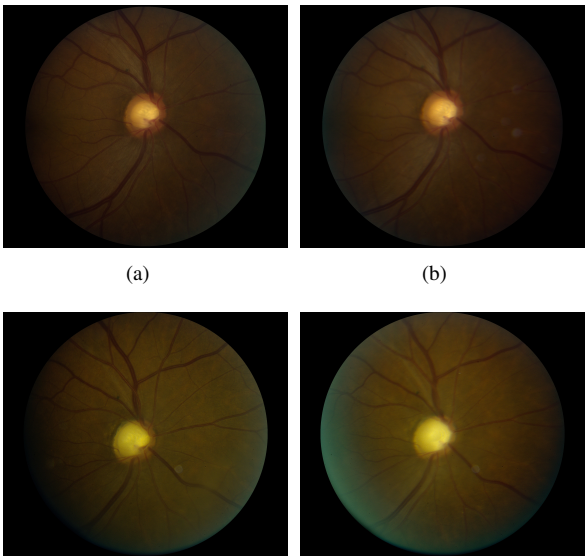
Figure 4. Left: outliers due to structure changes form clusters. Right: isolated outliers are caused by wrong feature matches.

6. Experiments and Discussions

To evaluate the approach, we assessed the automatic diagnosing results against clinical dataset of 21 image sequences. These images were selected and manually graded by consensus by a glaucoma-trained clinical coordinator and a glaucoma subspecialist. Each image sequence has two stereo pairs of fundus images that were taken at two different times of the same eye of a patient. Among the 21 image sequences, 7 sets from different patients were graded as having interval glaucoma changes. The other 14 image sequences, each from different patients, are known to have no changes because they comprised separate stereo image sets acquired on the same day.

The acquisition of these images occurred between 1990 and 2013. The average image quality is worse than that of the images from simultaneous stereo cameras used in [10] with more noise, and inconsistent image blur and large illumination changes between the images in the same stereo pair. However, the quality of the images we studied is typical of those acquired in clinical practice. As an example, Figure 5 shows an image sequence in our dataset with glaucoma changes in the optic disc.

The change detection result of our approach is summarized in Table 1. Out of the 7 images sequences that are graded as having glaucoma progression by the experts, our approach correctly labeled 6 of them as having changes and mistakenly labeled one sequence as having no changes. Out of the 14 image sequences that have no glaucoma changes,



(a) (b) (c) (d)

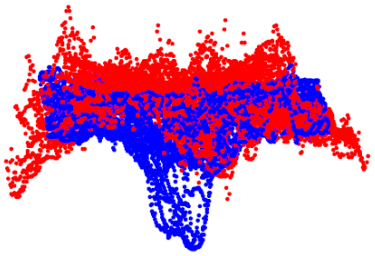
Figure 5. An example image sequence in our dataset. There is glaucoma progress in this sequence. First row: two stereo fundus images taken in 2005. Second row: two stereo fundus images taken in 2006 of the same eye.

the approach made the correct labeling for 12 of them and made the wrong labeling for the other 2 as having changes. Therefore, the precision and recall of our approach for identifying glaucoma progression on this dataset are 75% and 85.7% respectively.

True changes	Detected changes	Missed changes
7	6	1
True stables	Detected stables	Missed stables
14	12	2

Table 1. Change detection result of our approach on a dataset including 7 changed cases with glaucoma progression and 14 stable cases without changes.

To compare our approach with that in [4], we implemented the approach of [4] and ran it on the same dataset. As we mentioned above and also noted in [4], it is not possible to find a fixed threshold on distance in a 3D space up to an unknown perspective transformation for change detection; and therefore the approach in [4] cannot be fully automatic. In addition, we found the approach [4] had difficulty in achieving an acceptable 3D alignment between the reconstructed 3D models from different time sessions for most of the 21 image sequences in our dataset. By rendering the reconstructed 3D models with the approach in [4], we can see significant errors in the 3D alignment for 13 image sequences, and for only 8 of the 21 image sequences the 3D alignment appears reasonable. Figure 6 shows an example of a poor 3D alignment generated with the approach.



(a)

Figure 6. A poor alignment between the 3D models reconstructed from two time sessions with the approach of [4]. The red points are from session 1, and the blue points are from session 2.

No clinically accepted methods exist for automatic change detection of the optic disc in glaucoma. As a pilot study, we view the 75% precision and 85.7% recall of our approach as quite promising for a platform on which to develop a much needed method to improve the diagnosis of glaucoma progression. Future areas to explore include simultaneously acquired optic disc photos, improved camera resolution, external projection of texture on the optic disc through the camera system, fixing the camera focal position and/or assessment of color variation across the optic disc.

7. Conclusion

We present an automatic approach for glaucoma diagnosis based on detecting 3D structure changes in the optic disc area between two time sessions. For each time session, only two uncalibrated fundus images are required. The approach is designed to deal with the low textureness of fundus images, the near-planarity of the retinal surface, and noisy feature matches due to image blur and large illumination variations. Particularly, a two-step algorithm is used to match feature points that are either on vessel fragments or are strong Harris corners. A RANSAC process based on the 6-point algorithm is used to obtain a candidate list of metric 3D reconstruction solutions for each time session. The optimal 3D reconstruction is selected as the one that achieves the optimal 3D alignment after an iterative bundle adjustment. Structure changes are detected by examining reprojection errors in 2D images.

The approach is evaluated on a dataset of 21 image sequences with manual gradings from glaucoma experts as ground truth. The approach achieves 75% precision and 85.7% recall, which is significantly better than existing approaches. Our approach could serve as a platform for developing a clinically useful diagnostic method.

References

[1] T. Choe and G. Medioni. 3d metric reconstruction and registration of images of near-planar surfaces. *ICCV*, pages 1–8, 2007. 2, 3, 4

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2004. 2, 3, 4

[3] R. Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views: a parallax based approach. *DARPA IU Workshop*, 1994. 2

[4] S. Kuthirummal, M. Bansal, H. Sawhney, J. Eledath, D. J. Pearson, and R. A. Stone. 3d alignment and change detection from uncalibrated eye images. *Proc. of IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology*, pages 299–306, 2011. 2, 3, 4, 5, 6, 7

[5] D. G. Lowe. Distinctive image features from scale-invariant key points. *IJCV*, pages 91–110, 2004. 3

[6] T. Nakagawa, Y. Hayashi, Y. Hatanaka, A. Aoyama, T. Hara, A. Fujita, M. Kakogawa, H. Fujita, and T. Yamamoto. Three-dimensional reconstruction of optic nerve head from stereo fundus images and its quantitative estimation. *Proc. of IEEE EBMS*, pages 6748–6751, 2007. 2

[7] D. Nister and J. Stewenius. A minimal solution to the generalised 3-point pose estimation. *Journal of Mathematical Imaging and Vision*, 27(1):67–79, 2007. 5

[8] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *IJCV*, 2007. 4

[9] H. Stewenius, D. Nister, F. Kahl, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. *CVPR*, 2005. 2, 4

[10] L. Tang, M. K. Garvin, K. Lee, W. L. M. Alward, Y. H. Kwon, and M. D. Abramoff. Robust multiscale stereo matching from fundus images with radiometric differences. *TPAMI*, 33(11):2245–2258, 2011. 2, 3, 6

[11] L. Wang, K. Vinutha, B. Mayank, E. Jayan, S. Harpreet, K. Karen, D. J. Pearson, M. D. Mills, G. E. Quinn, and R. A. Stone. Interactive retinal vessel extraction by integrating vessel tracing and graph search. *MICCAI*, pages 567–574, 2013. 3

[12] D. Wong, J. Liu, J. Lim, X. Jia, F. Yin, H. Li, and T. Wong. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali. *IEEE Engineering in Medicine and Biology Society*, pages 2266–2269, 2008. 3

[13] C. Wu, S. Agarwal, B. Curless, and S. Seitz. Multicore bundle adjustment. *MICCAI*, pages 567–574, 2013. 6

[14] J. Xu, O.Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan. Optic disk feature extraction via modified deformable model technique for glaucoma analysis. *Pattern Recognition*, 40(7):2063–2076, 2007. 2

[15] Y. Zheng, D. Stambolia, J. Brien, and J. C. Gee. Optic disc and cup segmentation from color fundus photography using graph cut with priors, 2013. 2