2011 First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology

3D Alignment and Change Detection from Uncalibrated Eye Images

Sujit Kuthirummal, Mayank Bansal, Harpreet Sawhney, Jayan Eledath
SRI International Sarnoff
{*sujit.kuthirummal, mayank.bansal*}*@sri.com*
{*harpreet.sawhney, jayan.eledath*}*@sri.com*

Denise J. Pearson, Richard A. Stone
University of Pennsylvania
denise.pearson@uphs.upenn.edu
stone@mail.med.upenn.edu

***Abstract*—Analyzing change in the 3D structure of the optic disc over time has long been recognized as central to the diagnosis of glaucoma but has been inadequately addressed by computer vision methods. Currently, clinicians examine stereo pairs from different time instants for interval changes indicative of glaucoma. Due to the clinical procedures in capturing optic disc images, these stereo pairs are usually completely uncalibrated - the camera intrinsics and extrinsics are unknown. Clinicians have to account for these unknown factors and hence their diagnoses of optic disc stability or change are subjective.**

Changes in the 3D structure of the optic disc are typically accompanied by changes in the 3D structure of blood vessels in that region. Therefore, change in the 3D structure of blood vessels can be used for glaucoma diagnosis. In this paper, we introduce a projective geometry based approach that reconstructs and aligns 3D blood vessel networks given two stereo pairs of optic disc images. We demonstrate that this alignment can identify regions where the vessel structure has changed. Since calibration is unavailable, the 3D structures and the alignment have a projective ambiguity; and hence, we cannot use an absolute threshold on the alignment error to automatically identify change. We have therefore developed an interactive tool that highlights regions with the largest alignment errors. This tool demonstrates the utility of our approach and also can guide clinical observers to optic disc regions where they should look for changes. We believe that our approach can serve as a platform to develop much needed novel tools for glaucoma diagnosis.

***Keywords*-Glaucoma diagnosis, reconstructing 3D blood vessel network, change detection via 3D alignment.**

I. Introduction

Glaucoma is the second leading cause of blindness worldwide. The visual impairment in glaucoma is caused by a distinctive form of degeneration of the optic nerve. Clinically, this degeneration is manifest by a characteristic 3D structural change in the optic disc – the portion of the optic nerve within the eye [1]. For many decades, the "gold standard" of optic disc evaluation in glaucoma has been qualitative evaluation by clinical observers of stereo images of the eye. Figure 1 (a,b) shows such a stereo pair. The circular region in the center is the optic disc. Due to the process in which these images are typically captured, the stereo pairs are completely uncalibrated - both camera intrinsics and extrinsics are unknown. Thus, clinicical observations are inherently subjective and quite variable. Most prior works on using imaging methods for glaucoma diagnosis have focused on defining certain structural characteristics of the optic disc. For instance, many computer vision approaches and specialized instruments aim to measure parameters of the optic disc, such as the cup-to-disc ratio[1]. However, comparing these different approaches can give variable results and/or inconclusive clinical diagnoses. It has long been appreciated that an accurate method to detect changes in the 3D structure of the optic disc would comprise a major advance in glaucoma diagnosis, but no such robust, validated methods are currently available.

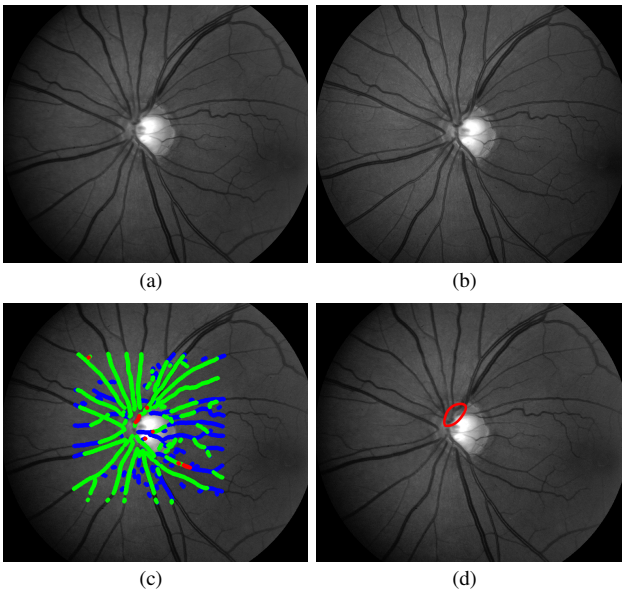


Figure 1. (a,b) Two images from an uncalibrated stereo pair. The circular region in the center is the optic disc; the more centrally located, lighter vertically oriented structure within the optic disc is a depressed region in 3D called the cup. (c) Color coded change detection result produced by our interactive tool. Vessel points that have changed are marked in red, points that are unchanged are marked in green, while points we are not confident about are marked in blue. (d) Change region marked by a clinical observer.

Changes in the 3D structure of the optic disc are typically accompanied by changes in the 3D structure of overlying blood vessels in that region. Therefore, for glaucoma diagnosis, change in the 3D structure of blood vessels can be used as a proxy for identifying structural changes in the optic disc.

[1]Cup-to-disc ratio is the ratio of the width of the central deeper portion to the entire diameter of the optic disc.

978-0-7695-4407-6/11 $26.00 © 2011 IEEE
DOI 10.1109/HISB.2011.1

We build on this idea and present a method that automatically reconstructs and aligns 3D blood vessel networks given two stereo image pairs. We then demonstrate that changes in 3D structure between the stereo pairs can be identified by comparing the aligned vessels. Thus, our approach can form the basis for improved glaucoma diagnosis.

To our knowledge, no previous work has compared 3D reconstructions over time to identify changes from stereo eye images. Our approach specifically exploits the 3D structure of blood vessels on the optic disc. It is motivated by the observation that, in glaucoma, changes in blood vessel structure are very local and reflect regional areas of optic nerve degeneration. Further, we also observe that the range of scene depths in these stereo images is small and that most of the retinal surface is near planar, except within the optic disc region. Therefore, we use a plane + parallax [2], [3] representation of the 3D structure for each stereo pair. We show that by establishing corresponding feature matches across all images in the two stereo pairs, we can estimate the 3D homography between their coordinate frames. This homography can then be used to align the 3D vessel networks computed from stereo. Since the homography is projective, we cannot use an absolute threshold on the alignment error to detect change automatically. We therefore developed an interactive tool which highlights regions with large alignment error and guides clinical observers to regions where they should seek change. In this tool, the observer has the option to set the change threshold and observe how regions marked as change vary with different thresholds. The tool demonstrates the potential diagnostic utility of our 3D alignment based approach.

To summarize we make the following contributions:

- We demonstrate how we can recover the 3D structure of vessel networks from uncalibrated stereo images. We recover the fine intricate geometry of the vessel networks and not just the rough retinal surface as in other previous works.
- We also show how we can align 3D vessel networks computed from two uncalibrated stereo pairs. This alignment can then be used to determine which vessel structures have changed. To our knowledge, ours is the first work that demonstrates such a capability.
- We have also developed an interactive tool that uses the estimated alignment to guide clinical observers to the vessels and hence optic disc regions that are most likely to have changed.
- A key component of our alignment process is finding matching points across all four images of two stereo pairs. The optics and capture process of the images yield unique challenges for feature matching – varying illumination, small depth of field, and noise. While we use standard techniques – SIFT and Harris features – we have devised a work flow that yields robust, plentiful and well distributed matches across all images.

II. Related Work

Several sophisticated systems are available for recovering the 3D geometry of the inside of the eye, such as optical coherence tomography (OCT) systems [4] and the Heidelberg Retina Tomograph (HRT) which is based on a confocal scanning laser ophthalmoscope [5]. These provide accurate measurements, but are expensive and require special capture sessions. In [6], an approach for detecting glaucomatous progression based on orthogonal decomposition is proposed. The approach uses absolute topographic measurements from the HRT and introduces several measures to compare measurements taken in follow-up exams against baseline measurements to detect progression and mark out change areas.

Computerized glaucoma diagnosis has also been approached using 2D image features only, typically, by looking for 2D optic disc and cup (the deepest structure inside the disc with detectable appearance differences) boundaries and then computing their relative size (radii) as a ratio. Xu et al. [7] proposed such an approach to automatically recover this ratio from single eye images. As stated above, most prior works on using imaging methods for glaucoma diagnosis focused on defining structural characteristics of the optic disc that are indicative of the disease and not on detecting 3D changes in the optic disc over time, as developed here.

Several works have applied stereo matching techniques to automatically recover the 3D geometry of the inside of the eye. Tang et al. [8] recently proposed a multi-scale stereo matching approach to reconstruct the shape of the optic nerve-head using stereo images. Their images are captured simultaneously and have a fixed stereo baseline. Thus, they do not need to resolve the epipolar geometry estimation issues that we need to address. Several other works [9], [10], [11] have also proposed stereo reconstruction approaches but not with the objective of comparing reconstructions over time. Medioni and his colleagues [12], [13], [14] have proposed approaches to handle the epipolar geometry estimation issues posed by the near-planar surface of the retina.

III. Plane + Parallax Representation

The range of scene depths in stereo images of the eye is small and the retinal surface is almost planar, barring the optic disc region. For such 3D structures the Plane + Parallax representation [2], [3] is a natural choice as it yields a reference coordinate system that is close to the structure. We use a plane to approximate the almost planar parts of the retinal surface and the rest of the 3D structure is represented as the distance from this plane. For completeness, we now give a quick overview of Plane + Parallax based 3D representation.

Consider two views $\mathcal{V}_l$ and $\mathcal{V}_r$ of the 3D structure. Let π be the reference scene plane which is used to define the 3D structure and let A be the 3×3 2D homography that maps points in $\mathcal{V}_l$ to $\mathcal{V}_r$ via π. For a scene point on π, A will exactly map its projection in $\mathcal{V}_l$ to its projection in $\mathcal{V}_r$.

(a) (b)

Figure 2. Two views of a 3D vessel network, computed from a stereo pair, in a Plane+Parallax representation.

However, for a scene point not on π, its projection in $\mathcal{V}_l$ mapped to $\mathcal{V}_r$ by A will not coincide with its projection in $\mathcal{V}_r$. The residual is termed as its disparity and represents how far the 3D point is from the reference plane. Let p_l and p_r be projections of a scene point P in views $\mathcal{V}_l$ and $\mathcal{V}_r$, respectively. It can be shown that [2], [3]

$$p_r \cong A p_l + \delta v_r, \tag{1}$$

where δ is the disparity with respect to the plane π, v_r is the unit vector in the direction $(p_r - e_r)$, e_r being the epipole in view $\mathcal{V}_r$, and $\cong$ denotes equal upto a scale. The disparity δ is infact invariant to the actual choice of view $\mathcal{V}_r$; it depends on the choice of the reference scene plane π. Therefore, we can define the 3D point P using its projection in $\mathcal{V}_l$ - $[p_{lx} \quad p_{ly} \quad 1]$ - and its disparity as

$$P = [p_{lx} \quad p_{ly} \quad 1 \quad \delta]. \tag{2}$$

We now describe how we compute 3D points on the vessel network given an uncalibrated stereo pair. First, we find matching features between the two images and estimate the Fundamental matrix between them (details are given in Sections VI-A and VI-B). The Fundamental matrix is used to compute transformations that rectify the images [15]. On the left rectified image (chosen as the reference), we detect points on the center of the vessels using a multiscale matched filter based method [16]. Figure 1(c) shows an example of the 2D vessel network extracted at this stage. It is for these pixels that we perform stereo matching (stereo matching details are in Section VI-D). The disparities produced by stereo matching can be interpreted as yielding corresponding points on vessels in the two rectified images. Undoing the rectifying transforms, we get correspondences in the original images. Now given the homography A that represents the mapping between the views via a reference scene plane π, we can use Equations 1 and 2 to map these correspondences to 3D points. Figure 2 shows two views of a 3D vessel network computed in this fashion.

IV. Feature Based Projective 3D Alignment

In the previous section, we have seen how we can recover a set of 3D points on vessels given a stereo pair. Given two stereo pairs, we can recover two sets of 3D points. However, we do not have one-to-one correspondence between them since vessel detection is performed independently in the left image of each stereo pair. In such a scenario, one might consider using techniques like Iterative Closest Point (ICP) to align the two point sets. However, since calibration is unavailable, the aligning transformation is more general than the rotation and translation transformations typically assumed for ICP.

We have developed an alternate approach that estimates the alignment transformation using corresponding points that are common across all four images of the two stereo pairs. The motivation behind using such correspondences is that when we combine the correspondences from the left and right images of a stereo pair to get 3D points, we get 3D point correspondences across the stereo pairs. These correspondences are not restricted to be on vessels and are distributed all over the image. That is, the alignment process does not use the vessel network directly.

Let us now see how we can align two sets of corresponding 3D points computed from two stereo pairs. Since the 3D points are defined in a projective frame, we know that they are related by a 3D homography [17]. More formally, if P_1 and P_2 are corresponding 3D points computed from stereo pairs 1 and 2 respectively (e.g. using Equation 2), then there is a 4×4 transformation H that satisfies

$$P_2 \cong H P_1, \tag{3}$$

The homography H is the same for all corresponding points and completely accounts for all the unknown camera intrinsics and extrinsics for the two stereo pairs[2].

We now describe the steps in our alignment procedure. For illustration, consider the two stereo pairs shown in Figure 3. These are separated by 3 years and the 3D structure has changed in that time. Let L_1 and R_1 denote the left and right images of stereo pair 1, and let L_2 and R_2 denote the left and right images of the stereo pair 2.

A. Finding 2D Image Correspondences

In the first step, we find correspondences across all images of the stereo pairs. We pick L_1 as the reference image and find Harris features in it. For these Harris features we try to find correspondences in each of the other images - R_1, L_2, and R_2 (details on feature matching are given in Section VI-A). We then retain only those correspondences that are common across all three pair-wise matching steps.

We separate these common 2D correspondences into two sets – ones outside the optic disc (bright region in the center), and ones inside the optic disc. Figure 3 shows the correspondences common across all four images where the red points are outside the optic disc while the green ones are close to or inside the optic disc[3]. This separation of correspondences into two sets is motivated by our earlier observation that the 3D structure of the retina is nearly planar outside the disc, while the inside of the disc shows a lot of variation in depth.

[2]It should be noted that H does not account for lens distortion. However, we have found that cameras typically used for such images come with high quality lenses with very low distortion.

[3]This separation is performed by using a mask of the optic disc region computed using the approach of [18].

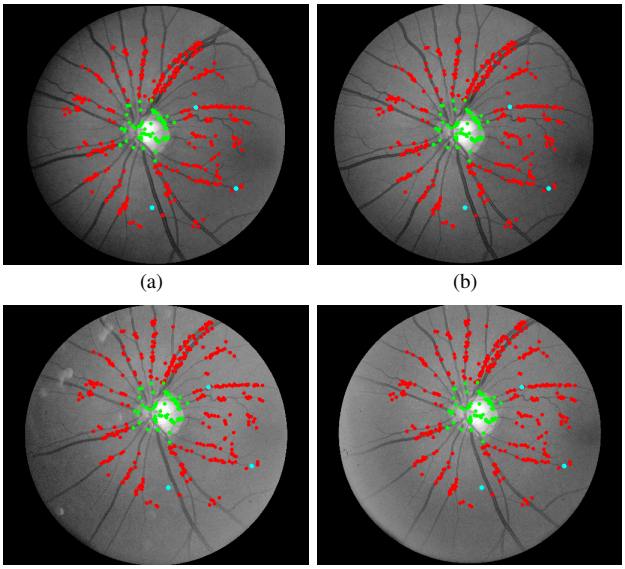


Figure 3. (a,b) Left and right images of stereo pair 1. (c,d) Left and right images of stereo pair 2. The 2D correspondences common to all four images are also shown. The correspondences outside the optic disc are marked in red, while the ones near or inside the optic disc are marked in green. The three correspondences that form the reference scene plane for the 3D homography that best aligns 3D structure computed from the two stereo pairs are marked in cyan.

Consequently, when picking correspondences to define the reference plane for the Plane+Parallax representation we would like to pick points from outside the disc, while to ensure that the homography estimation is well conditioned we would like to pick correspondences from inside the disc where there is greatest variation in 3D structure.

B. Computing 3D Points from 2D Correspondences

In order to determine the set of 3D points for stereo pair 1 from its 2D correspondences, we need the 2D homography associated with a reference scene plane (Section III). The 2D homography A used in Equation 1 has 8 degrees of freedom and we need 4 2D correspondences to define it. From the definition of Plane+Parallax the homography A should map the epipoles in the left and right views [2], [3]. So we can use the epipoles and three 2D correspondences between L_1 and R_1 to get the 8 constraints needed to define A. Once A is determined, we can use Equation 2 to get the 3D point corresponding to each 2D correspondence. We follow the same procedure to generate the 3D point set for stereo pair 2. For simplicity, we pick the same corresponding features to define the same reference plane in both stereo pairs.

C. Computing 3D Homography from 3D Correspondences

A 3D homography (Equation 3) has 15 degrees of freedom and we need atleast 5 corresponding 3D points to derive it. Note that to define the reference scene plane's homographies (A's), we picked three 2D correspondences. The 3D points derived from these 2D correspondences give us three of the required correspondences[4]. Now we need only two more correspondences. To ensure that the homography estimation is well conditioned we pick 3D points derived from two 2D correspondences from inside the optic disc region (marked in green in Figure 3). These 5 correspondences can be used to define the 3D homography H in Equation 3.

Till now we have assumed that the 2D correspondences across all four images of the stereo pair are correct. However, inevitably, the set of correspondences has outliers, which should be ignored. In order to robustly estimate the 3D homography, we have developed the following RANSAC-based procedure:

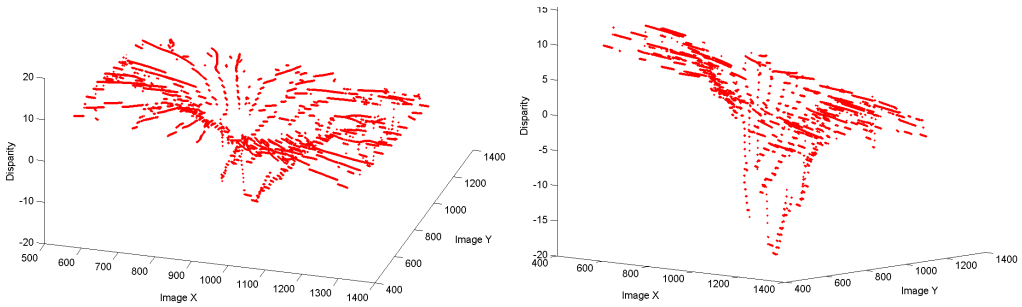
1) **Choose reference plane π_i:** We randomly pick three 2D correspondences from outside the optic disc to define 2D homographies corresponding to the reference scene plane π_i. These 2D homographies are used to construct the 3D point sets for the two stereo pairs.
2) **Ransac fit of H given π_i:** We now use a 2-point RANSAC to find the best 3D homography estimate with the reference scene plane π_i. We randomly pick two corresponding 3D points from inside the optic disc region to get a 3D homography $\hat{H}$ as described above. For all other 3D points, we then compute the fitting error with respect to $\hat{H}$. If P_1^j and P_2^j are corresponding 3D points from stereo pairs 1 and 2, respectively the symmetric fitting error is defined as

$$E^j = ||P_1^j - \alpha(\hat{H}P_2^j)||_2 + ||P_2^j - \alpha(\hat{H}^{-1}P_1^j)||_2, \quad (4)$$

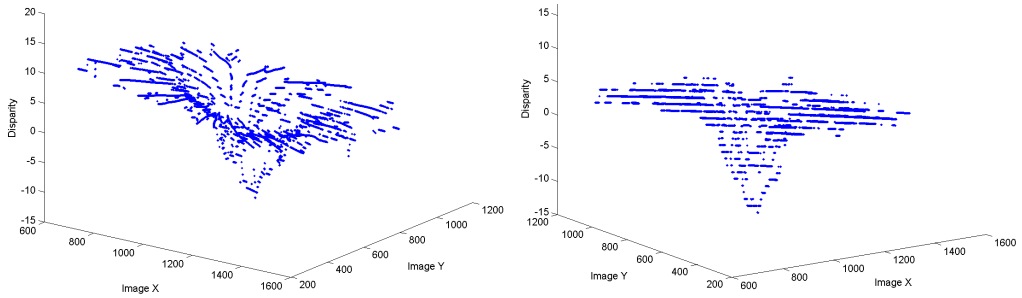
where $||.||_2$ denotes the L_2-norm and $\alpha(.)$ is a function that scale normalizes its vector argument so that its third component is 1. If the fitting error is less than a threshold $\mathcal{T}$, then we consider the correspondence to be an inlier, else it is considered an outlier. This step is repeated with different randomly chosen pairs of correspondences inside the optic disc. Of all the RANSAC trials, we pick the one with the maximum number of inliers - N_i. All the inlier correspondences in this trial are then used to get a least squares estimate H_i as per Equation 3.

In Step 1, we pick a reference plane π_i to define the set of 3D points and in Step 2, we find the best 3D homography that aligns 3D points in the coordinate frame defined by π_i. We repeat the above process T times and then pick the homography H_i with the maximum number of inliers N_i. This is the best homography H that aligns 3D structures computed using the associated reference scene plane π_i. We typically have around 200-400 correspondences (around 80% of them outside the optic disc) across all four images of the two stereo pairs and empirically we found that $T = 5000$ times gives good results. The triplet of correspondences chosen in Step 1 that yielded the best 3D homography for

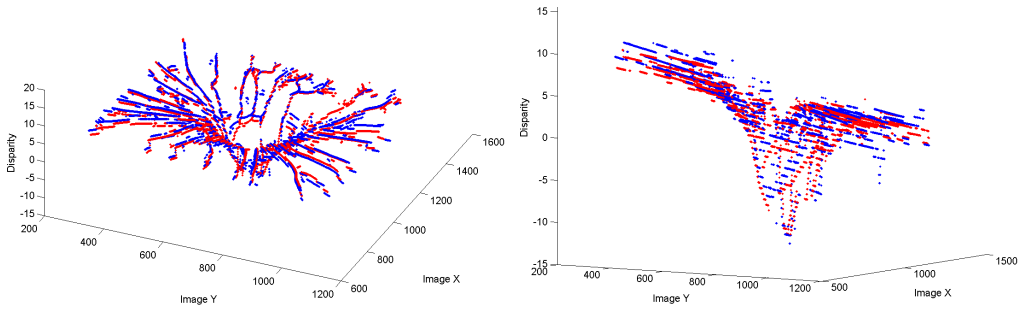
[4]Since these 3D points lie on the reference scene plane, they have zero disparity; $\delta = 0$ in Equation 1.



(a) Two views of the 3D vessel network computed from the top stereo pair shown in Figure 3.



(b) Two views of the 3D vessel network computed from the bottom stereo pair shown in Figure 3.



(c) Two views of the 3D vessel networks in (a) and (b), aligned in the same coordinate frame.

Figure 4. Computing 3D vessel networks from two stereo pairs (shown in Figure 3) and aligning them.

the stereo pairs in Figure 3 are marked in cyan.

Note that in principle we could have used a 5-point RANSAC process to estimate the best 3D homography. Our approach, as we mentioned earlier, is tailored to the particular geometry seen in these images. If we have 3D points on the vessels computed from the two stereo pairs (as described in Section III and in more detail in Section VI-D), we can use the estimated 3D homography to align the 3D points. Figures 4 (a) and (b) show views of the 3D vessel networks computed from respectively the top and bottom stereo pairs in Figure 3. Figure 4(c) shows views of both vessel networks in the same coordinate frame, where the alignment was computed as described above. The error in this 3D alignment would be indicative of change, which can be used in the diagnosis of glaucoma. We discuss this next.

V. Change Detection

Glaucoma progression is characterized by small local changes in the 3D structure of the optic disc. We assume that a 3D homography computed using image features located over the entire image would globally align 'stable' or unchanged 3D vessel points from two stereo pairs, but fail to align small local changes. Thus, by studying the alignment error we can detect vessels with changed structure.

A. Computing the Alignment Error

We now show how to compute the alignment error for 3D vessel points from two stereo pairs, using the 3D homography H estimated as described in Section IV-C. Let P_1^j and P_2^k represent 3D vessel points computed from stereo pairs 1 and 2, respectively. We apply H to P_1^j to get the corresponding point in the coordinate system of P_2:

$$P_{1\to 2}^j = H P_1^j. \tag{5}$$

$P_{1\to 2}^j$ is normalized so that its third component is 1. As noted earlier, the vessel points in the two stereo pairs need not correspond. So for each $P_{1\to 2}^j$ we need to find the 'closest' point P_2^l. In our implementation, P_2^l is the point that is closest to $P_{1\to 2}^j$ in the $x-y$ image plane

(first two dimensions of the vectors). We then compute the alignment error as the difference in the disparity values (fourth component) of $P^j_{1\to 2}$ and P^l_2. This is the alignment error in the coordinate frame of stereo pair 2. Using H^{-1} we can compute the alignment error in the coordinate frame of stereo pair 1 in an analogous manner.

B. Interactive Change Detection

Since calibration information is unavailable, the 3D reconstructions and alignment have a projective ambiguity, and so the alignment errors are relative to each individual coordinate frame. Thus, one cannot use an absolute threshold on the alignment errors to identify changed structure. Consequently, to demonstrate the diagnostic utility of our approach, we have developed an interactive tool that allows a clinical observer to vary the change threshold on the alignment error and determine the regions which are most likely to have changed. Thus, the tool serves as a guide in identifying changed regions.

Figure 5 shows results from our interactive tool for three different eyes. Each row has images for a particular eye and the first three columns have the change detection results for three progressively decreasing change thresholds. The results are shown by assigning vessel points one of three colors – (a) green, which signifies vessel points that have not changed, (b) red, which signifies vessel points that have changed, and (c) blue, which denotes vessel points about which we are not confident in making a judgment. The last category includes points on vessels that are aligned with the epipolar direction. These vessels appear horizontal in the rectified stereo images and hence we cannot get good correspondences for them. The vessel detection algorithm of [16], which we apply on the rectified images, identifies points along the center of vessels and also provides the local vessel orientation at each point. We use this orientation information to identify such low confidence points. For the above results, points on vessels with orientation $\pm 5^\circ$ from the horizontal are marked in blue. Vessel points that are detected in only one image are also marked in blue.

In the figures, we have marked out vessel points only near the optic disc. This is because in diagnosing glaucoma, clinicians only analyze vessel structure in the vicinity of the optic disc. The actual change regions marked by clinical observers are shown in the last column of Figure 5. One can see that as the change threshold is decreased the vessel region corresponding to the actual change is marked more clearly in red. For these three results we used stereo pairs separated by 3, 7, and 3 years, respectively.

VI. Other Components of the Alignment Process

In the previous sections, we have shown how to reconstruct and align 3D vessel networks to detect change. This process uses several components that use standard techniques for tasks like feature matching and epipolar geometry estimation. However, these had to be adapted to the unique challenges presented by eye images. For completeness, we describe some of these components in this section.

A. Feature Matching

There are a number of challenges to precisely detecting features across different subjects due to a large variation in their appearance and visible anatomical structures. Similarly, finding corresponding features between different images of the same eye is made difficult by their differing illumination, depth-of-field, focus and the actual retinal area photographed. Recent works like [13], [12] have proposed using different features like vessel bifurcations "Y-features" or SIFT features for the alignment task. However, considering the wide variety of images we have come across, we have found neither the density or the localization of these features as adequate for the precise 3D alignment problem we focus on in this paper. We follow a two-stage process where we first use large-scale SIFT features to bring the two images into a coarse alignment and then use Harris corners for a more precise estimation of the alignment parameters. Our matching approach is most similar to the hierarchical approach in [19] but we use different features in our two stage process and we employ a search strategy in the second stage as described next.

Let I_l and I_r be two images in which we need to find corresponding features and let I_l be the reference image.

SIFT-based Coarse Alignment: We first detect coarse-scale SIFT features in each image and then match those features using the 1-NN/2-NN ratio scoring measure [20] to get corresponding points. These correspondences are used in a RANSAC-based 2D homography estimation step to compute a homography that warps I_r towards I_l. Let us denote the warped image I'_r.

Harris Feature Matching: Given the continuous vessel structure in eye images, it is difficult to directly detect corresponding Harris features in both images. Hence, methods that try to establish correspondences between pre-detected features in both images, like [21], are inapplicable. Therefore, we employ a search strategy where the Harris features are detected only in the reference image I_l. For a Harris feature at (x, y), we take a $w \times w$ patch around it and search for the best matching patch in a local $q \times q$ 2D neighborhood around (x, y) in the warped image I'_r. This is possible because after the SIFT-based first stage, the images I_l and I'_r are 'close' to each other. For matching patches we use normalized cross-correlation. For each match we also compute a confidence measure. If s_1 and s_2 are the highest and second highest correlation scores in the $q \times q$ neighborhood, we define the confidence c as:

$$c = \frac{|s_1 - s_2|}{s_2}. \tag{6}$$

We discard matches with confidences below a threshold c_t.

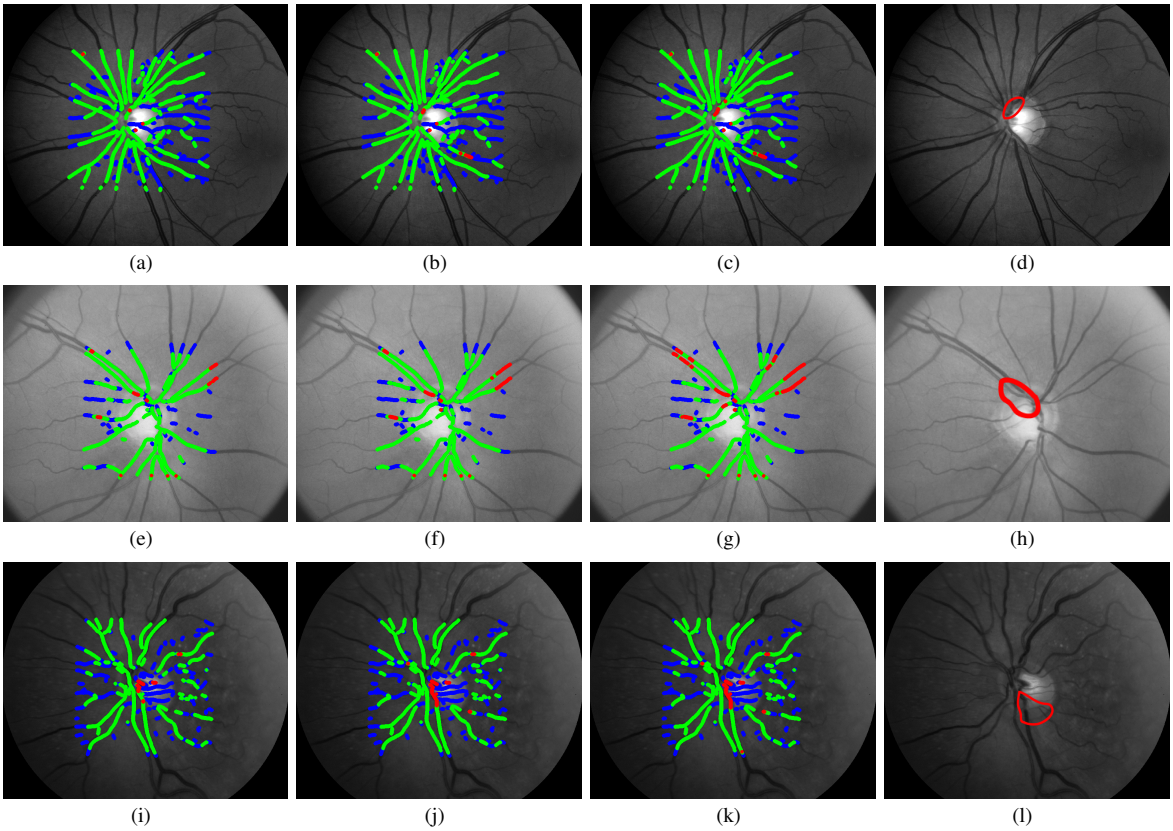


Figure 5. Interactive Change Detection. Each row corresponds to a particular eye. (columns 1-3) Change detection results for interactively chosen change thresholds. The thresholds decrease from column 1 to column 3. See text for details on the color coding of the change detection result. (column 4) Change regions marked by a clinical observer.

We have found that this two-stage process yields robust, plentiful and well distributed matches as can be seen in Figure 3. Our images are 2000×1712 pixels and for feature detection, we have used $w = 51$, $q = 61$, and $c_t = 0.025$.

B. Epipolar Geometry Estimation

The near-planarity of the retinal surface makes it challenging to accurately estimate 3D characteristics including the recovery of the Fundamental matrix between an image pair. Choe et al. [13] suggested a virtual parallax approach for robust recovery of the epipolar geometry for a near-planar retinal surface. In their approach, a 2D homography H_π (which maps points between the images via a plane π) is first estimated between the two images using a RANSAC-based feature matching procedure. Thereafter, another RANSAC procedure determines two additional correspondences such that their parallax vectors, with respect to the plane π, intersect at an epipole. In practice, we found this procedure to be inaccurate due to the small range of depths on the retinal surface. In many cases, the parallax vectors are too small in their magnitude to compute an intersection robustly.

As we pointed out in Section IV-A, points inside the optic disc exhibit a large variation in depth. Therefore, we restrict the search for correspondences in the second RANSAC step to be inside the optic disc. To aid this process, the optic disc is automatically segmented using the simple technique suggested in [18]. Additional feature correspondences are extracted in this region using the Harris-based technique described in Section VI-A. These correspondences are used to robustly estimate the epipole e. The fundamental matrix is then derived as $F = H_\pi [e]_\times$.

C. 2D Vessel Detection

For detecting image points on blood vessels, we have used the multi-scale match filter approach of [16], which in turn builds on the work in [19]. The algorithm in [16] uses a collection of matched filters tuned to a variety of vessel widths and orientations to estimate a filter response at each pixel in the input image. Then, it uses a ridge detector on the response map to determine pixels along the center of the vessels. These individual center pixels are then chained together in a greedy 'longest-vessel-first' strategy to get vessel segments. The algorithm also produces estimates of the local vessel width and orientation at each point. We use the orientation information in Section V to determine vessels that are close to horizontal in the rectified stereo images.

D. Stereo matching for Vessel Points

Given a pair of stereo images and the Fundamental Matrix that relates them, we can compute rectifying transformations

that warp the images such that the epipolar lines are horizontal [15]. We then use simple block based matching to find correspondences for all 2D vessel points extracted in the reference image using the process described in Section VI-C. For our results we have used 51×51 windows with zero-mean normalized cross correlation as the match measure. For robustness to rectification errors, for each disparity shift, we also compute the correlation obtained by offsetting the matching window by ± 1 scan line in the right image.

VII. Discussion

In this paper, we have demonstrated how to recover 3D vessel networks from uncalibrated stereo pairs. We have also shown how we can align 3D vessel networks computed from two stereo pairs with the aim of identifying changed vessel structures over time, using the alignment error. To our knowledge, this is (a) the first work that addresses recovering the fine intricate geometry of blood vessel networks from uncalibrated stereo pairs, as opposed to the rough retinal surface recoveries in previous works, and (b) the first work that determines change in 3D structure by automatically aligning 3D vessel networks from two uncalibrated stereo pairs. Since calibration information is unavailable, the reconstructions and the alignment have a projective ambiguity. Thus, we cannot use the alignment errors directly to identify change. Consequently, we developed an interactive tool that allows observers to vary the change threshold on the alignment error, thus identifying the regions which are most likely to have changed. This interactive tool demonstrates the potential utility of 3D vessel reconstructions for change detection. We believe that our approach would permit the future development of fully automated change detection algorithms, for instance by studying the distribution of relative alignment errors. Such automatic tools could potentially enable more accurate glaucoma diagnosis at a much lower cost, compared to the present subjective evaluations of optic disc images or expensive specialized imaging instruments.

Acknowledgments

The authors gratefully acknowledge support from NEI grant EY-017299, the Mackall Foundation Trust, and Research to Prevent Blindness.

References

[1] P. Airaksinen, A. Tuulonen, and E. Werner, "Clinical evaluation of the optic disc and retinal nerve fiber layer," *The Glaucomas. Editors: R. Ritch, M. Shields, T. Krupin*, pp. 617–657, 1996.

[2] H. S. Sawhney, "3D geometry from planar parallax," *In Proc. of CVPR*, pp. 929 – 934, 1994.

[3] B. Triggs, "Plane + Parallax, Tensors and Factorization," *In Proc. of ECCV*, pp. 617–657, 2000.

[4] A. Podoleanu, J. Rogers, and D. Jackson, "Three dimensional oct images from retina and skin," *Optic Express*, 2000.

[5] P. Sharp and A. Manivannan, "The scanning laser ophthalmoscope," *Phys. Med. Biol.*, pp. 951–966, 1997.

[6] M. Balasubramanian, S. Zabic, C. Bowd, H. Thompson, P. Wolenski, S. Iyengar, B. Karki, and L. Zangwill, "A framework for detecting glaucomatous progression in the optic nerve head of an eye using proper orthogonal decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 781–793, 2009.

[7] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. Chew Tec Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognition*, vol. 40, no. 7, pp. 2063–2076, 2007.

[8] L. Tang, Y. Kwon, W. Alward, E. Greenlee, K. Lee, M. Garvin, and M. Abràmoff, "3D reconstruction of the optic nerve head using stereo fundus images for computer-aided diagnosis of glaucoma," in *SPIE*, vol. 7624, 2010.

[9] 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," *Engineering in Medicine and Biology Society*, pp. 6747–6750, 2007.

[10] T. Nakagawa, Y. Hayashi, Y. Hatanaka, A. Aoyama, T. Hara, M. Kakogawa, H. Fujita, and T. Yamamoto, "Comparison of the depth of an optic nerve head obtained using stereo retinal images and HRT," *SPIE*, vol. 6511, 2007.

[11] E. Corona, S. Mitra, M. Wilson, P. Soliz, T. Krile, and Y. Kwon, "Digital Stereo Image Analyzer for Generating Automated 3-D Measures of Optic Disc Deformation in Glaucoma," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 1244–1253, 2002.

[12] T. E. Choe, I. Cohen, and G. Medioni, "3-d shape reconstruction of retinal fundus," *CVPR*, pp. 2277–2284, 2006.

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

[14] Y. Lin and G. Medioni, "Retinal image registration from 2d to 3d," *CVPR*, pp. 1–8, 2008.

[15] J. Mallon and P. F. Whelan, "Projective Rectification from the Fundamental Matrix," *Image and Vision Computing*, 2005.

[16] M. Bansal, S. Kuthirummal, J. Eledath, H. Sawhney, and R. Stone, "Automatic Blood Vessel Localization in Small Field of View Eye Images," *IEEE Engineering in Medicine and Biology Society*, 2010.

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

[18] 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*, pp. 2266–2269, 2008.

[19] M. Sofka and C. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1531–1546, 2006.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, pp. 91–110, 2004.

[21] L. Torresani, V. Kolmogorov, and C. Rother, "Feature Correspondence via Graph Matching: Models and Global Optimization," *In Proc. of ECCV*, pp. 596–609, 2008.