MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Analysis of Shape Assumptions in 3D Reconstruction of Retina from Multiple Fundus Images

Ataer-Cansizoglu, E.; Taguchi, Y.; Kalpathy-Cramer, J.; Chiang, M.F.; Erdogmus, D.

TR2015-051 April 2015

Abstract

Utilizing priors about the shape of retinal surface is important for accurate reconstruction. We present a detailed analysis of geometrical shape priors in the 3D reconstruction of retina. We first approximate the retinal surface either as a sphere inspired by the actual shape of the eyeball, or as a plane inspired by the 2D mosaicing approaches. Based on this approximation, we perform an initial camera localization with a 2D-to- 3D registration procedure. Then, parameters of the surface and the camera poses are refined through a nonlinear least squares optimization using different shape priors. The resulting 3D model and camera poses can be used for intuitively visualizing the retinal images with a model-guided browsing interface.

2015 IEEE International Symposium on Biomedical Imaging (ISBI)

ANALYSIS OF SHAPE ASSUMPTIONS IN 3D RECONSTRUCTION OF RETINA FROM MULTIPLE FUNDUS IMAGES

E. Ataer-Cansizoglu[1], *Y. Taguchi*[2], *J. Kalpathy-Cramer*[3], *M. F. Chiang*[4], *D. Erdogmus*[1]

[1]Cognitive Systems Laboratory, Northeastern University, Boston, MA, USA
[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA
[3]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
[4]Departments of Ophthalmology & Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA

ABSTRACT

Utilizing priors about the shape of retinal surface is important for accurate reconstruction. We present a detailed analysis of geometrical shape priors in the 3D reconstruction of retina. We first approximate the retinal surface either as a sphere inspired by the actual shape of the eyeball, or as a plane inspired by the 2D mosaicing approaches. Based on this approximation, we perform an initial camera localization with a 2D-to-3D registration procedure. Then, parameters of the surface and the camera poses are refined through a nonlinear least squares optimization using different shape priors. The resulting 3D model and camera poses can be used for intuitively visualizing the retinal images with a model-guided browsing interface.

Index Terms— retinal images, 3D reconstruction, visualization

1. INTRODUCTION

The need for computerized analysis of retinal images has been increasing with the wide clinical use of fundus photography. Diagnosis process involves the acquisition of multiple images taken from different viewpoints. These images are investigated one by one by doctors for final diagnostic decision. Furthermore, comparing two sets of images from different visits of the same patient might be vital to see the effect of treatment and/or to investigate the spread of the disease.

Existing studies about retinal image registration mainly deal with the issue of 2D registration to align a sequence of images to a reference frame. In [1, 2], this registration is followed by image mosaicing. Although mosaicing provides an enlarged view of retinal fundus, it involves a 2D approximation of the 3D transformation between images, which is less realistic. Moreover, a 3D retinal surface visualization would be easy and handy with the current advances in image visualization.

Although retinal image processing is a well studied field, there are a few studies about 3D reconstruction of retinal fundus [3, 4, 5, 6, 7, 8]. Liu et al. [3] estimate the epipolar geometry and projection matrices after a self calibration. Choe et al. [4] apply a plane+parallax algorithm to register images, which is followed by a mutual information-based disparity search stage. Both [3] and [4] work on stereo pairs, which have large overlapping areas. The method by Martinez-Perez et al. [5] involves an offline calibration stage and provides 3D reconstruction of the retinal vessels. Chanwimaluang et al. [6] present a retinal surface reconstruction technique using an affine camera model, since they work on retinal images with small field of view. Lin et al. [7] perform a 2D-to-2D registration followed by a multi-stage bundle adjustment.

Retinal surface reconstruction differs from traditional stereo reconstruction since fundus is observed through eye lens which produces distortion in the images. Aforementioned studies lack the consideration of this distortion effect. Deguchi et al. [8] assume a spherical shape for the eyeball. They model the mapping of this surface through the eye lens as a quadratic surface where camera localization is performed through registration of correspondences on this quadratic surface. Their method is based on two-stage optimization of the reprojection error. First they minimize the error with respect to the camera poses by keeping the surface equation fixed. This yields a good initialization for the camera poses to be used at the next step. Second, they minimize the error with respect to the surface equation and camera poses.

Utilizing priors about the shape of retinal surface is useful for accurate reconstruction. In this paper, we present a thorough analysis of 3D retinal surface reconstruction algorithms using different shape priors. Our goal is to reconstruct the real image of the retina surface transformed through the eye lens, which is modeled as a quadratic surface [8]. Since the ex-

EAC, JKC, MFC and DE are supported by grants R01 EY19474, R21 EY22387 from NIH and unrestricted departmental funding from Research to Prevent Blindness.

act shape of this quadratic surface is originally unknown, our method first approximates this surface either (i) as a sphere inspired by the actual shape of the eyeball, or (ii) as a plane inspired by the 2D mosaicing approaches. Based on this approximation, we construct a 3D model from the initial image and grow our model by localizing each image with a 2D-to-3D registration process. Second, the parameters of the quadratic surface equation and the camera poses are refined with a bundle adjustment (BA) procedure where the estimations from the first step are used as initialization. In order to assess the accuracy of the quadratic surface assumption, we also perform the second stage with other shape assumptions and compare the results. Finally, we present a model-guided browsing interface that helps navigate through multiple retinal images by using the reconstructed 3D model as the guide. Thus, the contributions of this paper are two fold: an in-depth analysis of shape priors in 3D retinal surface reconstruction and a novel model-guided interface that provides navigation of retinal images.

2. OUR METHOD

We define the world coordinate system as the coordinate system of the real image of retina transformed through the eye lens. Our goal is to reconstruct the retinal surface in the world coordinate system, which is modeled as a quadratic surface [8]. Our method consists of two steps. Starting from an initial shape assumption (sphere or plane) we first perform camera localization. We then refine the camera poses and the parameters of the quadratic surface using a nonlinear least squares optimization. In order to provide a comparative analysis of shape priors, the refinement stage is also carried out with planar and spherical shape assumptions.

2.1. Initial 3D Model Creation

Let $\mathbf{p}_k^w = (x_k, y_k, z_k)^T$ denote the kth 3D point in the world coordinate system. Assuming a spherical or planar shape of the retinal surface, we have the following constraints on the 3D points, respectively:

$$x_k^2 + y_k^2 + z_k^2 = r^2 \tag{1}$$

$$\beta_1 x_k + \beta_2 y_k + \beta_3 z_k + 1 = 0 \tag{2}$$

Here r is the radius of the sphere centered at origin, and $\beta_1, \beta_2, \beta_3$ are the plane parameters.

Let us denote the pose of the ith frame as $\mathbf{T}_i \in \mathbb{SE}(3)$, where $\mathbb{SE}(3)$ is the Euclidean group representing 3D rigid body motions. $\mathbf{T}_i$ consists of a rotation matrix $\mathbf{R}_i$ and a translation vector $\mathbf{t}_i$. Then a 3D point $\mathbf{p}_k^w$ in the world coordinate system is transformed to the camera coordinate system of the ith frame as

$$\mathbf{p}_k^i = \mathbf{R}_i \mathbf{p}_k^w + \mathbf{t}_i. \tag{3}$$

The 3D point $\mathbf{p}_k^i = (x_k^i, y_k^i, z_k^i)^T$ is projected to a pixel $\mathbf{q}_k^i = (u_k^i, v_k^i)^T$ on the image plane according to the pinhole camera model as

$$u_k^i = f x_k^i / z_k^i + c_x \tag{4}$$

$$v_k^i = f y_k^i / z_k^i + c_y, \tag{5}$$

where f is the focal length and (c_x, c_y) is the principal point. We denote this function that performs the transformation followed by the projection as $\mathbf{g}^{\mathbf{T}_i} : \mathbb{R}^3 \to \mathbb{R}^2$. On the other hand, given the pose of the camera and the parameters of the surface shape, we can backproject a pixel $\mathbf{q}_k^i$ to a 3D point $\mathbf{p}_k^w$ in the world coordinate system by computing the intersection between the surface and the ray corresponding to the pixel. Let us denote this backprojection function as $\mathbf{h}^{\mathbf{T}_i} : \mathbb{R}^2 \to \mathbb{R}^3$.

Initial model: We denote the first frame as the reference frame. For spherical surface, we assume its center is fixed at the origin and the optical axis of the reference frame passes through the sphere center, i.e., $\mathbf{R}_1 = \mathbf{I}_3, \mathbf{t}_1 = (0, 0, r - r')^T$, where $\mathbf{I}_3$ is the 3×3 identity matrix and r' is the distance to the surface. Similarly, the planar model is assumed to be orthogonal to the normal of the reference frame (i.e., $(\beta_1, \beta_2, \beta_3) = (0, 0, 1)$) and the distance between the reference frame and the plane is the same as the distance between the reference frame and the spherical model, i.e., $\mathbf{R}_1 = \mathbf{I}_3, \mathbf{t}_1 = (0, 0, -1 - r')^T$.

In this work, we use the bifurcation points of the vessels as feature points, and the correspondences between the feature points across images are manually selected. The feature points in the reference image are backprojected onto the retinal surface using the function $\mathbf{h}^{\mathbf{T}_1}$. Next, we proceed by registering each 2D retinal image with this 3D model. We exploit the standard P3P algorithm [9], which is widely used in structure from motion and simultaneous localization and mapping (SLAM) systems. We geometrically verify the given point correspondences in a RANSAC framework. RANSAC helps us handle the noise in picking pixel locations and noise in localization due to the geometrical shape assumption. After localizing each image, the inlier points and the new feature points that do not appear in the model, but appear in the localized image, are added to the 3D model. Thus, the 3D model is enlarged for each new image.

2.2. Bundle Adjustment

The spherical shape of the fundus is seen as a quadratic surface through the eye lens due to refraction [8]. Hence the following equation is satisfied for $\mathbf{p}_k^w$

$$\alpha_1(x_k^2 + y_k^2) + \alpha_2 z_k^2 + \alpha_3 x_k z_k + \alpha_4 y_k z_k + \alpha_5 z_k + 1 = 0, \tag{6}$$

where $\alpha_1, \ldots, \alpha_5$ are the surface parameters.

Assume the kth 3D point is observed in the ith and jth frames as $\mathbf{q}_k^i$ and $\mathbf{q}_k^j$, corresponding to each other. The reprojection error of this pair is given by

$$e(\mathbf{q}_k^i, \mathbf{q}_k^j) = \|\mathbf{g}^{\mathbf{T}_j}(\mathbf{h}^{\mathbf{T}_i}(\mathbf{q}_k^i)) - \mathbf{q}_k^j\|. \tag{7}$$

We carry out an optimization procedure [10] to minimize the sum of reprojection errors of all correspondence pairs from all images.

$$\underset{\mathbf{T}_1,\ldots,\mathbf{T}_M,\alpha_1,\ldots,\alpha_5}{\operatorname{argmin}} \sum_{i,j,k} e^2(\mathbf{q}_k^i, \mathbf{q}_k^j) + e^2(\mathbf{q}_k^j, \mathbf{q}_k^i). \quad (8)$$

The initialization for this procedure is taken from the estimates we get in the first step. Note that we optimize the poses of all cameras since the surface is defined in the world coordinate system independent of the camera poses.

2.3. 3D Visualization

After the bundle adjustment, we have the surface parameters and the refined pose of each image. We create a point cloud by backprojecting all the pixels in all the images onto the surface based on the refined poses, resulting in a highly dense point cloud. For a better visualization, we downsample the point cloud based on a voxel grid. We set a voxel grid with a size of (v_x, v_y, v_z) and each grid is represented with the centroid of the 3D points that falls inside this grid.

3. EXPERIMENTS AND RESULTS

We implemented the system in C++ on a Windows 8 environment. The experiments were carried out on sequences of retinal images acquired from healthy newborn infants. We used $r = 8.5$ mm as the radius of the sphere, since average eyeball diameter was reported to be 16–17 mm at birth [11]. Camera intrinsic parameters were provided by the imaging system[1]. We carried out our experiments on two image sequences, each of which consisted of 4 wide-angle retinal images. For each image, we had around 20 correspondences. Using a reprojection error threshold of 8 pixels the P3P algorithm returned about 50% inlier ratio for spherical model and 90% inlier ratio for planar model.

In order to compare different shape priors, initial camera localization was carried out using either spherical or planar surface model, and consequent refinement was carried out using a spherical, quadratic, or planar surface. Resulting reprojection errors can be seen in Table 1. Except the spherical model, reprojection errors were small. In both sequences the minimum reprojection error was acquired when we started with a spherical model and performed BA with the quadratic surface assumption.

Reconstruction results with different shape priors are displayed in Figure 1 for the sequence 2. As seen from the results, vessels were mostly matched in quadratic and planar surface reconstructions, while spherical model had some errors. Also, the resulting shape of the quadratic surface was close to planar. By looking at the small reconstruction errors in Table 1 and shape of the resulting quadratic surface, we observed that 2D mosaicing of these images might be a good approximation of the actual geometry.

Initial Model	Bundle Adjustment	Seq. 1	Seq. 2
Sphere	Sphere	237.24	323.60
Sphere	Quadratic	5.06	5.12
Plane	Quadratic	9.02	11.91
Plane	Plane	6.64	7.82

Table 1. Average reprojection errors in pixel square units.

Figure 2 shows an overview of our visualization system and the reconstruction result for the sequence 1 with an initial spherical model followed by BA on a quadratic surface. Here, the user can view the camera positions and navigate to the corresponding image by clicking the camera icon. (Please see `http://youtu.be/FMwdTLQX1fo` for a supplementary video.) Notice the matching vessels. As a vessel goes out of one view, it can be seen continually in a neighboring view.

[1]We did not use a distortion model as it was not provided by the imaging setup.

4. CONCLUSION AND DISCUSSION

We presented an in-depth analysis of shape assumptions in 3D retinal fundus reconstruction. First, cameras were localized based on an initial 3D model assumption (sphere or plane). Second, the parameters of the surface and the estimated poses of the images were simultaneously refined using a nonlinear least squares minimization of the reprojection errors. In this step, we assumed different shape models including planar, quadratic, and spherical surface. The experiments were carried out on two sequences of retinal images. As a result 3D reconstruction with a quadratic or planar surface model gave small reprojection errors. Moreover, small reprojection errors with planar assumption and the shape of resulting quadratic surface being close to plane suggested that 2D mosaicing of the images can provide a good approximation of the actual geometry. We also presented a novel visualization framework that can be used as a model-guided browsing interface for retinal images. In the future, we will extend our system by taking the center lines of vessels into account during registration. Improved filtering techniques for better 3D visualization are also in the scope of our future work.

5. REFERENCES

[1] S Lee, MD Abràmoff, and JM Reinhardt, "Retinal image mosaicing using the radial distortion correction model," in *Medical Imaging*, 2008, vol. 6914.

[2] PC Cattin, H Bay, L Van Gool, and G Székely, "Retina mosaicing using local features," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9, pp. 185–192, Jan. 2006.

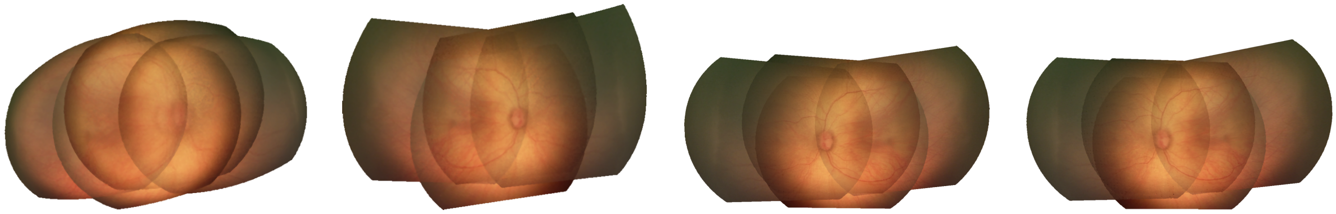

Fig. 1. Reconstruction results.

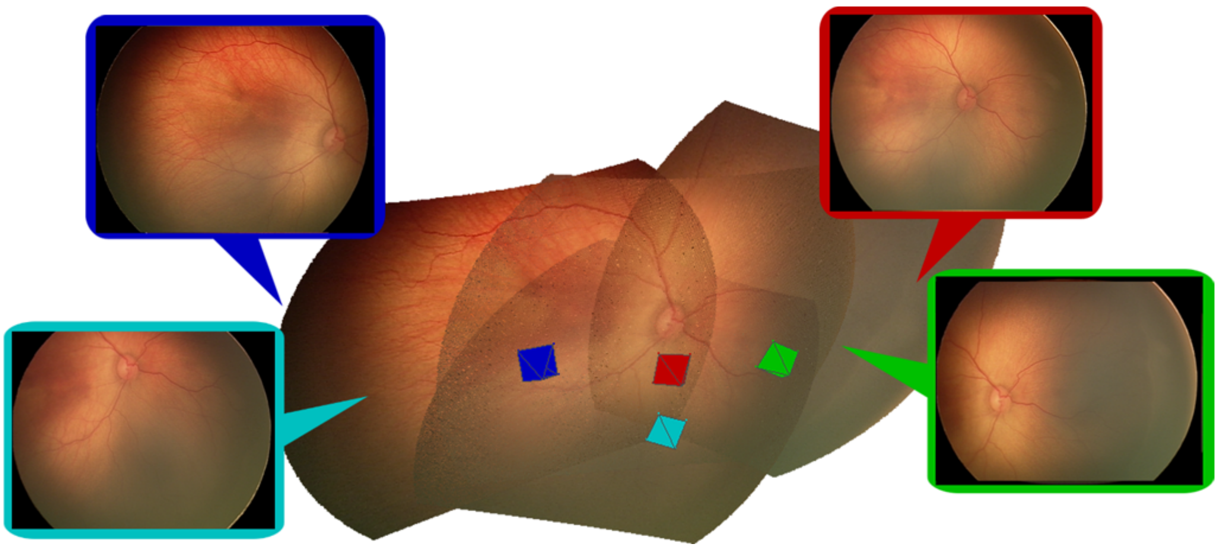

Fig. 2. Overview of our visualization system and the reconstruction result for the first sequence of images. The reconstructed model of the retinal fundus is visualized along with the estimated poses of cameras denoted by colored camera icons. The system provides a model-guided image browsing interface, where the user can navigate to each retinal image by clicking the corresponding camera icon (please see the supplementary video).

[3] D Liu, NB Wood, XY Xu, N Witt, AD Hughes, and Thom SAMcG, “3D reconstruction of the retinal arterial tree using subject-specific fundus images,” *Advances in Computational Vision and Medical Image Processing: Methods and Applications*, vol. 13, pp. 187, 2008.

[4] TE Choe, I Cohen, G Medioni, AC Walsh, and SR Sadda, “Evaluation of 3-D shape reconstruction of retinal fundus,” *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9, pp. 134–141, Jan. 2006.

[5] ME Martinez-Perez and A Espinosa-Romero, “Three-dimensional reconstruction of blood vessels extracted from retinal fundus images,” *Optics express*, vol. 20, no. 10, pp. 11451–11465, May 2012.

[6] T Chanwimaluang and G Fan, “Affine camera for 3-D retinal surface reconstruction,” in *Advances in Visual Computing*, pp. 19–30. Springer, 2006.

[7] Y Lin and G Medioni, “Retinal image registration from 2D to 3D,” in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[8] K Deguchi, J Noami, and H Hontani, “3D fundus pattern reconstruction and display from multiple fundus images,” in *International Conference on Pattern Recognition (ICPR)*, 2000, vol. 4, pp. 94–97.

[9] BM Haralick, C Lee, K Ottenberg, and M Nölle, “Review and analysis of solutions of the three point perspective pose estimation problem,” *International Journal of Computer Vision*, vol. 13, no. 3, pp. 331–356, 1994.

[10] JJ Moré, “The levenberg-marquardt algorithm: implementation and theory,” in *Numerical analysis*, pp. 105–116. 1978.

[11] S Hirano, Y Yamamoto, H Takayama, Y Sugata, and K Matsuo, “Ultrasonic observation of eyes in premature babies. part 6: Growth curves of ocular axial length and its components (author’s translation),” *Nippon Ganka Gakkai zasshi*, vol. 83, no. 9, pp. 1679–1693, 1979.