Constrained Optimization for Retinal Curvature Estimation
Using an Affine Camera

Thitiporn Chanwimaluang
National Electronics and Computer
Technology Center, Thailand

Guoliang Fan *
School of Electrical and Computer Engineering
Oklahoma State University

Abstract

We study retinal curvature estimation from multiple images that provides the fundamental geometry of human retina. We use an affine camera model due to its simplicity, linearity, and robustness. Moreover, the affine camera is suitable in this research because (1) NIH’s retinal imaging protocols specify a narrow $30^\circ$ field-of-view in each eye and (2) each field has small depth variation. A major challenge is that there is a series of optics involved in the imaging process, including an actual fundus camera, a digital camera, and the human cornea, all of which cause significant non-linear distortions in the retinal images. In this work, we develop a new constrained optimization procedure that considers both the geometric shape of human retina and lens distortions. Moreover, the constrained optimization is implemented in the affine space because it is computationally efficient and robust to noise. Specifically, we amend the affine bundle adjustment algorithm by including a quadratic surface fitting error and the lens distortion correction into the cost function for constrained optimization. The experiments on both synthetic data and real retinal images show the effectiveness and robustness of the proposed algorithm.

1. Introduction

Diabetes is the leading cause of blindness among working-age Americans, and many patients with vision-threatening diabetic retinopathy (DR) remain asymptotic until blindness occurs. The majority of this blindness can be prevented with proper eye examination by ophthalmologists or specialists who rely on the results of randomized clinical trials by the National Insinuates of Health (NIH), called Early Treatment Diabetic Retinopathy Study (ETDRS), to guide their treatment of DR patients. \(^1\) The ETDRS protocols require the retinal images to be captured from seven fields that cover a required area on the retina. The ETDRS imaging standard, specifying seven stereoscopic $30^\circ$ fields for each eye, is illustrated in Fig. 1(a). The long term goal of this research is to develop a visual 3-D retinal model as shown in Fig. 1(b), which can (1) assist ophthalmologists and specialists in diagnosing and evaluating the DR disease; (2) facilitate clinical studies; and (3) be used as a spatial map during laser surgical procedures.

3-D geometric reconstruction is a process to recover a 3-D scene or object from multiple images. It is usually referred to as the structure from motion (SfM). The SfM process usually recovers objects’ 3-D shapes, cameras’ poses (positions and orientations), and cameras’ internal parameters (focal lengths, principle points, and skew factors). Many possible camera models exist. A perspective projection is the standard camera model. However, other simplified projections, e.g., affine or orthographic models, are proved useful and practical for a distant camera. In this paper, we are interested in a specific SfM issue, i.e., retinal curvature estimation, which is concerned with the global geometry recovery of human retina. The retinal curvature provides fundamental geometric knowledge of the human retina that could be further used as the baseline reference for local depth recovery, especially in the pathological areas, or 3-D visualization of the human retina.

---

*This work was supported by an OHRS award (HR03-33) from the Oklahoma Center for the Advancement of Science and Technology (OCAST).
2. Related Work

2.1. 3D Retina Estimation

There are two main modalities used for 3D retina estimation, i.e., the stereoscopic fundus camera or the laser scanner. For example, the Heidelberg Retina Tomograph (HRT) is a confocal laser scanning system designed for 3D estimation of retinal topography. In this work, we are interested in retinal images captured by fundus cameras that are relatively cheap and more often used. Deguchi et al. [7, 8] modeled both the fundus camera and the human cornea as a virtual optical lens. It is assumed that a retinal surface has a spherical shape, and imaging a sphere through the virtual lens results in a quadratic surface. Then, the parameters of the virtual lens were estimated iteratively to recover fundus's spherical surface. Choe et al. [6] used PCA-based directional filters to extract candidate feature points, and the plane-and-parallax algorithm was employed to estimate the epipolar geometry based on which the stereo pair was rectified. Then, a Parzen window-based mutual information was used to generate a dense disparity map. Promising 3-D retinal reconstruction results were reported in [6].

In practice, we found traditional stereo reconstruction techniques may not work well for ETDRS stereo image pairs due to three major problems. First, each ETDRS image covers a small region on a retina, and the overlaps between seven fields are relatively small (as shown in Fig. 1), leading to sparse feature points. Second, there is a series of optics involved in ETDRS imaging, including a fundus camera, a digital camera, and the human cornea, all of which cause non-linear distortions in the images. Third, the curvature of a retinal surface in each field is relatively small, i.e., relative low depth variation, making the fundamental matrix estimation difficult. Some of these difficulties are also acknowledged in [6] where some erroneous results were reported using a traditional 8-point algorithm to compute the epipolar geometry. This research is concerned with the global geometry estimation of the human retina, and it is somewhat complementary to the one proposed in [6] which focuses on the recovery of relative local depth.

2.2. Affine Camera for Multi-view Geometry

Because an affine camera is assumed here, an affine SfM problem is the focus of this work. The affine camera model was first proposed by Koenderink and Van Doorn [13]. They have shown that two distinct views are enough to reconstruct a scene up to an arbitrary affine transformation without camera calibration. They have suggested the use of a local coordinate frame (LCF). Later their algorithm has been refined by Demy et al. [9], and Shapiro [19]. Tomasi and Kanade [22], then, proposed an affine factoriza-

2http://www.heidelbergengineering.com

tion method which eliminates the use of LCF and instead utilizes the entire set of points. An affine camera is an appropriate simplified model for retinal curvature estimation from ETDRS images. In other applications such as face modelling, the affine camera is also used. Sengupta et al. [17] assumed an affine camera and solved the SfM problem based on the affine epipolar geometry framework.

2.3. Constrained SfM

If there is some prior knowledge about the 3-D geometry, adding a surface model or geometrical constraints into the SfM process would yield a more geometrically meaningful solution. Fua et al. [10] addressed the SfM problem in the context of head modeling. Based on the prior knowledge of the head’s shape, the standard bundle adjustment is augmented with iterative re-weighted least square and regularization, and it does not output an explicit model. Shan et al. [18] proposed a model-based bundle adjustment algorithm for face modeling where a parametric surface model controlled by a small set of parameters is involved and the optimization is within the model space. The 2-D location of five feature points (two for eyes, two for mouth and one for nose tip) must be manually supplied into both algorithms. Gong et al. [11] used sequential quadratic programming (SQP) to recover 3-D quadratic surface parameters by using a quadratic surface as a constraint in a metric space. Bartoli et al. [1] merged the multi-coplanarity constraints with the traditional bundle adjustment approach. In [11, 1], both feature-based and model-based ideas are invoked where the features are firstly estimated as a first guess for structure and motion, then geometric primitives are estimate to correct the structure so that reconstructed features lie exactly on the geometric primitives. Inspired by previous work, we propose an affine SfM algorithm via bundle constrained adjustment which involves joint optimization of both features and a parametric model as well as lens distortion update for optimal structure estimation.

2.4. Registration for Correspondence Selection

Correspondence selection is a critical step for estimating the retinal geometry from multiple images, and it is largely depends on the results of retinal image registration. Retinal image registration has been well studied by many researchers recently, e.g., [2, 3, 20]. However, in the context of ETDRS retinal imaging, three major challenges are present. First, small overlaps between adjacent fields lead to inadequate landmark points (crossovers and bifurcations) for feature-based methods. Second, the contrast and intensity distributions within an image are not spatially uniform or consistent. This can deteriorate the performance of area-based techniques. Third, high-resolution ETDRS images contain large homogeneous nonvascular/textureless regions which result in difficulties for both feature-based and area-
based techniques. In this work, we adopt a hybrid registration approach proposed in [5] that integrates both feature-based and area-based techniques to register ETDRS image pairs and to select correspondences across images. Due to the small overlaps (less than 30%) in most fields, it is reliable to only use stereo image pairs in fields 1/2 or 2/3 (as shown in Fig. 1) for curvature estimation.

3. Affine Structure From Motion

We propose several SfM optimization algorithms in the affine space to study retinal curvature estimation. Specifically, three optimization procedures are tested, i.e., affine bundle adjustment (ABA), constrained affine bundle adjustment (CABA), and constrained affine bundle adjustment with lens distortion update (CABA-LDU). The CABA-LDU process optimizes all of the parameters, including camera’s parameters, 3-D points, the physical shape of a retinal surface, and lens distortion, simultaneously. We also introduce an efficient point-based linear approach to approximate the retinal spherical surface.

3.1. Initial Lens Distortion Removal

![Image](https://example.com/fig2.png)

Figure 2. The ETDRS imaging system (www.inoveon.com).

There is a series of optics involved in the retinal imaging process, which includes the actual fundus camera, the digital camera, and the human cornea, as shown in Fig. 2. All of these optics could be modeled as one virtual lens that contributes to certain lens distortion, e.g., radial distortion, in retinal images [7, 8]. The lens distortion has to be removed prior to 3-D retinal surface reconstruction. In this work, we employ the planar pattern calibration method proposed in [26]. The solution can be solved through minimizing an algebraic distance then refining it through the Levenberg-Marquardt algorithm with a following cost function:

\[ \sum_{i=1}^{f} \sum_{j=1}^{n} \| m_{ij} - \hat{m}(K, R_i, d_i, M_j) \|^2, \]

where we have \( f \) views/images and \( n \) correspondences. \( \hat{m}(K, R_i, d_i, M_j) \) is the projection of point \( M_j \) in the \( i \)th image. \( R_i \) denotes \( i \)-th row of rotation matrix \( R \) and \( d \) denotes \(-R^T \). \( k \) are coefficients for lens distortion. We have created a set of chessboard images using the actual fundus camera. Then we use the camera calibration toolbox \(^3\) to remove the lens distortion in real retinal images. However, we believe that only the lens distortions associated with the digital camera and the fundus camera can partially removed while the lens distortions caused by the human cornea are still present in the images and should be considered during the reconstruction process.

3.2. Initial Retinal Affine Surface

In this work, we assume an affine camera model because (1) the ETDRS imaging standard specifies a 30° field of view for each eye (narrow field of view); (2) each retinal image has small depth variation. We use affine factorization [22] method for initial reconstruction because the approach can accommodate multiple images and utilize the use of all feature points. Suppose there are \( f \) retinal images and \( n \) point correspondences from each image.

\[ W = PM, \]

where \( W \) denotes a \( 2f \times n \) matrix containing a set of 2D correspondences. \( M \) denotes a \( 3 \times n \) matrix containing the affine shape of the retinal surface. \( P \) denotes a \( 2f \times 3 \) matrix comprising \( f \) fundus camera models. With the rank theorem, \( W \) is at most rank three. Singular value decomposition (SVD) is used to factorized \( W \), therefore, \( P \) and \( M \) are the left and right eigenvectors corresponding to the three greatest eigenvalues.

3.3. Affine Bundle Adjustment (ABA)

Bundle adjustment is an optimization process of refining a visual reconstruction to produce jointly optimal structure and viewing parameters [24, 23]. It is usually formulated as a nonlinear least square problem. In our case, we want to estimate and refine affine cameras \( \hat{P} \) and the affine retinal surface \( \hat{M} \) simultaneously. \( \hat{m}(\hat{P}_i, \hat{M}_j) \) is a projection of point \( \hat{M}_j \) in the \( i \)th image. We try to minimize distance between the projected point \( \hat{m}(\hat{P}_i, \hat{M}_j) \) and the observed point \( m_{ij} \) by optimizing the 3-D point in the affine space and the affine cameras. If the 2D correspondences are noise free, this distance should be zero. If noise distribution is assumed to be zero mean, isotropic and Gaussian with certain variance, then the maximum likelihood estimation is equivalent to the solution to the minimum mean square error problem defined below:

\[ \min_{\hat{P}, \hat{M}} \sum_{i=1}^{f} \sum_{j=1}^{n} \| \hat{m}(\hat{P}_i, \hat{M}_j) - m_{ij} \|^2, \]

\(^3\)http://www.vision.caltech.edu/boug/
where we have $v$ views/images and $n$ correspondences. To solve this problem more efficiently, we can utilize the fact that a specific residual in one image is only dependent on one camera. This yields a sparse structure in the matrices involved in the optimization process. The estimated affine structure could be corrected into the Euclidean space that is possible given the minimum number of four images. In ETDRS retinal imaging, two stereo pairs from fields 1 and 2 will provide a set of four images where we have the largest overlaps for robust 3-D reconstruction [4].

### 3.4. Constrained ABA (CABA)

If point correspondences are error free and lens distortions can be completely removed, motion parameters and a 3-D structure can be accurately obtained through a standard SfM procedure. Point correspondences are very sensitive to noise and some lens distortion is still present. Moreover, the standard optimization procedure, i.e., bundle adjustment, does not carry any 3-D geometrically meaningful description. In many SfM problems, some prior knowledge could be involved to compensate the errors from point correspondences and lens distortions by developing a geometrically meaningful cost function. In our case, we do have some prior knowledge of the shape of human retina.

Most of the works proposed in the subject of geometric constrained optimization are done in a final step or in the Euclidean space. The exceptions are the ones that assume planarity constraints [1, 21]. Szeliski et al. have suggested that prior geometric knowledge, which is incorporated in the early reconstruction process, can improve the quality of the estimated structures. In this work, we deal with retinal images which are captured from the back of eyeballs. Because the human eyeball can be approximated by a sphere [7], a spherical constraint could be considered for retinal curvature estimation. Since the optimization is implemented in the affine space, the spherical constraint becomes an ellipsoid constraint instead. Since the SfM optimization process is in the affine space, it requires less computational complexity compared to that in the Euclidean space, we choose to implement the constrained optimization in an affine space. Therefore, the ellipsoid surface constraint is included into the original cost function to improve the robustness and accuracy of the SfM results as follows,

$$\min_{\tilde{P}_i, \tilde{M}_j, Q_e, \rho} \sum_{i=1}^{v} \sum_{j=1}^{n} \left\| \tilde{m}(\tilde{P}_i, \tilde{M}_j) - m_{ij} \right\|^2 + \rho \sum_{j=1}^{n} \tilde{M}_j^T Q_e \tilde{M}_j,$$

where we have $v$ views/images and $n$ correspondences. $Q_e$ is a symmetric $4 \times 4$ matrix representing an ellipsoid surface. $\tilde{m}(\tilde{P}_i, \tilde{M}_j)$ is a projection of a point $\tilde{M}_j$ from an image $i$. $m_{ij}$ represents a retinal image point and $\rho$ is a Lagrange multiplier.

$Q_e$ is initialized as a spherical surface by using the point-based linear method introduced in Section 3.7. During the iterations, parameters in matrix $Q_e$ are updated to represent an ellipsoid surface. The Lagrange multiplier $\rho$ acts like a weighting parameter in the cost function. It can be chosen based upon the dynamic ranges of the two terms in Equation (4) where the first part is the residual error in all 2-D images and the second term the surface approximation error in the 3-D affine space. In our case, the two error terms have the similar dynamic range, and we initialize $\rho$ to be 1 that is updated during the iterations.

### 3.5. CABA with Lens Distortion Update (CABA-LDU)

All of the works dedicated to the subject of 3-D constrained optimization assume either lens distortion is removed before the reconstruction process, or it is insignificant and can be ignored. In the case of retinal imaging, however, lens distortion is too prominent to be neglected. Although, we have removed partial lens distortion prior to the affine reconstruction, we believe that the lens distortions caused by the human cornea still exist in the images. Therefore, we propose a new constrained optimization algorithm which includes lens distortion correction in the cost function. According to [12], only two types of lens distortions, radial and tangential, are significant in normal optical images. Radial and tangential distortions are shown in Equations (5) and (6) respectively.

$$\begin{align*}
\delta x_r &= k_{c1} r + k_{c2} r^2 + k_{c3} r^3, \\
\delta y_r &= k_{c1} r + k_{c2} r^2 + k_{c3} r^3, \\
\delta x_t &= 2k_{c4} x \tilde{y} + k_{c5} (r + 2x^2), \\
\delta y_t &= k_{c6} (r + 2y^2) + 2k_{c5} \tilde{x} \tilde{y},
\end{align*}$$

where $(\tilde{x}, \tilde{y})$ are image coordinates in the metric unit, $r = \tilde{x}^2 + \tilde{y}^2$, and $k_{c1}, k_{c2}, k_{c3}$ are coefficients for radial distortion. The expression for tangential distortion is

$$\begin{align*}
\delta x_t &= 2k_{c4} x \tilde{y} + k_{c5} (r + 2x^2), \\
\delta y_t &= k_{c6} (r + 2y^2) + 2k_{c5} \tilde{x} \tilde{y},
\end{align*}$$

where $k_{c4}$ and $k_{c5}$ are coefficients for tangential distortion. Therefore, we amend the ABA algorithm in two manners. First, an ellipsoid surface constraint is included into a cost function to improve robustness and accuracy (CABA). Second, the lens distortion update is incorporated into the CABA (CABA-LDU) to further remove remaining lens distortions. With both the 3-D geometric constraint and the lens distortion integrated in the optimization process, the cost function of CABA-LDU becomes:

$$\begin{align*}
\min_{\tilde{P}_i, \tilde{M}_j, Q_e, \rho, k_c} \left( \sum_{i=1}^{v} \sum_{j=1}^{n} \left\| \tilde{m}(\tilde{P}_i, \tilde{M}_j, \delta_r, \delta_t) - m_{ij} \right\|^2 + \rho \sum_{j=1}^{n} \tilde{M}_j^T Q_e \tilde{M}_j \right),
\end{align*}$$

where we have $v$ views/images and $n$ correspondences. $\tilde{m}(\tilde{P}_i, \tilde{M}_j, \delta_r, \delta_t)$ is a projection of a point $\tilde{M}_j$ in the $i$th
image following by the radial $\delta_r \triangleq [\delta_{x,r}, \delta_{y,r}]$ and tangential distortions $\delta_t \triangleq [\delta_{x,t}, \delta_{y,t}]$ defined in Equations (5) and (6) respectively. $m_{ij}$ represents a retinal image point. $Q_k$ is a symmetric $4 \times 4$ matrix representing an ellipsoid surface. $\rho$ is a Lagrange multiplier. Equation (7) shows that we associate two types of errors, both the 2-D error (the first term) and the 3-D error (the second term), into one optimization process. The cost function also incorporates both geometrically meaningful definition and lens distortion update. The procedure optimizes all of the parameters, camera’s parameters, 3-D points, the physical shape of human retina, and the lens distortion parameters, simultaneously.

### 3.6. Euclidean Reconstruction of Retinal Surface

After the affine reconstruction, we need to recover retina’s Euclidean surface from the affine structure obtained by above optimization algorithms. Several different solutions for different affine camera projections were proposed. Tomasi and Kanade [22] proposed a solution for orthographic projection. Weinshall and Tomasi [25] introduced a solution under weak-perspective camera. Poleman and Kanade [15] proposed a solution for paraperspective projection. Quan [16], Kurata et.al. [14] attempted to congregate those solutions into one unified framework for general affine camera without having to calibrate the camera. We employ the method proposed by Quan [16] to recover the Euclidean surface.

### 3.7. Point-Based Surface Approximation

Estimating the underlying spherical model from a set of 3-D points involves an error-prone nonlinear optimization process. We introduce a linear point-based sphere fitting method. The method is accomplished by first selecting a reference point $M_k = (X_k, Y_k, Z_k)$ from the 3-D point cloud. Every point ($j \in \{1, \ldots, n\}$) is expect to satisfy the sphere equation.

$$((X_j - A)^2 + (Y_j - B)^2 + (Z_j - C)^2) = R^2,$$

where $(A, B, C)$ and $R$ are the sphere’s center point and radius respectively. By subtracting the equation of reference point $M_k$ on both side of Equation (8) and rearranging the terms, we get

$$((X_k - X_j)^2 + (Y_k - Y_j)^2 + (Z_k - Z_j)^2) = 2(X_k - X_j)(A_j) + 2(Y_k - Y_j)(B_j) + 2(Z_k - Z_j)(C_j),$$

where is in a linear form. Sphere’s center point $(A, B, C)$ can be obtained by solving multiple linear equations. Then, radius $R$ can be computed in a least mean square sense. Ideally speaking, every point has to satisfy the sphere equation. An error at a particular point $j$ is calculated by the following equation:

$$E_j^{(k)} = M_j^T Q_k M_j,$$

where $Q_k$ is a $4 \times 4$ spherical matrix by a reference point $k$. $E_j^{(k)}$ represents an error at point $j$ by using $Q_k$. By minimizing the following equation, the optimal sphere surface with best fitness to all points can be achieved.

$$\hat{Q} = \arg \min_{k,1 \ldots n} \sum_{j=1}^{n} E_j^{(k)},$$

where $E_j^{(k)}$ is defined in Equation (10).

### 4. Experimental Results

We tested three variations of bundle adjustment, affine bundle adjustment (ABA), constrained ABA (CABA), and CABA with lens distortion update (CABA-LDU), on both synthesized data and real retinal images. The synthetic data allows us to measure the algorithm performance numerically because quantitative evaluation is impossible to obtain from real retinal images. We have generated a 3-D partial sphere point cloud with the spreading angle of 90°. Then, four virtual cameras are positioned according to the ETDRS imaging setting, as shown in Fig. 3(a). The four synthetic images captured by these cameras are shown in Fig. 3(b). In the following experiments, we add various lens distortion coefficients and different noise variances to correspondence measurements in order to compare different approaches under various circumstances. Four images are the minimum settings for Euclidean reconstruction using an affine camera.

Two numerical criteria are used to evaluate the effectiveness of the 3-D retinal surface reconstruction, i.e., the surface approximation error defined in and the spreading angle.
The surface approximation error is defined as follows,
\[
E = \frac{1}{n} \sum_{k=1}^{n} E_k,
\]  
(12)
and
\[
E_k = M_k^T \hat{Q} M_k,
\]  
(13)
where \( \hat{Q} \) is defined in Equation (11). \( \bar{E} \) gives the average surface fitness error with respect to the optimal sphere surface \( \hat{Q} \), and \( E_k \) is the fitness error of point \( M_k \). Both of them are used for performance evaluation in the following.

From Fig. 4(a) illustrates the top view of the reconstructed 3-D point clouds. We approximate the 2-D radius (circle’s radius) by using four farthest points, \( a, b, c, \) and \( d \), shown in Fig. 4(a) to calculate \( r_1 \) and \( r_2 \). In the case of synthesized data or a perfect sphere, the final circle’s radius is computed by averaging the two radii \( r = (r_1 + r_2)/2 \). In the case of real retinal images, radii \( r_1 \) and \( r_2 \) correspond to the size of overlaps vertically and horizontally, respectively. Once the sphere’s center \((A, B, C)\), the sphere’s radius \( R \), and the circle’s radius \( r \) can be estimated, we can approximate the spreading angle \( \theta \), shown in Fig. 4(b), of the reconstructed surface as follow
\[
\theta = 2 \arcsin \left( \frac{r}{R} \right).
\]  
(14)

After the spreading angles along both directions, \( \theta_1, \theta_2 \), are estimated, we can map a retinal image onto an approximated partial sphere specified by \( \theta_1, \theta_2 \).

4.1. Surface Approximation on Synthetic Data

We evaluate and compare three optimization algorithms in terms of above two criteria on the synthesized data of three different conditions, i.e., synthesized data with noise only, synthesized data with artificial lens distortion only, and synthesized data with both noise and lens distortion.

- In the first case (with noise only), zero-mean, isotropic, Gaussian noises with different variances are added to the position of correspondences in all images. Fig. 5(a) shows the errors of surface approximation versus noise variances. Fig. 5(b) shows errors of spreading angle in percentage versus noise variances. At each noise level, the algorithm is performed ten times to obtain an average error. It is shown that CABA improves the reconstruction performance and sustains good performance even under strong noises.

- In the second case (with lens distortion), we generate the noise-free synthesized data with the lens distortion coefficients set to \( k_c = [0.03, -0.08, 0.02, 0.01, -0.02] \). This time, we use the point-based approximation defined in Equation (13). In the first column of Fig. 6(a), first row compares errors between the procedure without optimization versus ABA. Second row relates errors between ABA and CABA. The last row associates errors between CABA and CABA-LDU. It is demonstrated that the surface approximation errors can be improved step by step through appropriate optimization procedures. CABA-LDU yields the best performance.

- In the last case (with both noise and lens distortion), we tested on CABA and CABA-LDU procedures. Zero-mean, isotropic, Gaussian noises with different variances and lens distortion coefficients set to \( k_c = [0.03, -0.08, 0.02, 0.01, -0.02] \) are added to images. Fig. 7 shows the errors of surface approximation versus noise variances. The plot demonstrates that CABA-LDU produces better surface approximation accuracy compared with CABA.

4.2. Surface Approximation on Retinal Images

We also tested three optimizations, ABA, CABA, and CABA-LDU on two sets of retinal retinal images each of which have four images (two stereo pairs from fields 1 and 2). The point-wise surface approximation error \( E_k \) defined in Equation (12) are shown in the second column of Fig.
Figure 6. (a) Noise-free synthesized data with lens distortion coefficients set to $k_e = [0.03, 0.08, 0.02, 0.01, -0.02]$. (b) The experimental results on two sets of retinal images.

Figure 7. Synthesized data with both noise and lens distortion: the errors between reconstructed points and surface approximation versus noise variances for CABA and CABA-LDU.

Figure 8. Two retinal model examples that are obtained by mapping retinal images onto the estimated sphere surface whose curvature is derived from the curvature estimated from the overlap.

4.3. More Discussions

There are two limitations in this work that need further investigation. First, since the optimization is in the affine space, the lens distortion parameters cannot be accurately estimated, although the retinal curvature estimation is improved by considering the lens distortion update. Second, the spherical constraint is enforced indirectly as surface fitness in the optimization process, and the surface approximation error in the cost functions of CABA and CABA-LDU algorithms may not directly reflect the shortest Euclidian distance between the 3-D points and the reconstructed surface. That means the cost function is partially geometrically meaningful. Nevertheless, this research is able to provide accurate and robust estimation of retinal curvature that can be further combined with other techniques for more detailed and accurate 3-D retinal reconstruction and visualization.
5. Conclusions

This paper presents a new constrained optimization algorithm for retinal curvature estimation where we have considered both the prior knowledge about 3-D geometry of human retina and the virtual lens distortion introduced by fundus camera and the human cornea. The affine camera is used for 3-D surface reconstruction. Specifically, we have defined a new optimization function for affine bundle adjustment that incorporates both the geometrically meaningful surface approximation error and the lens distortion update. The proposed algorithm can yield more accurate and robust retinal curvature estimation compared with the standard affine bundle adjustment. This work is the first step toward our long-term goal to build a visual 3-D retinal model.

References


