

3-D Retinal Curvature Estimation

Thitiporn Chanwimaluang, Guoliang Fan, *IEEE, Senior Member*, Gary G. Yen, *IEEE, Fellow*,
and Stephen R. Fransen

Abstract—We study 3-D retinal curvature estimation from multiple images that provides the fundamental geometry of the human retina and could be used for 3-D retina visualization and disease diagnosis purposes. An affine camera model is used for 3-D reconstruction due to its simplicity, linearity, and robustness. A major challenge is that a series of optics is involved in the retinal imaging process, including an actual fundus camera, a digital camera, and the optics of the human eye, all of which cause significant non-linear distortions in retinal images. In this work, we develop a new constrained optimization method that considers both the geometric shape of the human retina and non-linear lens distortions. Moreover, we examine a variety of lens distortion models to approximate the optics of the human eye in order to create a smooth spherical surface for curvature estimation. The experimental results on both synthetic data and real retinal images validate the proposed algorithm.

Index Terms—Structure from Motion, retinal curvature estimation, affine camera, bundle adjustment, lens distortion

I. 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 occurs¹. 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 Institutes of Health (NIH), called Early Treatment Diabetic Retinopathy Study (ETDRS), to guide their treatment of DR patients [1]. The ETDRS protocols specify seven stereoscopic 30° fields in each eye that motivate our research. Our long term goal is to develop a diagnostic tool for 3D retinal visualization, which can (1) assist ophthalmologists and specialists in diagnosing and evaluating the DR disease; (2) facilitate clinical studies with annotated anatomical and pathological features; and (3) be used as a spatial map during laser surgical procedures [6]. A key factor to build such model is the retinal curvature that provides the global geometry shape and could be further used as the baseline reference for local detailed depth recovery for anatomical and pathological structures.

T. Chanwimaluang was with the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater, Oklahoma. She is now with the National Electronics and Computer Technology Center (NECTEC), Thailand. E-mail: thitiporn.chanwimaluang@nectec.or.th. G. Fan and G. Yen are with the School of Electrical and Computer Engineering, OSU, Stillwater, Oklahoma. E-mail: {guoliang.fan,gyen}@okstate.edu. S. R. Fransen is with the Dean McGee Eye Institute, Oklahoma City, Oklahoma.

This work was presented in part in the International Symposium in Visual Computing (ISVC) in 2006 and the IEEE International Workshop on Beyond Multiview Geometry: Robust Estimation and Organization of Shapes from Multiple Cues in 2007. This work was supported by an OHRS Award of Project number HR03-33 from the Oklahoma Center for the Advancement of Science and Technology (OCAST).

¹<http://www.nei.nih.gov/diabetes/>

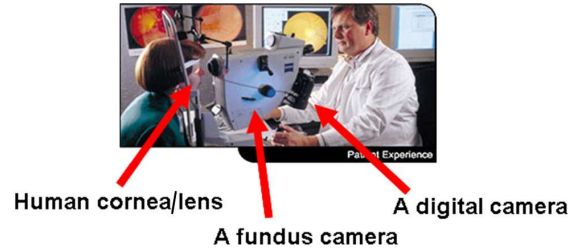


Fig. 1. The retinal imaging setup (Courtesy of Inoveon, www.inoveon.com).

In this work, we are interested in estimating the 3-D retinal curvature from multiple ETDRS images that is a specific issue of structure from motion (SfM). The SfM process usually recovers objects' 3-D shapes, cameras' extrinsic parameters (positions and orientations), and intrinsic parameters (focal lengths, principle points, and skew factors). Many possible camera models exist in literature [16]. A perspective projection is the standard model that is often used. Other simplified ones, e.g., affine or orthographic models, have been proven useful and practical for a distant camera. Particularly, we choose the affine camera due to its simplicity and linearity as well as its suitability to the ETDRS imaging protocols.

Three technical issues are addressed in this paper. First, we want to add a geometry-sensitive term in the cost function that allows the SfM result to resemble the shape of the human retina. Second, a series of optics is involved in ETDRS imaging, including an actual fundus camera, a digital camera, and the optics of the human eye (Fig. 1), all of which cause significant non-linear distortions in retinal images. Most SfM methods usually assume that lens distortion is insignificant (not true in our case) or can be removed first. Instead, we want to include the correction of lens distortion in the SfM process. Third, most SfM works have the geometric constrained optimization imposed in a final step or in the Euclidean space. Some exceptions are the ones that assume planarity constraints [3], [30]. We opt to implement the constrained optimization in the affine space, where the computational complexity is lower. It is worth mentioning that the geometry term in the cost function has to be considered in the affine space. It is fortunate that the shape geometry shares a similar mathematical form in both the Euclidean and affine spaces, simplifying the problem formulation and the optimization process.

The remainder of this paper is organized as follows. Section II briefly reviews the related work. We discuss the affine camera in Section III. The proposed algorithms are presented in Section IV. Algorithm validation is presented in Section V. The experimental results are shown in Section VI. We discuss conclusions and future work in Section VII.

II. RELATED WORKS

A. 3-D Retina Estimation

There are two main modalities used for 3-D retina estimation, i.e., the stereoscopic fundus camera and laser scanner (e.g., Heidelberg Retina Tomograph (HRT))². We are interested in the former one that are cheaper and more often used. Deguchi et al., [10], [11] modeled both the fundus camera and optics of the human eye as a single virtual optical lens. Then, the parameters of the virtual lens were estimated iteratively to recover fundus's spherical surface. Choe et al., [9] used the plane-and-parallax algorithm to estimate the epipolar geometry to rectify the stereo pair, and then applied a Parzen window-based mutual information to generate a dense disparity map. In practice, we found traditional SfM techniques may not work well for ETDRS images due to three reasons. First, each image covers a small region with sparse feature points. Second, serious non-linear lens distortion may corrupt the constraints of multi-view geometry. Third, the retinal curvature in each field is small (relative flat), making the fundamental matrix estimation difficult. Some of these challenges were also mentioned in [9] where erroneous results were reported using a traditional 8-point algorithm to compute the epipolar geometry.

B. Affine Camera for Multiview Geometry

The affine camera was first proposed by Koenderink and Van Doorn [21]. It was shown that two views are enough to reconstruct a scene up to an arbitrary affine transformation without camera calibration. They suggested the use of a local coordinate frame (LCF) to simplify feature representation. Their algorithm has been refined by Demey et al. in [12] and Shapiro et al. in [28]. Tomasi and Kanade [31] proposed an affine factorization method that eliminates the use of LCF and utilizes the entire set of features. The affine camera is also used in face modeling, e.g., [26] where the SfM problem is solved based on the affine epipolar geometry framework.

C. Constrained SfM

Bundle adjustment (BA) is referred to as a process of refining a 3D reconstruction to produce jointly the optimal structure and camera parameters [33]. If some prior knowledge about the underlying geometry is known, the SfM performance could be improved by involving some geometrical constraints in the cost function. Fua et al. [14] addressed the SfM problem in head modeling where the standard BA is augmented with iterative regularization to incorporate the generic knowledge of the head shape. Shan et al. [27] proposed a model-based BA algorithm for face modeling where a parametric surface model controlled by a small set of parameters is involved. Gong et al. [15] used sequential quadratic programming (SQP) to recover 3-D quadratic surface parameters by using a quadratic constraint. Bartoli et al. [3] incorporated the multi-coplanarity constraints. However, most of previous SfM works ignored the non-linear lens distortion in the images. In this work, we propose an affine SfM algorithm via constrained BA algorithm, which involves joint optimization of both features and a parametric 3-D model as well as lens distortion.

²<http://www.heidelbergengineering.com>

D. Image Registration and Correspondence Selection

Correspondence selection is a critical step, and it largely depends on retinal image registration that has been well studied e.g., [4], [5], [29]. In the context of ETDRS, three major challenges present. First, small overlaps between adjacent fields lead to inadequate landmark points for feature-based methods. Second, the non-uniform contrast and intensity distributions within an image can deteriorate the performance of area-based techniques. Third, high-resolution images contain large homogeneous nonvascular/textureless regions, which result in difficulties for both feature-based and area-based techniques. In this work, we used our previously proposed hybrid registration approach [7], [8] that integrates both feature-based and area-based techniques to register ETDRS image pairs and to select correspondences. Due to the small overlaps in most field pairs, it is reliable to use stereo image pairs in fields 1/2 or 2/3 defined in ETDRS protocols [1] for curvature estimation.

III. PROBLEM FORMULATION

In this section, we discuss two issues related to our problem formulation. One is why the affine camera is appropriate for our research, and the other is how to approximate the retinal shape when observed via a series of optics.

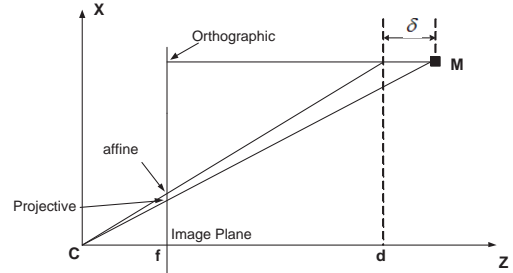


Fig. 2. Image formation in 1D. Orthographic projection is along z -axis. Projective one is projected along the principal ray direction. In the affine camera, a point is first projected along z -axis to a plane $Z = d$, and then undergoes projective projection onto the image plane.

A. Affine camera

Three projection models are shown in Fig. 2, i.e., projective, orthographic projection, and affine [16]. We want to derive the condition for good SfM performance using the affine camera. Let us start with a general projective camera as follows [16]:

$$\mathbf{T}_{proj} = \begin{bmatrix} f_x R_1^T + s R_2^T + c_x R_3^T & f_x D_x + s D_y + c_x D_z \\ f_y R_2^T + c_y R_3^T & f_y D_y + c_y D_z \\ R_3^T & D_z \end{bmatrix} \quad (1)$$

where R_i^T is the i -th row of the rotation matrix \mathbf{R} and $D_x = -R_1^T t_x$, $D_y = -R_2^T t_y$, $D_z = -R_3^T t_z$ where t_x , t_y and t_z are translation parameters. $f = (f_x, f_y)$ represents the camera's focal length. (c_x, c_y) and s are the principal point and skew angle of the image's pixel, respectively. The image and world coordinates are related by

$$m = \begin{bmatrix} (f_x R_1^T + s R_2^T) M + f_x D_x + s D_y + c_x \\ f_y R_2^T M + f_y D_y + c_y \\ R_3^T M + D_z \end{bmatrix}, \quad (2)$$

where m and M denote image and world homogeneous coordinates, respectively. If we assume the relative depth in principal ray direction $R_3^T M$ is small compared to D_z , then (2) can be rewritten as:

$$m = \begin{bmatrix} (f_x R_1^T + s R_2^T) M + f_x D_x + s D_y + c_x \\ f_y R_2^T M + f_y D_y + c_y \\ D_z \end{bmatrix}. \quad (3)$$

Therefore, we get an affine camera model defined as

$$\mathbf{T}_{\text{affine}} = \begin{bmatrix} f_x R_1^T + s R_2^T & f_x D_x + s D_y + c_x D_z \\ f_y R_2^T & f_y D_y + c_y D_z \\ \mathbf{0}_3^T & D_z \end{bmatrix}. \quad (4)$$

Comparing (2) and (3), the difference is only in the last row. Hence, we will rewrite both equations as follows

$$m_{\text{projective}} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ d + \delta \end{bmatrix} \quad m_{\text{affine}} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ d \end{bmatrix}, \quad (5)$$

$$\tilde{m}_{\text{projective}} = \begin{bmatrix} \frac{\tilde{x}}{d+\delta} \\ \frac{\tilde{y}}{d+\delta} \\ \frac{\tilde{d}}{d+\delta} \end{bmatrix} \quad \tilde{m}_{\text{affine}} = \begin{bmatrix} \frac{\tilde{x}}{d} \\ \frac{\tilde{y}}{d} \\ \frac{\tilde{d}}{d} \end{bmatrix}, \quad (6)$$

where $\tilde{x} = (f_x R_1^T + s R_2^T) M + f_x D_x + s D_y + c_x$ and $\tilde{y} = f_y R_2^T M + f_y D_y + c_y$, $d = D_z$ and $\delta = R_3^T M$. \tilde{m} is the non-homogeneous coordinate. Then,

$$\frac{\hat{m}_{\text{projective}}}{\hat{m}_{\text{affine}}} = \frac{d}{d + \delta}. \quad (7)$$

From (7), we conclude that if δ , or the relative depth, is small compared with D_z , then the affine camera is a good approximation. In our experiment, we found that each ETDRS image covers a spherical region that spans around 45° horizontally and 35° vertically. Under a field-of-view of 30° for the fundus camera, the ratio defined in (7) is normally in a range of 3% – 6% that validates the affine camera for this research.

B. Retinal Shape Approximation

It is generally assumed that the human retina is spherical in shape [2]. However, due to a series of optics involved, we need to re-consider this assumption. Deguchi et al. [10], [11] used a single virtual lens to collectively characterize all optics involved, and pointed out that imaging a sphere through the virtual lens results in a quadratic surface, which can be further generalized by the ellipsoidal shape as:

$$M^T Q_e M = 0, \quad (8)$$

where M denotes 3-D world homogeneous coordinates; Q_e is a symmetric 4×4 matrix representing an ellipsoid surface, and has the form

$$\begin{bmatrix} q_1 & q_4/2 & q_5/2 & q_7/2 \\ q_4/2 & q_2 & q_6/2 & q_8/2 \\ q_5/2 & q_6/2 & q_3 & q_9/2 \\ q_7/2 & q_8/2 & q_9/2 & 1 \end{bmatrix}, \quad (9)$$

An ellipsoid has 9 degrees of freedom which correspond to the 10 independent elements in the Q_e matrix (less one for scale). Equ. (8) can be re-written as $q_1 x^2 + q_2 y^2 + q_3 z^2 + q_4 xy + q_5 xz + q_6 yz + q_7 x + q_8 y + q_9 z + 1 = 0$. In the following, condition (8) will be incorporated as a geometrical constraint into the SfM process to encourage the formation of an ellipsoid shape.

IV. PROPOSED FRAMEWORK

The proposed algorithm involves five steps, as shown in Fig. 3, *camera calibration/correspondence selection*, *affine factorization*, *constrained affine bundle adjustment (CABA)*, *Euclidean reconstruction*, and *surface approximation*.

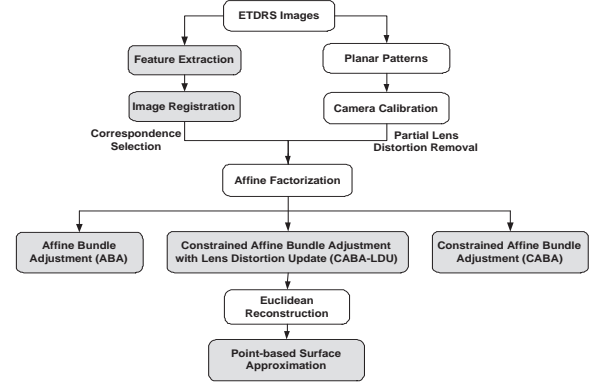


Fig. 3. The flowchart of the proposed framework.

A. Camera Calibration and Correspondence Selection

There is a series of optics involved in ETDRS imaging. To remove the lens distortion introduced by the fundus/digital cameras, the calibration method proposed in [38] is used. Specifically, we applied the camera calibration Matlab toolbox developed in Caltech³. It is believed that the lens distortions caused by optics (lens and cornea) of the human eye are still present after camera collaboration and should be considered during SfM. After calibration, we use the hybrid registration approach [7], [8] to register multiple images and select correspondences that are prominent blood vessel bifurcations shared by all images. Particularly, the affine transformation along with the iterative closest point (ICP) algorithm [29] are used for feature matching and correspondence selection. To ensure the latter SfM performance, we hope to have enough correspondences that can spread over the overlapping area. Therefore, we use two stereo pairs in fields 1 and 2 (totally four images) in this work where we see the maximum overlap and the most-populated features in an ETDRS image set.

B. Initial Retinal Affine Surface

The affine factorization method introduced in [31] is used for initial affine reconstruction. Suppose there are v retinal images and n point correspondences from each image. Then,

$$\mathbf{W} = \mathbf{P}\mathbf{M}, \quad (10)$$

where \mathbf{W} denotes a $2v \times n$ matrix containing a set of 2-D correspondences. \mathbf{M} denotes a $3 \times n$ matrix containing the affine shape of the retinal surface. \mathbf{P} denotes a $2v \times 3$ matrix comprising v affine cameras. With the rank theorem, \mathbf{W} is at most rank three. Singular value decomposition (SVD) is used to factorize \mathbf{W} , therefore, \mathbf{P} and \mathbf{M} are the left and right eigenvectors corresponding to the top three eigenvalues. The affine shape \mathbf{M} provides a starting point for SfM optimization.

³http://www.vision.caltech.edu/bouguetj/calib_doc/

C. Affine Bundle Adjustment (ABA)

Bundle adjustment is an optimization process of refining a visual reconstruction to produce a *jointly optimal* structure and viewing parameters [32], [33]. It is usually formulated as a nonlinear least square problem. In our case, we want to estimate and refine affine cameras $\hat{\mathbf{P}}$ and the affine retinal surface $\hat{\mathbf{M}}$ 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} (the j th point in the i th image) by optimizing the 3-D point in the affine space and the affine cameras as:

$$\min_{\hat{P}_i, \hat{M}_j} \sum_{i=1}^v \sum_{j=1}^n \|\hat{m}(\hat{P}_i, \hat{M}_j) - m_{ij}\|^2, \quad (11)$$

where $\hat{m}_{ij} = \hat{P}_i \hat{M}_j$. The computational complexity of normal bundle adjustment is $\mathcal{O}(N^3)$ in the number of unknowns. To solve this problem more efficiently, we can utilize the fact that a specific residual in one image is dependent on one camera. The sparse matrices involved in the optimization process leads to considerable computational benefits [33]. The computational complexity of ABA depends on the number of variables in (11), i.e., $(6v + 3n)$, where each affine camera and 3-D point have six and three parameters respectively.

D. Constrained ABA (CABA)

In many SfM algorithms, some prior knowledge could be involved to compensate the errors from point correspondences and lens distortions by developing a geometrically meaningful cost function. In most cases, the constrained optimization is done in the final step or in the Euclidean space. The exceptions are the ones that assume planarity constraints e.g., [3], [30] where prior geometric knowledge of a planar scene is incorporated in the reconstruction process to improve the quality of the estimated structures. In this work, we deal with retinal images that are captured from the back of eyeballs. Because the human eyeball can be approximated by a sphere [10], a spherical constraint can be considered that becomes an ellipsoid constraint in the affine space. We implement the constrained affine bundle adjustment algorithm as follows,

$$\min_{\hat{P}_i, \hat{M}_j, Q_e} \sum_{i=1}^v \sum_{j=1}^n \|\hat{m}(\hat{P}_i, \hat{M}_j) - m_{ij}\|^2 + \rho \sum_{j=1}^n \hat{M}_j^T Q_e \hat{M}_j, \quad (12)$$

where Q_e is a symmetric 4×4 matrix representing an ellipsoid surface defined in (9), and ρ is a regularization factor that controls the relative importance between the 2-D and 3-D errors. Q_e is initialized as a spherical surface by using the point-based linear method introduced in Section IV-G, and it is updated during iterations. The regularization factor ρ acts like a weighting parameter in the cost function. It can be chosen based upon the dynamic ranges of the two terms in Equ. (12), where the first part is the residual error in all 2-D images and the second term is the surface approximation error in the 3-D affine space. In our case, we set ρ to be 1 due to the similar dynamic ranges of the two terms in (12), and it is possible to use ρ to adjust the strength of the geometric penalty. CABA involves $(6v + 3n + 9)$ variables in (12) (Q_e is a symmetric 4×4 matrix with one element to be 1).

E. CABA with Lens Distortion Update (CABA-LDU)

Although we have removed partial lens distortion prior by offline camera calibration, we believe that the lens distortions caused by the human eye are still present that could invalidate the multi-view geometry. Therefore, we propose CABA-LDU algorithms to correct lens distortion during optimization.

1) *Lens Distortion*: According to [13], [17], [35], [37], there are three major lens distortions, *radial*, *decentering*, and *thin prism*. The radial distortion is caused by the imperfect lens shape. The decentering distortion is caused by the various degree of de-centering of the actual optical systems [23], [35]. The thin prism distortion is usually developed by the camera assembly [23]. The total distortion is the combination of those aforementioned distortions and it generates both radial and tangential errors in point positions defined in (13) and (14),

$$\begin{aligned} \delta x_r &= k_{c1}r + k_{c2}r^2 + k_{c3}r^3, \\ \delta y_r &= k_{c1}r + k_{c2}r^2 + k_{c3}r^3, \end{aligned} \quad (13)$$

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

$$\begin{aligned} \delta x_t &= 2k_{c4}\tilde{x}\tilde{y} + k_{c5}(r + 2\tilde{x}^2), \\ \delta y_t &= k_{c4}(r + 2\tilde{y}^2) + 2k_{c5}\tilde{x}\tilde{y}, \end{aligned} \quad (14)$$

where k_{c4} , and k_{c5} are coefficients for the tangential error.

We examine five different lens distortion models as shown in Table I in terms of their appropriateness and accuracy to represent the optics of the human eye. The numerical criterion for algorithm validation is the surface approximation error defined later, and the main assumption is that the shape of the human retina can be approximated by a spherical surface.

TABLE I
FIVE LENS DISTORTION MODELS

Five models	Non-zero coefficients	Number of variables
CABA-LDU ₍₁₎	k_{c1}, k_{c2}, k_{c3}	$6v + 3n + 9 + 3$
CABA-LDU ₍₂₎	k_{c1}	$6v + 3n + 9 + 1$
CABA-LDU ₍₃₎	k_{c4}, k_{c5}	$6v + 3n + 9 + 2$
CABA-LDU ₍₄₎	k_{c4}	$6v + 3n + 9 + 1$
CABA-LDU ₍₅₎	$k_{c1}, k_{c2}, k_{c3}, k_{c4}, k_{c5}$	$6v + 3n + 9 + 5$

2) *The Cost Function*: We can incorporate one of five distortion models into the CABA, and the new algorithm is called CABA-LDU that involves a cost function defined as

$$\min_{\hat{P}_i, \hat{M}_j, Q_e, k_c} \left(\sum_{i=1}^v \sum_{j=1}^n \|\hat{m}(\hat{P}_i, \hat{M}_j, \delta_r, \delta_t) - m_{ij}\|^2 + \rho \sum_{j=1}^n \hat{M}_j^T Q_e \hat{M}_j \right), \quad (15)$$

where $\hat{m}(\hat{P}_i, \hat{M}_j, \delta_r, \delta_t)$ is the projection of a point \hat{M}_j in the i th image followed by the radial $\delta_r = [\delta_{xr}, \delta_{yr}]$ and tangential distortions $\delta_t = [\delta_{xt}, \delta_{yt}]$ defined in (13) and (14), respectively. Equ. (15) shows that the cost function integrates both the geometric fitting error and the online lens distortion update. The procedure optimizes all the parameters, camera's parameters, 3-D points, the physical shape of the human retina, and the lens distortion parameters (i.e., 2-D correspondences) simultaneously. The number of variables involved depends on the lens distortion model used, as shown in Table I.

F. Euclidean Reconstruction of Retinal Surface

After affine reconstruction, we need to recover the Euclidean surface from the affine structure. Several solutions for different affine camera projections were proposed, for example, [24], [31], [34]. Quan [25] and Kurata et.al., [22] attempted to congregate those methods into one unified framework for general affine camera without camera calibration. We employ the method proposed in [25] in this work. If the images are assumed to be taken by the same affine camera, then the intrinsic parameters \mathbf{K} are the same for every view. The following constraint is obtained:

$$\arg \min_{\mathbf{X}} \sum_{i=1}^{v-1} \left(\frac{u_i^T \mathbf{X} u_i}{w_i^T \mathbf{X} w_i} - \frac{u_{i+1}^T \mathbf{X} u_{i+1}}{w_{i+1}^T \mathbf{X} w_{i+1}} \right)^2 + \left(\frac{u_i^T \mathbf{X} w_i}{w_i^T \mathbf{X} w_i} - \frac{u_{i+1}^T \mathbf{X} w_{i+1}}{w_{i+1}^T \mathbf{X} w_{i+1}} \right)^2, \quad (16)$$

where $\mathbf{P}_i = \begin{bmatrix} u_i^T \\ w_i^T \end{bmatrix}$ and $\mathbf{X} = \mathbf{D}\mathbf{D}^T$ where \mathbf{D} is any 3×3 non-singular matrix. Equ. (16) can be minimized by a maximum likelihood estimation with the assumption of zero mean, isotropic, and Gaussian noise. See more details in [25].

G. 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 (John D'Errico, Personal Communication, 2006). 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 ($j \in \{1, \dots, n\}$) is expected to satisfy the sphere equation:

$$(X_j - A)^2 + (Y_j - B)^2 + (Z_j - C)^2 = R^2, \quad (17)$$

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 sides of Equ. (17) and rearranging the terms, we get

$$\begin{aligned} (X_k^2 - X_j^2) + (Y_k^2 - Y_j^2) + (Z_k^2 - Z_j^2) = \\ 2(X_k - X_j)A + 2(Y_k - Y_j)B + 2(Z_k - Z_j)C, \\ j \in 1, \dots, n, j \neq k, \end{aligned} \quad (18)$$

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

$$E_j^{(k)} = M_j^T \mathbf{Q}_k M_j, \quad (19)$$

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

$$\hat{\mathbf{Q}} = \arg \min_{k \in 1, \dots, n} \sum_{j=1}^n E_j^{(k)}, \quad (20)$$

where $E_j^{(k)}$ is defined in Equ. (19).

V. ALGORITHM VALIDATION

We define three quantitative measures to validate our algorithms, the 3-D reconstruction error, the surface approximation error, and the curvature estimation error. The first one is used for synthetic data where ground-truth 3D information is available, while the latter two are used for retinal images.

A. 3-D Reconstruction Error

This measurement quantifies the 3-D reconstruction accuracy by aligning the reconstructed point cloud with the original one under the same coordinate system and computing the matching error after alignment. We have implemented an efficient method proposed in [18] where a closed-form solution is used for 3-D registration. Given two sets of 3-D points, M_i^A and M_i^B ($i = 1, \dots, n$), in coordinate systems A and B , we want to minimize the 3-D registration error as:

$$e_1 = \min \sum_{i=1}^n \|M_i^A - s\mathbf{R}_{B,A}(M_i^B) - \mathbf{t}\|^2, \quad (21)$$

where s is a scale factor, \mathbf{t} is a translational offset, and $\mathbf{R}_{B,A}$ denotes the rotation matrix between two coordinate systems, all of which can be estimated analytically in [18].

B. Surface Approximation Error

Given the optimal sphere surface is determined as (20), we can define the surface approximation by

$$e_2 = \sum_{j=1}^n M_j^T \hat{\mathbf{Q}} M_j. \quad (22)$$

Under the spherical shape assumption, a smaller e_2 indicates a better SfM result and more trustable curvature estimation.

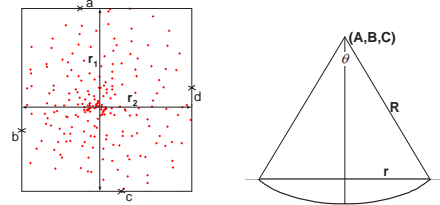


Fig. 4. 3-D point clouds: (Left) Top view. (Right) Side view.

C. Curvature Estimation Error

In order to estimate 3-D curvature, we first approximate the spatial coverage of the point cloud by using four farthest points in the top-down view, a , b , c , and d , to calculate height r_1 and width r_2 as shown in Fig. 4(left). In retinal images, radii r_1 and r_2 correspond to the size of overlaps vertically and horizontally, respectively that can be used to estimate the curvature along both directions. Given the estimated radius R , we can approximate the curvature θ as follows (Fig. 4(right)):

$$\theta = 2 \arcsin\left(\frac{r}{R}\right). \quad (23)$$

The well known anatomy reveals that the angle between the fovea and optic disk edge should be around 11.8° [2], [20]. This number is used to validate the estimated retinal curvature, in conjunction with the one defined in (22).

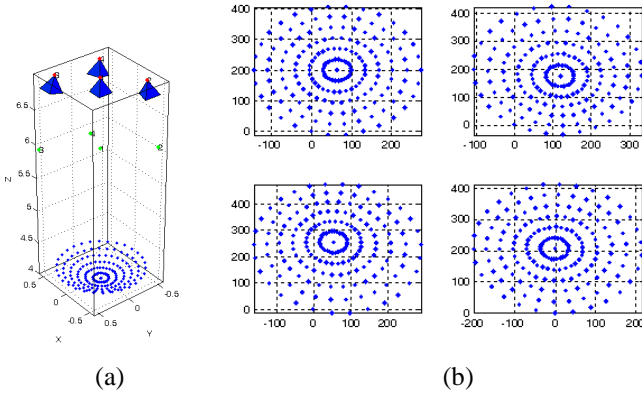


Fig. 5. (a) The experimental set up. A synthetic point cloud is constructed on a spherical surface with a curvature of 90° . (b) Four images generated by the four cameras shown in (a).

VI. EXPERIMENTAL RESULTS

We have tested ABA, CABA, CABA-LDU algorithms on both synthesized data and real retinal images. We have generated a 3-D uniformly distributed partial sphere point cloud with the curvature of 90° . Then, four virtual cameras are positioned according to the ETRS imaging setting, as shown in Fig. 5(a). The four synthetic images captured by these cameras via perspective projection, are shown in Fig. 5(b). The computational time of ABA, CABA, and CABA-LDU (with $v = 4$ and $n = 300$) of the Matlab (V.7) implementations is around 6, 15, and 55 seconds, respectively, on a 1.86GHz Intel Core2 Duo computer. In the following, we want to study three issues: (1) whether CABA can improve 3-D reconstruction results by incorporating the geometrical constraint compared with ABA; (2) whether CABA-LDU algorithms can further improve 3-D reconstruction performance by correcting lens distortion; and (3) which lens distortion model is appropriate to represent the optics of the human eye?

A. Surface Approximation on Synthetic Data

We have four experimental conditions: (1) noise-free and lens-distortion-free, (2) noise only, (3) lens distortion only, and (4) both noise and lens distortion. The 3-D registration error defined in (21) (e_1) is used for algorithm evaluation.

- In the first case (no noise and no lens distortion), Fig. 6 shows the 3-D reconstruction result of CABA that resembles the original one well (the numerical result is shown in Fig. 7(a)(b) with zero noise variance). The slight improvement from CABA over ABA may be because the geometric constraint in CABA compensates for the imperfection of the affine camera used in SfM.

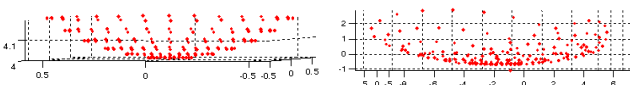


Fig. 6. The left shows the original point set of a partial sphere with a curvature of 90° , and the right is the surface approximation by using CABA.

- In the second case (with noise only), Fig. 7(a) shows the 3-D reconstruction error versus noise variances. It is clear that CABA outperforms the one without optimization and ABA. We also tested CABA-LDU₍₅₎ (the full lens distortion model) that produces the similar results as CABA with very small lens distortion parameters (usually in the range of 10^{-3} and 10^{-4} under low noise variances).

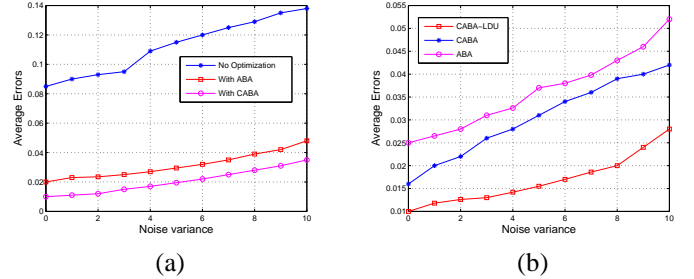


Fig. 7. The 3D reconstruction results on synthesized data with noise only (a) and with both noise and lens distortion (b). At each noise level, the algorithm is performed ten times to obtain an average.

- In the third case (with lens distortion), we have tried different lens distortion coefficients to simulate a real camera, and one typical result is shown here. We generated the noise-free synthesized data with the lens distortion coefficients set to $k_c = [0.03, -0.08, 0.02, 0.01, -0.02]$. The 3-D reconstruction errors of three algorithms are compared in Fig. 8(a)(b), where the X-axis and Y-axis show the 3-D registration errors for 3-D points using two different algorithms. Fig. 8(a) compares ABA with CABA, and Fig. 8(b) CABA with CABA-LDU₍₅₎. It is clear that the surface approximation errors can be improved step by step, and CABA-LDU₍₅₎ yields the best performance.

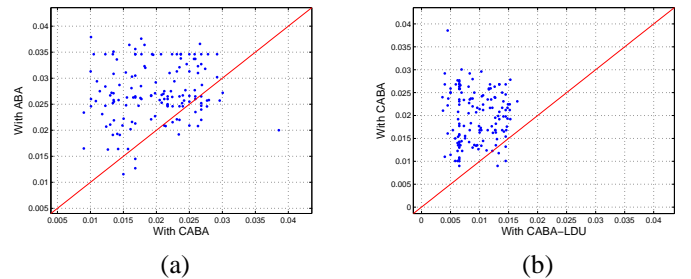


Fig. 8. The comparisons of three algorithms on the noise-free synthesized data with lens distortion. The X-axis and Y-axis represent the 3-D reconstruction error of a 3D point using algorithms A and B. (A is better than B if most points are above the line $y = x$.) (a) shows the comparison between ABA and CABA, and (b) presents the comparison between CABA and CABA-LDU₍₅₎.

- In the last case (with both noise and lens distortion), we tested CABA and CABA-LDU₍₅₎. 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(b) shows the errors of surface approximation versus noise variances. The plot demonstrates that CABA-LDU₍₅₎ produces better surface approximation accuracy compared with CABA, showing the usefulness of online lens distortion update.

B. Surface Approximation on Retinal Images

We have also examined CABA and CABA-LDU on six ETDRS image sets, each of which have four images. Two criterion are used for algorithm evaluation. One is the 3-D spherical surface approximation error defined in (19) (e_2), and the other is the curvature estimation error in terms of the angle between fovea and optic disc. The former one is used to find the optimal CABA-LDU algorithm by comparing five lens distortion models (as shown in Table II), and the latter one validates the effectiveness of the optimal CABA-LDU algorithm for retinal curvature estimation (as shown in Table III).

Table II reveals three observations. (1) All five CABA-LDU algorithms outperform CABA, showing the usefulness to incorporate lens distortion in SfM. (2) The tangential distortion models (CABA-LDU_(3,4)) are not very useful, showing that the optics of the human eye have little decentering and/or thin-prism distortions. It is quite understandable, because they are usually caused by camera assembly. (3) The lower-order radial distortion model (CABA-LDU₍₃₎) yields the best result, implying that the human eye has a near perfect lens shape.

TABLE II
THE 3-D SPHERICAL SURFACE APPROXIMATION ERROR (e_2).

Algorithms	Errors (e_2) for six retinal image sets (10^{-3})						Mean
	I	II	III	IV	V	VI	
CABA	258	401	356	312	151	186	277
CABA-LDU ₍₁₎	3	21	18	16	8	9	14
CABA-LDU ₍₂₎	8	11	11	13	5	7	8
CABA-LDU ₍₃₎	91	153	112	115	50	65	98
CABA-LDU ₍₄₎	126	201	173	162	96	98	143
CABA-LDU ₍₅₎	25	80	43	67	10	24	42
Average	87	145	119	114	53	65	

Table III presents the curvature estimated in the overlap part as well as the angle between fovea and optic disc for six image sets ($14.11^\circ \pm 2.21^\circ$). The well known anatomy reveals that the angle should be around 11.8° [2], [20]. The error could be due to magnification and refraction of the eye [2].

TABLE III
ESTIMATED CURVATURE

Image sets	Estimated curvature vertical/horizontal	Angle between fovea-optic disc
I	$25.11^\circ / 19.53^\circ$	14.21°
II	$25.81^\circ / 20.07^\circ$	11.17°
III	$24.96^\circ / 19.41^\circ$	13.37°
IV	$23.34^\circ / 18.15^\circ$	12.61°
V	$26.20^\circ / 20.38^\circ$	16.98°
VI	$25.02^\circ / 19.46^\circ$	16.29°
Ave.	$25.07^\circ / 19.50^\circ$	14.11°
Standard Deviation	$0.98^\circ / 0.77^\circ$	2.21°

We visualize curvature estimation results for six image sets for field 1 in Fig. 9 with estimated horizontal/vertical curvatures. It is not our intention to represent the optic nerve in this 3-D map, and the retinal curvature shows the global geometry that could serve as a baseline reference for local depth estimation of the optic nerve or other pathological areas. In that sense, our research is complementary to the work where a dense depth map is created for the retina surface, i.e., [9].

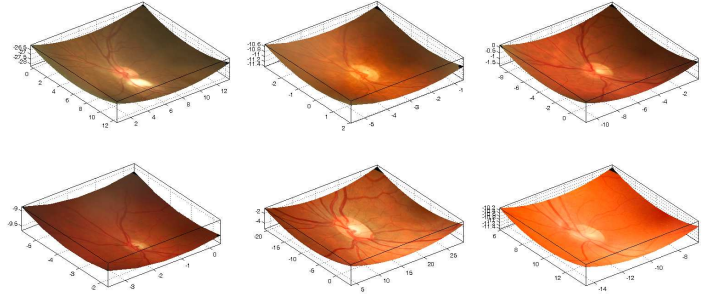


Fig. 9. Six results of mapping retinal images onto the estimated sphere surface (the optic nerve is not represented in this visualization).

C. More Discussion

There are four issues that need further investigation. First, because the SfM is optimized in the affine space, the estimated radial/tangential parameters may not well represent the optics involved. Second, the spherical constraint is enforced indirectly in the affine space, and the surface approximation error in the CABA and CABA-LDU may not directly reflect the shortest Euclidian distance between the reconstructed 3-D points and estimated surface. It means the cost functions in CABA and CABA-LDU are partially geometrically meaningful. Third, the assumption of the spherical model may not be universally valid for some eyes. It is possible to generalize the proposed method to handle the general ellipsoidal shape that requires some modification to the point-based surface approximation discussed in Section IV-G. Fourth, although the lower-order radial distortion model is found the best among five models tested in terms of the surface approximation error, there are still two questions unanswered. One is the whether the surface approximation error is an effective metric, and the other is whether there is a more accurate lens distortion model for the optics of the human eye that can be fully supported by the anatomical structure of the human eye (in stead of the existing lens distortion models).

The estimated curvature provides the important global shape information that has not been exploited in the past, and could be used as an additional mid-level cue for disease diagnosis. Moreover, the surface approximation error indicates the regularity of the retinal shape, and a large error may imply significant shape deformation due to certain eye disease [1].

VII. CONCLUSIONS AND FUTURE WORK

We have studied several affine camera-based algorithms, ABA, CABA, and CABA-LDU for 3-D retinal curvature estimation, and we have considered five lens distortion models in CABA-LDU the are used to compensate the non-linear distortion introduced by the human eye. Three quantitative measurements are used to validate the algorithms. The experimental results on both synthetic data and retinal images show that the geometrical constraint is useful to improve the SfM result, and a lower-order radial distortion (CABA-LDU₍₂₎) is able to further enhance 3-D retinal reconstruction by modeling the optics of the human eye.

This work is the first step toward our long-term goal to build a visual 3-D retinal model for comprehensive disease diagnosis. There is still much room for further investigation. First, it is known that the affine factorization method does not tolerate any missing correspondences, limiting its capability of handling more images. It is possible to use other robust factorization methods, i.e., the one in [19], to incorporate images from other fields for more robust and complete 3-D reconstruction. Second, the estimated curvature provides some auxiliary information for DR evaluation that could be added as a new cue in our previously developed retinal image sorting system [36]. Third, it is possible to combine existing local depth estimation algorithms that focus on pathological or anatomical structures with this research to produce a complete and detailed depth map. Although this research may have little immediate impact on disease diagnosis, it is an essential step to build a visual 3-D retinal model that could serve as a clinically viable tool for ophthalmologists and eye specialists.

ACKNOWLEDGEMENTS

The authors also thank the anonymous reviewers for their valuable comments and suggestions that improved this paper.

REFERENCES

- [1] Early Treatment Diabetic Retinopathy Study (ETDRS). footnote-<http://www.nei.nih.gov/neitrials/viewStudyWeb.aspx?id=53>.
 - [2] D. Atchison and G. Smith. *Optics of the human eye*. Butterworth-Heinemann, 2000.
 - [3] A. Bartoli and P. Sturm. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *Int'l Journal of Computer Vision*, 52(1):45–64, April 2003.
 - [4] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(3):347–364, March 2002.
 - [5] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based technique for joint, linear estimation of higher-order image-to-mosaic transformations: Mosaicing the curved human retina. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(3):412–419, March 2002.
 - [6] T. Chanwimaluang. *Advanced Retinal Imaging: Feature Extraction, 2D Registration and 3D Reconstruction*. PhD thesis, Oklahoma State University, Stillwater, OK, 2006.
 - [7] T. Chanwimaluang, G. Fan, and S. Fransen. Hybrid retinal image registration. *IEEE Trans. Information Technology in Biomedicine*, 10(1):129–142, January 2006.
 - [8] T. Chanwimaluang, G. Fan, and S. Fransen. Corrections to hybrid retinal image registration. *IEEE Trans. Information Technology in Biomedicine*, 11(1):110, January 2007.
 - [9] T. Choe, I. Cohen, and G. Medioni. 3-D shape reconstruction of retinal fundus. In *Proc. IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2277–2284, June 2006.
 - [10] K. Deguchi, D. Kawamata, K. Mizutani, H. Hontani, and K. Wakabayashi. 3D fundus shape reconstruction and display from stereo fundus images. *IEICE Trans. Inf. & Syst.*, E83-D(7):1408–1414, July 2000.
 - [11] K. Deguchi, J. Noami, and H. Hontani. 3D fundus pattern reconstruction and display from multiple images. In *Proc. Int'l Conference on Pattern Recognition*, volume 4, pages 94–97, September 2000.
 - [12] S. Demey, A. Zisserman, and P. Beardsley. Affine and projective structure from motion. *Proc. British Machine Vision Conference (BMVC)*, pages 49–58, 1992.
 - [13] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 1, pages 125–132, June 2001.
 - [14] P. Fua. Regularized bundle-adjustment to model heads from image sequences without calibration data. *Int'l Journal of Computer Vision*, 38(2), July 2000.
 - [15] R. Gong and G. Xu. Quadratic surface reconstruction from multiple views using sqp. *Integrated Image and Graphics Technologies*, pages 197–217, 2004.
 - [16] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision: Second Edition*. Cambridge University Press, 2003.
 - [17] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *Proc. IEEE Int'l Conference Computer Vision and Pattern Recognition*, pages 1106–1112, June 1997.
 - [18] B. K. P. Horn. Close-form solution of absolute orientation using unit quaternions. *Journal of Optical Society of America A*, 4(4):629–642, April 1987.
 - [19] D. Q. Huynh, R. Hartley, and A. Heyden. Outlier correction in image sequences for the affine camera. In *Proc. of IEEE International Conference on Computer Vision*, 2003.
 - [20] W. John Moran Eye Center. Facts and figures concerning the human retina. <http://webvision.med.utah.edu/Facts.html>.
 - [21] J. J. Koenderink and A. J. V. Doorn. Affine structure from motion. *Journal of Optical Society of America*, 8(2):377–385, February 1991.
 - [22] T. Kurata, J. Fujiki, and K. Sakae. Affine epipolar geometry via factorization method. In *Proc. of Int'l Conference on Pattern Recognition*, volume 1, pages 862–866, August 1998.
 - [23] A. S. of Photogrammetry. *Manual of Photogrammetry*. 1980.
 - [24] C. J. Poleman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(3):206–218, March 1997.
 - [25] L. Quan. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision*, 19(1):93–110, 1996.
 - [26] K. Sengupta and C. C. Ko. Scanning face models with desktop cameras. *IEEE Trans. Industrial Electronics*, 48(5):904–912, October 2001.
 - [27] Y. Shan, Z. Liu, and Z. Zhang. Model-based bundle adjustment with application to face modeling. In *IEEE Int'l Conf. on Computer Vision ICCV*, volume 2, pages 644–651, July 2001.
 - [28] L. S. Shapiro. *Affine Analysis of Image Sequences*. PhD thesis, Sharp Laboratories of Europe, Oxford, Oxford, UK, 1995.
 - [29] C. V. Stewart, C. L. Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Imag.*, 22(11):1379–1394, November 2003.
 - [30] R. Szeliski and P. H. S. Torr. Geometrically constrained structure from motion: Points on planes. In *Proc. Workshop on 3D Structure from Multiple Images of Large-scale Environment (SMILE)*, June 1998.
 - [31] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int'l Journal of Computer Vision*, 9(2):137–154, Nov. 1992.
 - [32] P. Tresadern and I. Reid. Uncalibrated and unsynchronized human motion capture: a stereo factorization approach. In *Proc. IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 1, pages 128–134, June 2004.
 - [33] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. *Bundle adjustment: A modern synthesis. Vision Algorithms: Theory And Practice*, Springer-Verlag, 2000.
 - [34] D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(5):512–517, May 1995.
 - [35] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(10):965–980, October 1992.
 - [36] G. G. Yen and W. F. Leong. A sorting system for hierarchical grading of diabetic fundus images: a preliminary study. *IEEE Trans. Information Technology in Biomedicine*, 12(1):118–130, Jan. 2008.
 - [37] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proc. IEEE Int'l Conference on Pattern Recognition*, volume 1, pages 407–411, August 1996.
 - [38] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proc. IEEE Int'l Conference on Computer Vision*, volume 1, pages 666–673, September 1999.
- **Thitiporn Chanwimaluang, Guoliang Fan and Stephen R. Fransen:** Author photographs and biographies can be found from *Hybrid Retinal Image Registration* published in Jan. 2006 in IEEE Trans. Information Technology in Biomedicine, Vol. 10, No. 1, pp129-142.
 - **Gary G. Yen:** Author photograph and biography can be found from *A sorting system for hierarchical grading of diabetic fundus images: a preliminary study* published in Jan. 2008 in IEEE Trans. Information Technology in Biomedicine, Vol. 12, No. 1, pp118-130.