

2-D Registration and 3-D Shape Inference of the Retinal Fundus from Fluorescein Images

Tae Eun Choe^{1*}, Gerard Medioni¹, Isaac Cohen¹, Alexander C. Walsh², Srinivas R. Sadda²

¹ Institute for Robotics and Intelligent Systems,
University of Southern California
3737 Watt way, Los Angeles, CA, 90248, USA
{tchoe, icohen, medioni}@usc.edu

² Doheny Eye Institute,
1450 San Pablo Street, Los Angeles, CA, 90033, USA
{AWalsh, SSadda}@doheny.org

Abstract

This study presents methods to 2-D registration of retinal image sequences and 3-D shape inference from fluorescein images. The Y-feature is a robust geometric entity that is largely invariant across modalities as well as across the temporal grey level variations induced by the propagation of the dye in the vessels. We first present a Y-feature extraction method that finds a set of Y-feature candidates using local image gradient information. A gradient-based approach is then used to align an articulated model of the Y-feature to the candidates more accurately while optimizing a cost function. Using mutual information, fitted Y-features are subsequently matched across images, including colors and fluorescein angiographic frames, for registration. To reconstruct the retinal fundus in 3-D, the extracted Y-features are used to estimate the epipolar geometry with a plane-and-parallax approach. The proposed solution provides a robust estimation of the fundamental matrix suitable for plane-like surfaces, such as the retinal fundus. The mutual information criterion is used to accurately estimate the dense disparity map, while the Y-features are used to estimate the bounds of the range space. Our experimental results validate the proposed method on a set of difficult fluorescein image pairs.

Keywords: 3-D reconstruction; Mutual information; Global registration; Retinal fundus image

*Corresponding author. Tel.: +1 213 210 9944; fax +1 213 740 7877. E-mail address: tchoe@gmail.com (Tae Eun Choe)
Address: IRIS, University of Southern California, PHE 224, 3737 Watt way, Los Angeles, CA, 90089

1 INTRODUCTION

We present methods that facilitate the analysis of retinal images. Automated analysis of retinal images can assist in the diagnosis and management of blinding retinal diseases, such as diabetic retinopathy, age-related macular degeneration (AMD), and glaucoma. For evaluating and imaging patients with retinal diseases, clinical photographers usually first capture color images of the retina using a specialized fundus camera. Subsequently, a fluorescein dye is injected into a vein in the subject's arm, and as the dye propagates through the retinal blood vessels, the photographer takes serial (over a 5-10 minute period)

pictures of the retina. The view point changes as the eye moves, and the camera may also move. Registration of the retinal image sequence needs to be performed to align the images in order to facilitate the study of the evolution of the dye propagation. Unfortunately, the intensities of the color image and the acquired grey level angiograms are not consistent, making registration of these images non trivial. Figure 1 shows various examples of color and fluorescein images of the retina. Significant variations in the intensity (and in some cases morphology) of the fluorescein pattern of retinal structures can be observed over the course of the angiogram, and thus registration of color and angiographic sequence can be considered truly multimodal. For example, the retinal veins in Figure 1 are red in the color image (Figure 1-(a)), hypointense in the early angiogram (Figure 1-(b)), increase in intensity (fill with dye) in mid angiogram (Figure 1-(c)), and then decrease in intensity (“fade”) in the late phases (Figure 1-(d)).

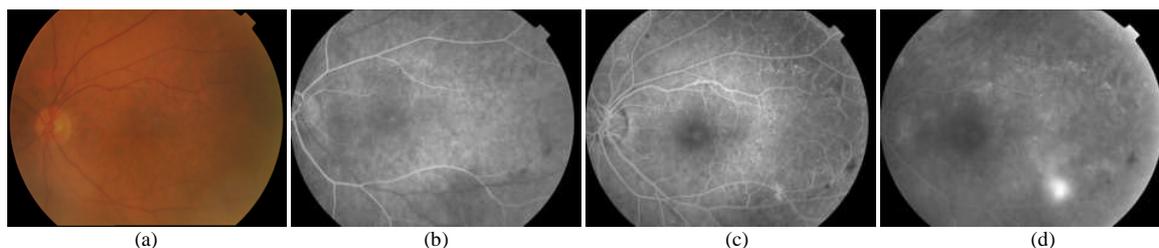


Figure 1. Color and Fluorescein Images of a Retinal Fundus

In the literature, most of the previous research in retinal images analysis has focused primarily on the issue of 2-D registration, which means registration at the image level. Since the retinal fundus has depth variations, 2-D registration cannot account for the true 3-D transformation, and residual errors must be expected. So, the inference of the 3-D shape of the retinal fundus may be useful in diagnosing and monitoring changes in macular edema, choroidal neovascular membranes, and the morphology of the optic nerve head in patients with optic neuropathies such as glaucoma. Research in the 3-D reconstruction of retinal images and in particular macular images, however, has been very limited. Fluorescein images of the retinal fundus have unique properties that prevent the use of classical stereo algorithms to estimate the 3-D shape of the retinal fundus from a pair of images. First, as demonstrated in Figure 1, the intensity and color of the same physical positions may vary considerably across consecutive images. Second, the shape of the retinal fundus is almost planar, and its 3-D depth is relatively shallow,

which makes classical methods unreliable to estimate the depth.

The methods proposed here cover two main topics. The first topic deals with the 2-D registration of color images and fluorescein angiographic image sequences. The second presents the 3-D reconstruction of the retinal fundus from a pair of images. Both methods utilize Y-features as matching features and *mutual information* for matching criteria; these methods are intended to improve the accuracy and reproducibility of diagnosis for ophthalmologists. The novel contributions of this paper are: (1) the use of Y-features; (2) a global 2-D registration method using the all pairs' shortest path algorithm; (3) the estimation of epipolar geometry by a plane-and-parallax method and search space reduction using accurately fitted Y-feature position; (4) the estimation of a dense disparity map by *mutual information*, and, (5) the evaluation of the accuracy of the 3-D reconstruction using ground-truth data from optical coherence tomography.

The paper is organized as follows. Section 2 reviews different approaches in retinal image analysis. Section 3 describes the 2-D registration of color and fluorescein retinal images with model-based Y-feature extraction, matching Y-features, and global registration. Section 4 describes the 3-D reconstruction of retinal fundus images with epipolar geometry estimation and dense stereo matching. Future research directions and the conclusion are outlined in section 5.

2 PREVIOUS WORK

2.1 2-D Registration

The key element for matching and registering color and fluorescein images is the extraction of invariant geometric features. Commonly used robust features are the optic disc, blood vessels, and Y-features. The optic disc is a frequently-used and robust feature to which all vessels and nerve fibers converge. Although many studies (Pallawala et al., 2004; Walter and Klein, 2001) have implemented image registration with the detection of the optic disc, its use is limited since it requires the optic disc to be visible in the image, and the images to be registered by a simple translation model (Foracchia et al.,

2004). The assumption of a simple translation model imposes a stringent restriction which is often not valid with fluorescein angiograms.

Another important feature in retinal images is blood vessels. One of the common method to extract vessels in a 2-D image is to use directional filters (Zana and Klein, 1999, 2001). Exploiting the property of linearity in vessels, blood vessels can be extracted by using a set of shifted Gaussian filters rotated at different orientations. Several investigators (Matsopoulos et al., 1999; Zana and Klein, 1999; Zana and Klein, 2001) have demonstrated the detection of blood vessels through a mathematical morphology approach using linear structural elements. This method performs well in detecting bright vessels. For detection of dark vessels, the method is applied on the negative version of the image. This method, however, has difficulty when there is great variation in the vessel colors; for example, during the early phase of the fluorescein angiogram, when both dark and bright vessels are present in the same image (Figure 1-(b)). Besides morphological methods and directional filters, blood vessels can be extracted by level sets (Malladi et al., 1995; Bemmell et al., 2003), vesselness measure (Frangi et al., 1998), and many other methods. Existing methods for vessel detection are not always suitable for extracting vessels in color and fluorescein images where the color, size, and the brightness of vessels vary across fluorescein images. These methods are also affected by low image contrast, which is a common attribute of the late phase images of the fluorescein sequence.

The retinal vascular tree has a regular pattern of bifurcations which form Y-features, and thereby has the potential to reliable features in the retinal image. Can et al. (1999a, 1999b, 2000) presented a method for Y-feature detection in retinal images. Y-features are extracted while tracing the vessels by finding the centerline between the left and right boundaries of the vessel. An important advantage of using the centerline of the vessel is that the Y-feature detection is less affected by variations in retinal vessel diameter which can occur as a result of the normal pulsatility or retinal blood flow. Because the centerline cannot be reliably extracted in the branch area, this tracing method produces multiple and inaccurate center positions, which subsequently cause errors in image registration. Tsai et al. (2003,

2004) improved this approach by specifying a branching area and extracting the center position of Y-feature within the area. With this method, the center position is estimated from the closest point of the three linearly approximated traces. Nevertheless, this method still produces multiple Y-features in one branch. Zana and Klein (1999) propose a method that detects the Y-features using the mathematical morphology approach with structural elements of Y and T shapes. However, this approach does not estimate the orientation of the Y-features accurately and it tends to generate a large number of false detections. Choe and Cohen (2005) have proposed a method for which the initial positions are estimated using directional filters and the accurate position and shape of the Y-features are obtained by fitting an articulated model to the image boundary. This is the method we use here.

The most common method for registering image sequences is pairwise registration, which sequentially registers images. Stewart et al. (2003) and Tsai et al. (2003) propose a dual-bootstrap ICP algorithm to register a pair of images using blood vessels as features. They improved the accuracy of the registration and solved the initialization problem of the ICP algorithm. The downside of this method is that when an incorrectly registered image is added to the mosaic image, the registration error propagates to all other images. A global registration method solves this problem by considering a graph that connects all overlapping images and estimates the global registration error. We discuss global registration in Section 3.3.

2.2 3-D Reconstruction

To obtain the 3-D shape of the retina, practitioners often use specialized devices such as a scanning laser ophthalmoscope (SLO) or an optical coherence tomography (OCT) system. Based on the principle of low-coherence interferometry, OCT provides an *in vivo* cross-sectional image of the retina that simulates microscopic visualization and has axial resolutions under $3\mu\text{m}$ (Walsh et al. 2005). Unfortunately, the cost of OCT equipment and the expertise required for its interpretation has limited the widespread adoption of this technology, particularly in developing countries. In addition, angiographic studies using these OCT systems have thus far proved to be challenging. An alternative to subjective

inspection and high-cost imaging consists of inferring the 3-D shape of the retina using images acquired with a commonly-available, lower-cost fundus camera.

A few research efforts have been reported on the concept of reconstructing the 3-D shape of optic disc from the retinal images (Kai, 2005). Previous methods are relatively simple as they only consider images that are synchronously acquired using a calibrated rig and therefore relative calibration is performed offline using a calibration pattern. The problem of 3-D shape reconstruction of the retinal fundus from a sequence of fluorescein images has not been addressed in the literature. We believe that a successful 3-D reconstruction of the retina from a pair of fluorescein images should provide a cost-effective and useful diagnostic tool for ophthalmologists.

The first step to obtain a dense stereo map is rectification, a process which aligns the matching points along horizontal scanlines. In order to rectify two stereo images, the epipolar geometry, represented by the fundamental matrix, is estimated. The classical approaches for estimating the fundamental matrix F (using seven or eight points) (Bartoli and Sturm, 2004; Hartley and Zisserman, 2000; Longuet-Higgins, 1981), are not accurate because the shape of the retina is relatively flat (Hartley and Zisserman, 2000). Even in disease states such as macular edema, where the central retina can be “dramatically” thickened, the retinal surface elevation in these situations is modest compared with the overall radius of curvature of the retina. Furthermore, although the retinal fundus is a curved 3-D shape, its surface curvature is relatively small. Kumar et al. (1994) propose a method for computing the fundamental matrix using the plane-and-parallax approach. The method consists of first estimating the 2-D projective motion, or the homography, and then inferring the epipoles by using two other matching points belonging to residual parallax regions. The fundamental matrix is then estimated using the epipoles and the homography. Werner and Matas (2005) use the RANSAC (RANdom SAmple Consensus) method (Fischler and Bolles, 1981) to select a set of 7 matching points that are not on a plane and estimate the fundamental matrix using the plane-and-parallax algorithm.

A comprehensive review of several stereo estimation methods is provided by Scharstein and Szeliski

(2002). In particular, graph cut (Kolmogorov and Zabih, 2001) and belief propagation (Sun et al., 2003) methods have shown good performance on test sets of stereo images with scenes having large depth changes, similar intensity for identical pixels, and textured regions. Retinal images, however, exhibit unique and specific challenges, such as varying intensities, shallow depth retinal fundus surface (especially in relatively normal retina), and low texture. Existing methods in the literature are not well adapted for stereo estimation of the retinal fundus from fluorescein images, as attested by our comparative study presented in the experimental results section.

3 2-D REGISTRATION

In this section, we propose a method for automatic registration of color and fluorescein angiograms in 2-D. The method consists of three main steps: First, seed positions for Y-feature are computed using a PCA-based analysis of directional filters responses. Second, an articulated model of a Y-feature is fitted to the image features using a gradient descent method. Third, the extracted Y-features are matched across the images by maximizing *mutual information*, and the geometric transformation between images is established using an affine model.

3.1 Estimating Initial Y-feature Positions

In this section, we present a method for locating Y-feature candidates in the image. Y-features are characterized by regions in the image where three vessels converge. The Y-feature should have exactly 3 strong responses from directional filters spanning 180° . The PCA based analysis of filter outputs allows us to locate the positions of the Y-features in the image.

3.1.1 Directional Filtering

The approximate position of the Y-features is obtained using a set of directional filters. We use a 2-D Laplacian of Gaussian (LoG) filter:

$$LoG = \frac{1}{2\pi\sigma_x\sigma_y} \left(\frac{x^2}{\sigma_x^4} + \frac{y^2}{\sigma_y^4} - \frac{1}{\sigma_x^2} - \frac{1}{\sigma_y^2} \right) e^{-\left(\frac{x}{2\sigma_x^2} + \frac{y}{2\sigma_y^2}\right)} \quad (1)$$

With the proper selection of σ_x and σ_y , ($\sigma_x=16$ and $\sigma_y=4$ for a 640×480 image), the elongated shape of the LoG filter is created. With this base kernel, directional filters are formulated by rotating the kernel from 0 to 180 degrees. A large number of directional filters generate finer responses from the Y-features in the image. However, this can cause duplication of the responses from one vessel. On the other hand, using a small number of filters could result in some missed Y-features. As a tradeoff, we use 6 directional filters at 6 different angles. In this paper, all parameters related with image size, such as the filter size, the length of the Y-feature, the search space, and window size, are specified for a 640×480 image. These parameters can be changed proportionally to account for the size of an image.

3.1.2 Principal Components Analysis

The initial position of Y-features is estimated from the 6 filter responses. Since a Y-feature should respond at exactly 3 different directions, we use the Principal Components Analysis (PCA) method to detect the pixels with large responses in exactly 3 different filter outputs. A local analysis of the directional filter responses permits integration of responses from neighboring pixels. The 3×3 neighboring pixels for each filter response are then transformed to 1-D 9 pixels, and autocorrelation of the 6×9 matrix is calculated. The resulting 6×6 autocorrelation matrix is decomposed to eigenvalues and eigenvectors using the Singular Value Decomposition (SVD) method. For each pixel, we examine the third largest eigenvalue; if that eigenvalue is large, the pixel is considered as a good candidate for a Y-feature. By sorting pixels by the third largest eigenvalues, we select the best 100 features for each image. Clearly, this is an approximation. We produce points which do not correspond to vessel bifurcation, and miss some also, but we found that the method produce a sufficient number of correct features in practice. In Figure 2, we show the extracted seed points using the PCA analysis of directional filter outputs.

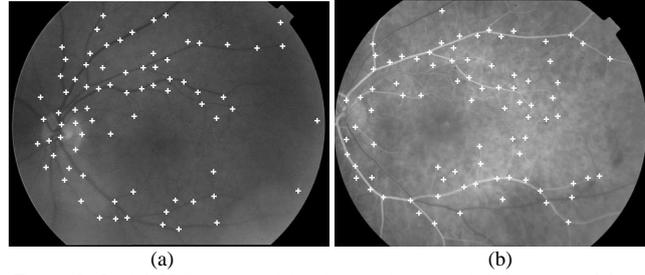


Figure 2. Initial Y-feature position in two images of different modalities. Only the best 100 features are selected.

3.2 Fitting Articulated Y-feature Model

We propose a Y-feature extraction method based on fitting an articulated model. The articulated model is fitted by maximizing the local intensities inside the template and gradient information on the boundary of the template.

3.2.1 Articulated Model for Y-features

The considered articulated model for the Y-features has 8 degrees of freedom (DOF), $\mathbf{x}=(x,y,\theta_1,\theta_2,\theta_3,w_1,w_2,w_3)$ which include the center position, three angles, and three widths for each branch. The length of each branch is fixed.

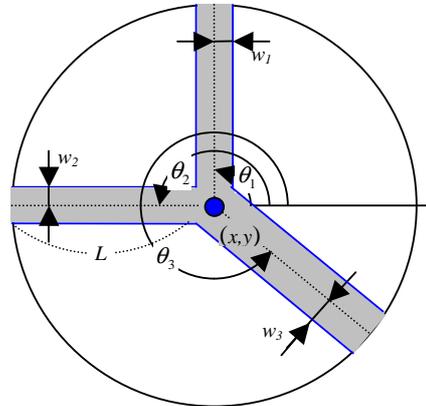


Figure 3. Articulated model of a Y-feature.

Using the geometric properties of the Y-features in the retinal images, we constrain the arms to be neither too close nor too far from each other. The half width of each arm, w_i ($i=1,2,3$), is the distance between the center of the vessel and its boundary. The width is also constrained to be between the minimum and maximal size of the vessels we are interested in detecting.

3.2.2 Initialization of Y-feature Shape

Given the seed points extracted by the PCA analysis of the response of the directional filters, the initial orientations of the three arms of the model need to be estimated. Can et al. (1999a) proposes the use of a rectangular grid and local detection of the maximum intensities for each vertical and horizontal line over the whole image. This approach, however, has several potential flaws: (1) a large number of grids must be dealt with, (2) Y-features located at the edges of the grid cannot be detected, and (3) the selection of local maxima from the four grid lines is combined in an ad-hoc manner into the three best vessels. Our method detects three bright or dark vessels from the circular boundary of the Y-feature located on the seed points. Along the circle, peaks or valleys of intensities are detected. Peaks represent bright vessels and valleys correspond to dark vessels. For each peak or valley, the best three arms are selected based on the following error function:

$$F_i(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^3 \int_{-w_i}^{w_i} \int_0^L [I(x_i, y_i)]^2 dl dw \quad (2)$$

where L is the length of the branch, w_i is the half width of the branch, i is the index of the considered arm, and $I(x_i, y_i)$ is the image intensity at $x_i = x + w \cdot \sin \theta_i + l \cdot \cos \theta_i$, $y_i = y + w \cdot \cos \theta_i + l \cdot \sin \theta_i$, where (x, y) is the center position of the model. $F_i(\mathbf{x})$ describes the sum of intensities inside the Y-feature model, represented by the shaded area in Figure 3. The angles of the Y-feature model are determined by connecting the center of the model and three vessels on the boundary. In the case of a dark vessel, three valleys that minimize $F_i(\mathbf{x})$ are selected, and, for a bright vessel, three peaks maximizing $F_i(\mathbf{x})$ are considered for the initial angles of the Y-feature model.

3.2.3 Fitting Y-features

After initializing the position of the Y-feature and the angle of the branches, we fit the articulated Y-features using gradient descent. In addition to the constraints defined by Equation (2), the gradient value along the boundaries of the vessels is utilized to enhance the accuracy of the estimated Y-feature model. We minimize the following energy:

$$F(\mathbf{x}) = (-1)^m \frac{1}{2} \sum_{i=1}^3 \int_{-w_i}^{w_i} \int_0^L [I(x_i, y_i)]^2 dl dw - \frac{1}{2} \sum_{i=1}^3 \sum_{w \in \{-w_i, w_i\}} \int_0^L [G(x_i, y_i)]^2 dl \quad (3)$$

where $G(x_i, y_i)$ is the gradient value on the boundary, and $m=0$ for a dark vessel, $m=1$ for a bright vessel. $G(x_i, y_i)$ is computed by the gradient of Gaussian filter output. When an articulated Y-feature is correctly fit, the sum of the intensity values, $F(\mathbf{x})$, in a Y-feature will be maximum for a bright vessel and the boundaries of the model are on the edge of the image. In addition, we constrain the angle and width of the branch to be within a specified range of values. The relative angles between the branches of the Y-model are defined by $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$, and $g_3(\boldsymbol{\theta})$ as follows:

$$\begin{aligned} b_l \leq g_1(\boldsymbol{\theta}) = \theta_2 - \theta_1 \leq b_u, \\ b_l \leq g_2(\boldsymbol{\theta}) = \theta_3 - \theta_2 \leq b_u, \text{ and} \\ b_l \leq g_3(\boldsymbol{\theta}) = \theta_1 - \theta_3 + 2\pi \leq b_u, \end{aligned}$$

where $\boldsymbol{\theta}=(\theta_1, \theta_2, \theta_3)$ and angle boundaries are set to $b_l = \pi/9, b_u = 10\pi/9$. These bounds were defined empirically after observing that no Y-feature goes beyond the specified boundary angles in the working set of retinal images for this study. Similarly, the half width of each branch is constrained by defining an upper and lower bound on the width:

$$\omega_l \leq w_i \leq \omega_b, \quad i = 1, 2, 3$$

In our experiments, half width boundaries are set to $\omega_l = 1, \omega_b = 4$. The length of the Y-feature model and the width of vessel boundaries are determined to be proportional to the size of the image. These constraints are enforced using Lagrange multiplier, thus solving the constrained optimization problem with the above inequalities (Bertsekas, 1999). We use a penalization approach using a barrier function. The inequality constraints on the angles $\boldsymbol{\theta}$ and the width of the arms $\mathbf{w}=(w_1, w_2, w_3)$ are translated to the penalty function:

$$B(\boldsymbol{\theta}, \mathbf{w}) = \sum_{j=1}^3 \left(\frac{1}{\alpha(g_j(\boldsymbol{\theta}) - b_l)(b_u - g_j(\boldsymbol{\theta}))} \right)^s + \sum_{j=1}^3 \left(\frac{1}{\beta(w_j - \omega_l)(\omega_b - w_j)} \right)^s \quad (4)$$

Figure 4 shows plots of the barrier function $B(x)$ with different values of s , which controls the smoothness of the barrier function. As s increases, the curve becomes steeper near the boundaries. In our

experiments we set $s = 1$. α and β are parameters to homogenize radian and pixel into metric quantities.

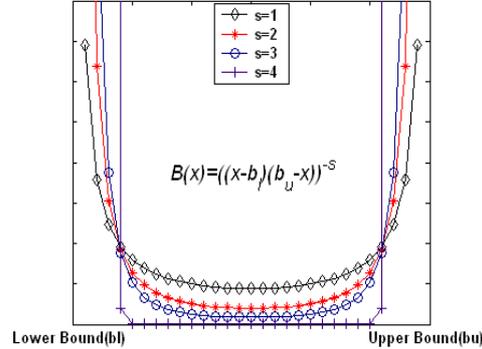


Figure 4. The barrier functions with lower and upper boundary w.r.t. to s . When the value x is closer to the lower or upper bounds, the function penalizes x by associating a large value.

The extraction of Y-features in the image consists of initializing the model using the feature point location and orientations and fitting the articulated Y-model to the image features by minimizing the function:

$$E(\mathbf{x}) = F(\mathbf{x}) + \lambda B(\theta, \mathbf{w}) \quad (5)$$

where λ is the trade-off between the goodness of fit to image features and the constraints on the orientation of the arms and their thickness. The function E is minimized iteratively using a gradient-based approach.

Fitting is performed based on Equation (5). Since the articulated Y-feature has 8 degrees of freedom, if we update all 8 parameters at once, the gradient algorithm could be trapped in a local minimum. To prevent this, we fix the three endpoints of the branch and update only the position and the width of the arms. Once the center point of the Y-feature converges to a stationary location, the angles and widths of the branches are updated. The gradient-based minimization scheme is iterated until it reaches a maximum number of iterations, set to 25, or the parameters reach stationary values. Figure 5 shows the fitting process. As the PCA analysis detects both types of vessel (i.e. bright/dark), the model is fitted twice to detect Y-features, first assuming dark vessels and second assuming bright vessels. After alignment, the Y-feature candidate with the higher gradient value on the boundary is selected.

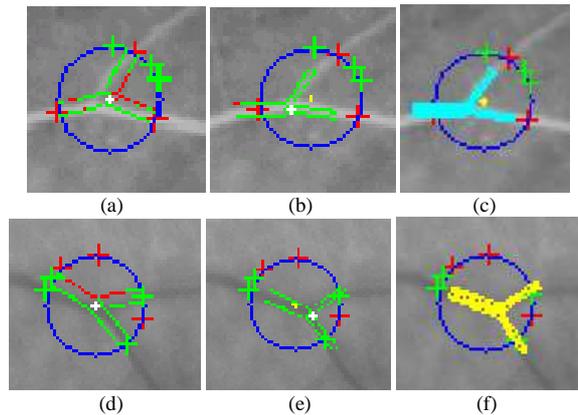


Figure 5. Fitting a Y-feature model to the image data. The yellow and white points correspond respectively to the seed point, and the estimated center point. The green and red crosses indicate respectively local valleys, and peaks in the grey level distribution. (a)(d) Initial Y-feature model of bright or dark vessel. (b)(e) Y-feature model after fitting. (c)(f) Detected Y-features. Every valid Y-feature is classified into bright or dark vessels.

All fitted Y-features do not necessarily correspond to real bifurcations of vessels. Indeed, many initial Y-features are selected based on whether they have a high response to the directional filters, as described in Section 3.1.2). A large number of seed points are considered in order not to miss any real Y-feature in the image. After fitting the articulated model to the selected image regions, we discard some of the extracted Y-feature points where the boundary of the arms of the articulated model does not lie on strong edge points. If 2/3 of the boundary of any arm is not on the edge, the Y-feature model is discarded.

3.3 Matching and Global Registration

The circulation of the fluorescein dye in the retinal vessels induces different grey level properties in the color and angiogram images, making the matching of extracted Y-features challenging. We propose to match extracted Y-features across modalities and through different phases of the circulation using the local maximization of the *mutual information* (Wells et al., 1996; Viola and Wells, 1997). To match individual Y-features, geometric descriptions are too coarse, and intensity matches are unreliable due to the intensity changes. Instead, we do match intensities, but use *mutual information* rather than correlation. A RANSAC method (Fischler and Bolles, 1981) is used for pairwise registration of images. This, however, does not provide robust registration for the entire image sequence, and a global registration is required. We propose a method of achieving a robust global registration of color and fluorescein frames

using the all pairs' shortest path algorithm.

3.3.1 Matching Y-features using Mutual Information

We consider a rectangular window enclosing the Y-feature model in the source image and compare it to other windows in the target image within a search area, which is set to a fifth of the image size. Y-features for which the *mutual information* is maximal among the set of candidates are paired. *Mutual information* for a pair of rectangular windows \mathbf{z}_A in the source image and \mathbf{z}_B in the target image is defined by Viola and Wells (1997):

$$MI(\mathbf{z}_A, \mathbf{z}_B) = H(\mathbf{z}_A) + H(\mathbf{z}_B) - H(\mathbf{z}_A, \mathbf{z}_B) \quad (6)$$

where $H(\mathbf{z}) = -\int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z}$ is Shannon's entropy of the image window \mathbf{z} , and p is the distribution of the grey levels in the considered window. We approximate the entropy as:

$$H(\mathbf{z}) \approx -\frac{1}{N_z} \sum_{z_i \in \mathbf{z}} \ln p(z_i)$$

where N_z is the size of the window \mathbf{z} , and the density function $p(\mathbf{z})$ is estimated based on Parzen window density estimation. We consider a Gaussian density function for the Parzen window W_p , and the distribution of the grey levels is locally approximated as follows:

$$p(z) \approx \frac{1}{N_p} \sum_{z_j \in W_p} g_\psi(z - z_j)$$

where N_p is the number of samples in the Parzen window W_p , and $g_\psi(z)$ is the uni- or bi-variate Gaussian density function with diagonal covariance matrix ψ (Wells et al., 1996). The entropy function is then rewritten as:

$$H(\mathbf{z}) \approx -\frac{1}{N_z} \sum_{z_i \in \mathbf{z}} \ln \frac{1}{N_p} \sum_{z_j \in W_p} g_\psi(z_i - z_j) \quad (7)$$

3.3.2 Global Registration

Using matched pairs of Y-features, images are registered using an affine transform. The RANSAC method is used to find the inliers among the matched features. The best three matching pairs are selected to obtain the best affine transform, which minimizes the geometric errors of every match:

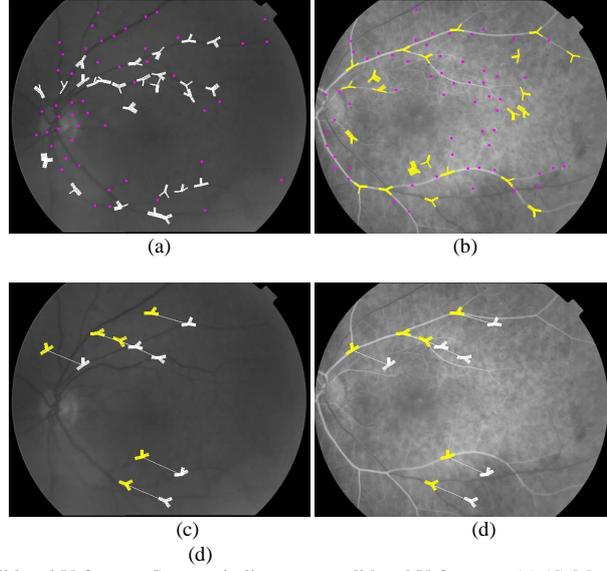


Figure 6. (a) (b) Fitted and validated Y-feature. Squares indicate non-validated Y-features. (c),(d) Matching pairs of Y-features across modalities

$$\mathbf{A}_{best} = \arg \min_{\mathbf{A}} E_R(\mathbf{A}) = \arg \min_{\mathbf{A}} \frac{1}{2} \sum_i \left\{ (\mathbf{x}_i - \mathbf{A}\mathbf{y}_i)^2 + (\mathbf{y}_i - \mathbf{A}^{-1}\mathbf{x}_i)^2 \right\} \quad (8)$$

where \mathbf{x}_i and \mathbf{y}_i are a matching pair and \mathbf{A} is affine transform obtained from selected 3 pairs of Y-features. With the selected affine transform, the geometric error for each matching pair is computed. Based on the error, the outliers of the matching pairs are removed, and the inliers are considered for estimating the affine transform.

Figure 6-(a) and (b) displays the extracted Y-features. The small squares in the image indicate locations selected as an initial Y-feature but later rejected due to low image gradients along the edges of the Y-feature arms. Figure 6-(c) and (d) show the corresponding inlier pairs after matching and registration.

Image sequences of angiograms can be registered sequentially using pairwise registration. However, if one image is not correctly registered, this error propagates to the remaining images. Furthermore, registration across modalities varies in accuracy according to the phase of circulation of the fluorescein dye in the retinal vessels. Global registration is introduced to automatically reduce the errors of registration of an unordered set of images within and across modalities. Global registration is intended to identify the best registration among all pairs of images, while minimizing the global registration error.

To address this problem, we construct a complete and undirected graph, where the nodes correspond to the images to be registered and the edges correspond to pairwise registration of the images. A cost is associated to each edge representing registration error computed by Equation (8) obtained by the pairwise registration described above. The global registration problem is formulated as finding the shortest path from every node to all the other nodes in this complete and undirected graph. The all-pairs' shortest paths are calculated using the Floyd-Warshall algorithm (Floyd-Warshall, 1962). With this algorithm, the shortest paths from every node to all other nodes are obtained. A node with the lowest sum of costs is selected as a reference frame. And the shortest paths from the reference frame to all other frames are simply decided by the all pairs' shortest path algorithm (Choe et al. 2006a). Figure 7 illustrates the difference between pairwise and global registration.

3.4 Experimental Results

We conducted experiments on our registration method on 11 sets of retinal images. The 11 image sets were chosen by an experienced retinal specialist (Dr. Alexander Walsh) to be representative of the typical range of images encountered in clinical practice, encompassing a range of common retinal diseases. Each set has one color image and approximately 20 angiographic images, including very dark ones with poor contrast. In most of the previous work dealing with retinal angiogram registration, these images are simply discarded and not registered. Furthermore, in some pathological cases, irregular spots with very strong contrast bias the extraction of Y-features.

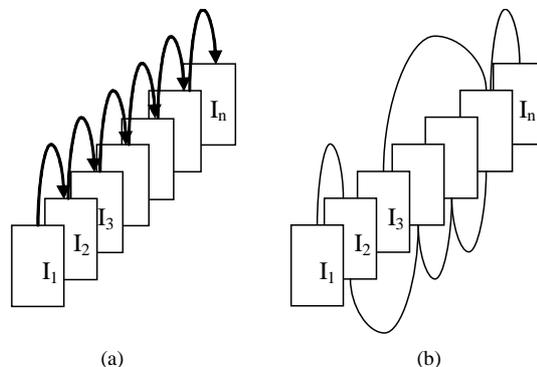


Figure 7. Global registration of frames across modalities (a) Pairwise registration of consecutive images (b) Global registration constructed from the proposed method.

Figure 8 shows the pairwise registration of two images using an affine transform. Global registration was derived using the all pairs' shortest path algorithm described above. One example of the global registration tree is shown in Figure 10. The node is each picture and the weight of the edge is the geometric error. Figure 9 shows the final mosaic obtained after registering a set of images. The images are α -blended images, where $\alpha = 1 / \#$ of images.

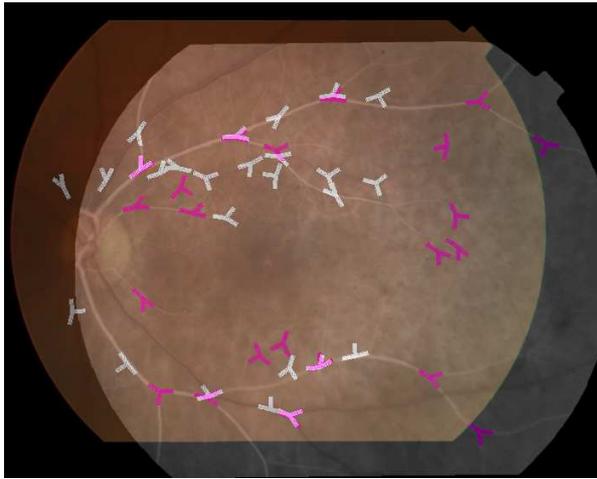


Figure 8. A view of registered color and dark images



Figure 9. Mosaic image of registered angiograms

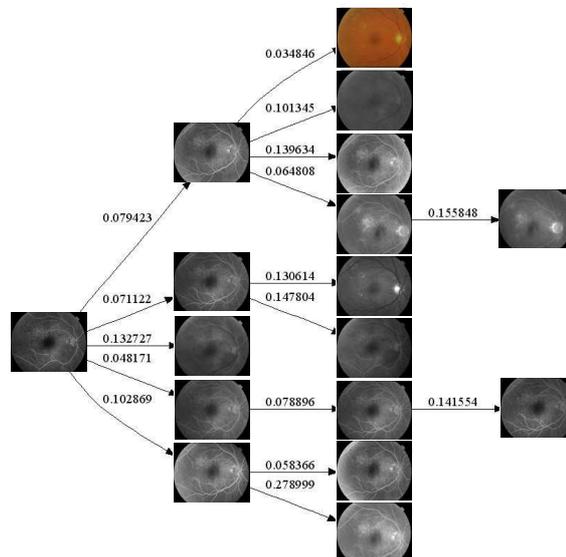


Figure 10. An example of global registration from the selected reference frame to all other frames using the shortest path algorithm. Nodes of images are connected by arcs with average pixel error between the two images.

We have conducted an evaluation to quantify the accuracy of the detection of the Y-feature using the

proposed articulated model. We have measured the number of detected Y-features, as a variable of the total number of seeds points N_s detected by the PCA-based algorithm. The obtained ROC curve is displayed in Figure 11. The ROC curve is saturated at a false detection rate of approximately 0.17 with a positive detection rate of more than 0.96. Ground truth was provided by hand-tagging the location of the Y-features in the image set.

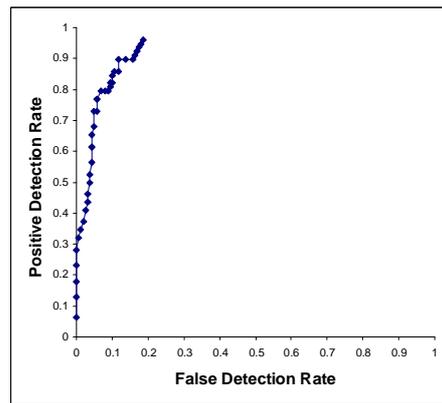


Figure 11. ROC curve of Y-feature detection according to the number of seed points, N_s

Evaluating the accuracy of the registration across modalities is a difficult task because of lack of ground truth. We defined the ground truth by having Dr. Alexander Walsh, an experienced retinal specialist, annotate corresponding points manually, and computed the affine transform matrix based on these points.

TABLE 1
COMPARISON OF AVERAGE GEOMETRIC PIXEL ERRORS USING: PAIRWISE REGISTRATION OF CORNERS, PAIRWISE REGISTRATION OF EXTRACTED Y-FEATURES, AND GLOBAL REGISTRATION USING Y-FEATURES.

	PAIRWISE REGISTRATION WITH PCA-BASED FEATURES	PAIRWISE REGISTRATION WITH Y-FEATURES	GLOBAL REGISTRATION WITH Y-FEATURES
Total Error	4.335719	3.792287	2.757204

We have also compared the proposed approach based on the matching of Y-features to other different methods that use other image features. In Table 1, we show the registration error obtained by: PCA-based features, pairwise registration of Y-features fitted to image characteristics, and using the global registration method. 91.09% of images (235 out of 258 images) were accepted for the registration, and 8.91% of images (23 out of 258 images) were automatically discarded from the registration since they

had high registration errors. We notice an improvement of the accuracy by using the proposed all pairs' shortest path algorithm. The geometric error of the matching points was calculated from the manually estimated ground truth affine transform. The average pixel error of the proposed method is 2.757204. This value was obtained by averaging over the whole set of images considered in the 11 sets of images (224 pairs). However, this evaluation of 2-D registration using manual annotation has limitations. First, manual annotation cannot be perfect, since annotation is performed by a human. Second, the actual shape of the retina is a 3-D surface, but we registered the images by 2-D affine transformation. Also, it should be noted that the distribution of Y-features plays a role in registration. If Y-features are absent from a large portion of the images, registration may not be accurate. These sources of errors contribute to the results reported in Table 1. We address these limitations by reconstructing the 3-D shape of a retina.

4 3-D RECONSTRUCTION

In this section, we propose an automatic method for 3-D reconstruction from a pair of retinal images. The method consists of three steps: First, the matched Y-features are used to estimate the epipolar geometry based on a plane-and-parallax approach. Second, the images are rectified by the estimated fundamental matrix; after rectification, the search space on the scanline for stereo matching is estimated based on the Y-feature correspondences. Subsequently, a dense disparity map is estimated using *mutual information*. Figure 12 shows examples of retinal image pairs used in this study.

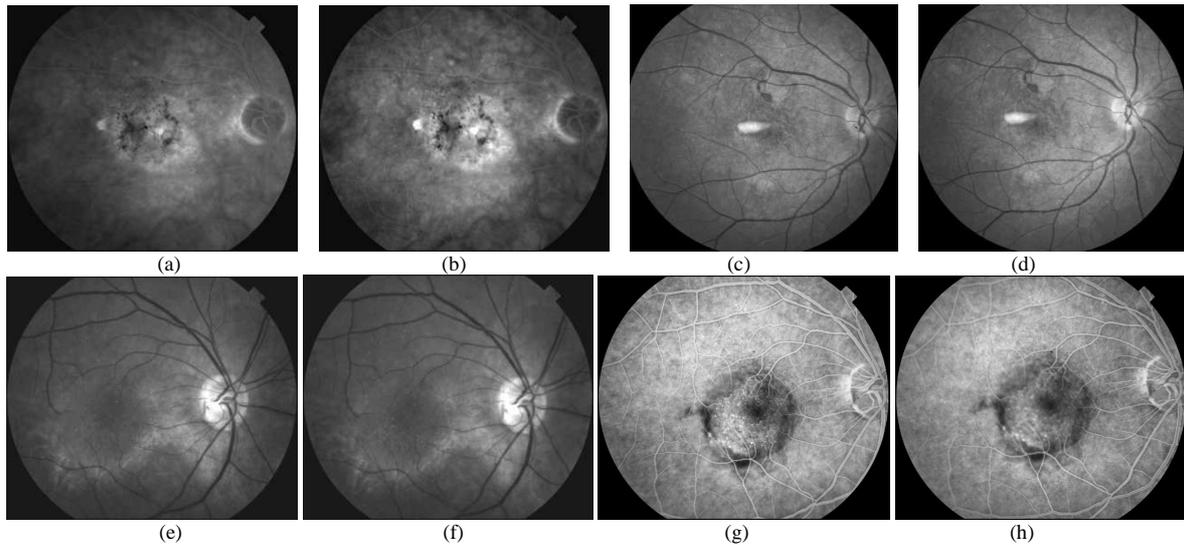


Figure 12. Various pairs of retinal images (a)(b) Images of Retina_A with strong illumination changes. (c)(d) Images of Retina_B (e)(f) and (g)(h) Two pairs of stereo retinal images of Retina_C from a patient with age-related macular degeneration (e)(f) A characteristic ‘blister’ in the macula is harder to see in the red-free images (e)(f) than it is in the images taken after injection of fluorescein dye (g)(h) into the antecubital vein. The image in (h) is blurry which complicates stereoscopic measurements.

4.1 Estimation of Epipolar Geometry

The pair of images with the lowest registration error between them is selected for 3-D shape reconstruction among a sequence of images. Among all pairs, this selected pair is likely to have more Y-features and a similar density of Y-features. The homography and the fundamental matrix are estimated from a set of Y-feature correspondences between two images. This section describes the method for estimating the epipolar geometry from a pair of fluorescein images of the retinal fundus.

Using the matched pairs of Y-features, images are registered by a homography. The RANSAC method is used to detect the inliers among the matched features. The best four corresponding pairs of Y-features are selected to estimate the homography, which minimizes the geometric error:

$$\mathbf{H}_{best} = \arg \min_{\mathbf{H}} \left[\frac{1}{2} \sum_i \left\{ (\mathbf{x}_i - \mathbf{H}\mathbf{y}_i)^2 + (\mathbf{y}_i - \mathbf{H}^{-1}\mathbf{x}_i)^2 \right\} \right] \quad (9)$$

where \mathbf{x}_i and \mathbf{y}_i are 3×1 vectors of the matching pair and \mathbf{H} is the 2-D homography.

At least seven point correspondences are necessary to estimate the fundamental matrix of the epipolar geometry. However, various standard implementations of the 7-points and 8-points algorithm were tested and did not provide satisfactory results (Hartley, 1995; Kumar et al., 1994). Figure 13 shows the result of

the corresponding epipolar lines defined by the fundamental matrix estimated by the 8-point algorithm. In Figure 13, we show examples of erroneous estimation of the epipolar geometry from a set of correctly matched Y features.

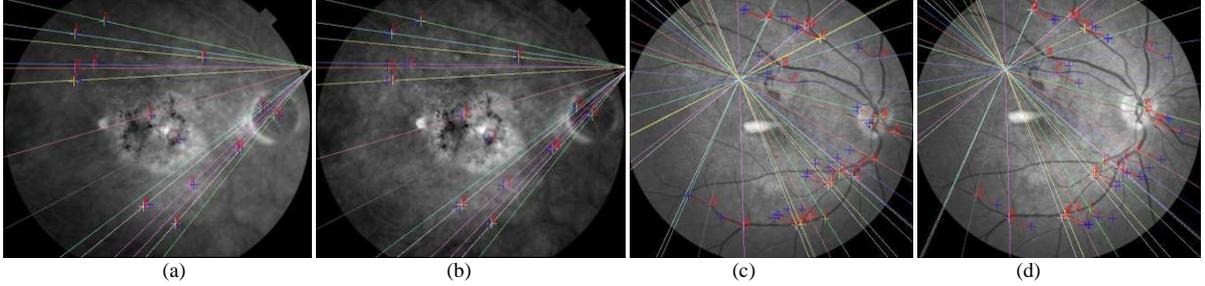


Figure 13. Examples of estimated epipolar geometry using an 8-point algorithm in the case of a translation of the camera. The estimated epipolar geometry is inaccurate. (a)(b) Retina_A (c)(d) Retina_B

In the examples above, the motion of the camera between two acquisitions is close to a pure translation. There are two common situations where a degenerate case occurs in the estimation of the fundamental matrix (Hartley and Zisserman, 2000): Cases in which both camera centers and the 3-D points are on a ruled quadric, or when all the points lie on a plane. In the case of retinal images, the later case is common because the surface of the retina is relatively flat (except in a few patients with severe retinal disease). Although the points are not on the same plane, the range of 3-D depth is too shallow to get a satisfactory fundamental matrix from the 8-point algorithm.

To overcome this limitation of the 8-point algorithm on retinal images, we use the plane-and-parallax algorithm proposed by Kumar et al. (1994). Given 4 corresponding points, first a homography is calculated. Adding 2 more point correspondences belonging to residual parallax regions enables us to estimate the location of the epipoles. The fundamental matrix can then be estimated by:

$$\mathbf{F} = [\mathbf{e}']_{\times} \mathbf{H} \quad (10)$$

where \mathbf{H} is the homography obtained from the Equation (9) and the epipole \mathbf{e}' is the intersection of $(\mathbf{H}\mathbf{x}_1) \times \mathbf{y}_1$ and $(\mathbf{H}\mathbf{x}_2) \times \mathbf{y}_2$ where \mathbf{x}_i and \mathbf{y}_i ($i=1,2$) are a matching pair of features, and $[\cdot]_{\times}$ is a matrix notation for the calculation of the cross product. Figure 14 illustrates the plane-and-parallax algorithm to estimate the fundamental matrix.

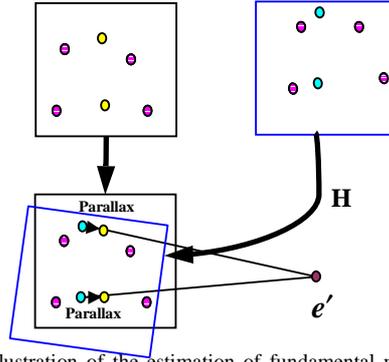


Figure 14. Illustration of the estimation of fundamental matrix using the plane-and-parallax algorithm. Four corresponding points define the projective geometry, H . Two additional points in the residual parallax regions define the epipole e' , which then provides the estimate of the fundamental matrix $F=[e' \uparrow, H]$

We have implemented a RANSAC-based algorithm for the plane-and-parallax method, similar to the one used for the estimation of the homography. For each estimated homography, we randomly select two additional features with large geometric errors reported by Equation (9) to form a fundamental matrix. Subsequently, we select the fundamental matrix that satisfies the following equation.

$$\mathbf{F}_{best} = \arg \min_{\mathbf{F}} \left\{ \frac{1}{2} \sum_i (\mathbf{x}_i \mathbf{F} \mathbf{y}_i)^2 \right\} \quad (11)$$

where \mathbf{x}_i and \mathbf{y}_i is the corresponding pair. Figure 15 shows the corresponding epipolar lines obtained by the plane-and-parallax approach. The plane-and-parallax algorithm can reconstruct not only flat-like surfaces but also surfaces with greater 3-D variations. After estimating the fundamental matrix, stereo images are rectified using Gluckman and Nayar's method (2001) for the estimation of the depth map.

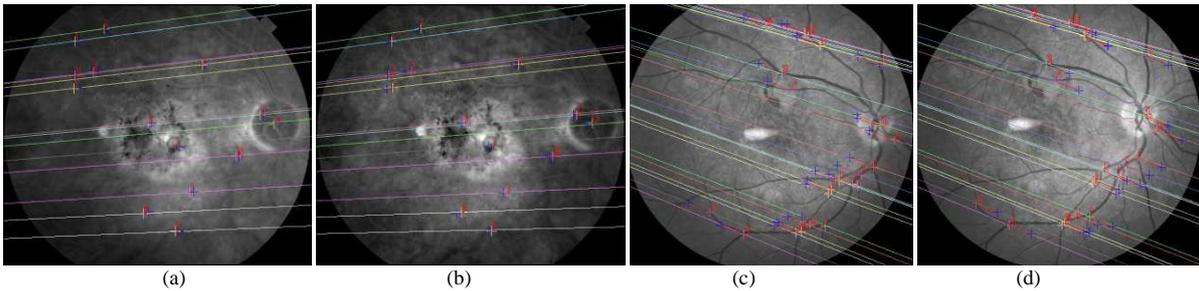


Figure 15. The epipolar lines obtained from the plane-and-parallax based fundamental matrix estimation.

4.2 Stereo Matching and 3-D Surface Estimation

4.2.1 Estimating the Search Space

In this section, we address the problem of estimation of depth range from the set of matched Y-features.

The definition of a suitable search space within the scanlines for stereo matching is important: when the search space is too narrow, the corresponding points cannot be identified, on the other hand, when the search space is too wide, it may be possible to detect the corresponding pixels in the scanline, but the detection is not necessarily accurate. A proper range for the search space gives a high confidence in matching; and it also reduces the computation time. Most of the existing algorithms manually select the search space to get the best results. One approach for automatically estimating the search space is based on using a coarse-to-fine search (Magarely and Dick, 1998). This method assesses the approximate offset of the search space from the coarser level, and does not provide an accurate estimate. In addition, error from wrong matches in the coarser level can be propagated to the finer level. Moallem and Faez (2001) define the search space based on the maximum of the disparity gradient value. This requires scenes or surfaces with strong disparity in depth, and therefore may not be applicable to many retinal images. We propose a method that automatically defines the search range by using the matched Y-features.

Although some cases of severe retinal disease may have “dramatic” distortion of the retinal morphology with focal elevations of the retina several hundreds of microns above the expected retinal surface, for many patients the retina has a relatively small 3-D depth variation. As such, the disparity search space is fairly shallow. Since disparities for the sparse key features are known, disparity at other points can be easily extrapolated because the retinal surface is an almost planar surface. Assuming that corresponding Y-features are well distributed across the images, the search space can be easily estimated from the set of known disparities. Figure 16-(a)(b)(c) shows the histogram of the dense disparity map. Figure 16-(d)(e)(f) shows the histogram of Y-features’ disparities. We note that the shape of the distribution can be well approximated by a Gaussian distribution. Even though the number of sample points defined by the Y-features is much smaller, the histogram follows the same pattern as the one estimated on the dense disparity map. The average and the variance of the Y-features’ disparity are also similar to those of the disparity map.

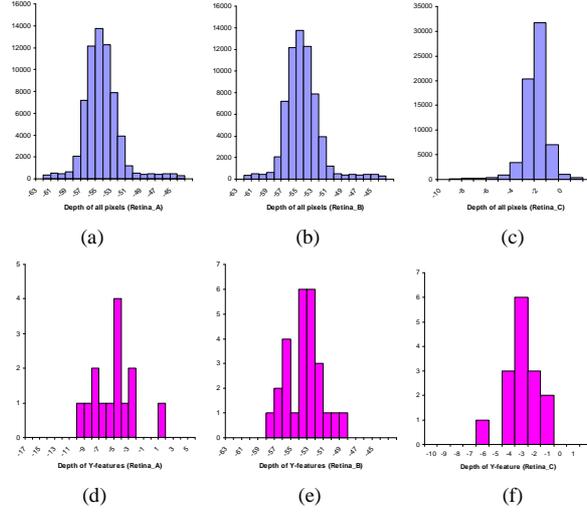


Figure 16. (a)(b)(c) Histograms of all pixels' disparity for each stereo pair. (d)(e)(f) Histograms of Y-features' disparity. The mean(μ) and the standard deviation(σ) of each histogram is (a) $\mu=-6.1, \sigma=2.221$ (d) $\mu_Y=-6.0, \sigma_Y=2.801$ (b) $\mu=-55.1, \sigma=2.459$ (e) $\mu_Y=-54.3, \sigma_Y=2.121$ (c) $\mu=-2.8, \sigma=0.983$ (f) $\mu_Y=-3.5, \sigma_Y=1.336$

After rectifying the position of matching Y-features, we can obtain the mean(μ_Y) and the standard deviation(σ_Y) of disparities among all transformed Y-feature correspondences. The search space S is then determined by the following inequality:

$$\mu_Y - 4\sigma_Y < S < \mu_Y + 4\sigma_Y \quad (12)$$

which has 99.994% confidence interval in the Gaussian distribution.

4.2.2 Dense Disparity Map using Mutual Information

After rectification, all matching pixels are located on the same scanlines. On the scanline, the lower and the upper bound of the search space are estimated by the method described in the previous section. This method produces a 1-D shallow search space which enables a more accurate computation of the dense disparity map. In case of the retinal fundus where the intensities of the matching areas vary across the stereo images, a general cross-correlation based algorithm does not provide satisfactory results. Instead, we propose using *mutual information* for matching points along the scanlines and estimating the depth map.

Kim et al. (2003) used only joint entropy $H(\mathbf{z}_A, \mathbf{z}_B)$ rather than *mutual information*. The proposed method does not work well on low-textured areas. When mostly textureless areas are compared to each

other, joint entropy has a high value on textureless area, which is incorrect. However, marginal entropies $H(\mathbf{z}_A)$ and $H(\mathbf{z}_B)$ help to boost the *mutual information* value on the textured structures. Hirschmüller (2005) estimates the discretized density function $p(\mathbf{z})$ in order to reduce the time complexity. $p(\mathbf{z})$ is calculated at the given point and it is convolved with a 2-D Gaussian to simulate the density function. However this simplification reduces the accuracy of *mutual information*.

We utilize Equation (6) to implement *mutual information*. In *mutual information*, the marginal entropies $H(\mathbf{z}_A)$ and $H(\mathbf{z}_B)$ are included and the density function $p(\mathbf{z})$ is computed individually using a Parzen window estimation with Gaussian assumption. The disparity of each pixel is determined by following equation.

$$disparity = \arg \max_{d \in S} MI(d) = \arg \max_{d \in S} MI(\mathbf{z}_A, \mathbf{z}_{B,d}) \quad (13)$$

where $MI(\mathbf{z}_A, \mathbf{z}_{B,d})$ is *mutual information* from Equation (6) and $\mathbf{z}_{B,d}$ is the window of d pixel distance from the window \mathbf{z}_B in the second image. Each pixel is centered in a window \mathbf{z}_A , which is then compared with a window $\mathbf{z}_{B,d}$ in the other image within the search space S . The pair of windows that has maximum *mutual information* determines the disparity of each pixel position.

Since the range of disparity values is shallow, the subpixel resolution of disparity is essential in 3-D reconstruction of retinal images to avoid “staircasing” in the disparity map. We estimate the disparity map with subpixel resolution using a quadric interpolation relying on neighboring *mutual information* values as follows:

$$Subpixel = disparity + \frac{1}{2} \frac{c - a}{b - \min(a, c)} \quad (14)$$

where $a=MI(disparity-1)$, $b=MI(disparity)$, $c=MI(disparity+1)$, and *disparity* is the value of selected disparity in Equation (13). Every disparity is determined individually by the *mutual information* matching criterion. We did not apply any smoothness constraint to calculate disparities. Based on the taxonomy of stereo algorithms (Scharstein and Szeliski, 2002), only the matching function is applied. Neither aggregation nor optimization methods are used. We have implemented a number of methods to

perform this task (Kolmogorov and Zabih, 2001; Scharstein and Szeliski, 2002; Webpage of Middlebury source code). In our experiments we have observed that such methods only degrade the accuracy of the matching.

4.3 Experimental Results

The reconstructed 3-D shape is evaluated by comparing our results to aligned OCT data from a commercial instrument (Carl Zeiss Meditec, Dublin, CA). Our data differs slightly from OCT data since we have a disparity map – not a true depth map. A more accurate depth map would require knowledge of the internal camera parameters, which are currently unavailable.

4.3.1 3-D shape of the retinal fundus

We conducted experiments on more than 20 sets of stereo image and 5 challenging pairs of fluorescein images as selected by an experienced retina specialist (4 pairs are shown in Figure 12). Retina_A has a large elevated lesion in the center of the image. The stereo images of Retina_A have a very strong illumination change. Retina_B is the concave shape with a very small lesion in the upper middle part of the image. Three different stereo pairs of Retina_C are tested, and two sets are shown in Figure 12-(e)(f) and (g)(h). The images demonstrate a retinal pigment epithelial detachment (‘blister’) in the macula secondary to choroidal neovascularization from age-related macular degeneration. The ‘blister’ in the Figure 12-(h) is out of focus, which makes the image reconstruction difficult. After the Y-features in each image are extracted and matched, the epipolar geometry is estimated using the plane-and-parallax algorithm (Figure 14). The stereo images are rectified based on the fundamental matrix, and the search space is estimated from the set of matching Y-features (Figure 16 and Equation (12)). Using *mutual information*, subpixel resolution dense disparity maps are estimated and shown in Figure 17.

In Figure 17-(a)(d), (b)(e) and (c)(f), 3-D shape of Retina_A, Retina_B, and Retina_C are shown from different views. In these images, the disparity values have been scaled (by a factor of 30) to magnify the 3-D depth. The estimated 3-D depth maps have little noise even though we did not apply any post-matching smoothing step. The calculated depth values near the boundary are incorrect due to occlusions.

The size and degree of the elevated lesion area of Retina_A is demonstrated in Figure 17-(a)(d). Figure 17-(b)(e) shows the concave shape of the fundus in this myopic patient's eye. In Figure 17-(c)(f), 3-D shape of a blister-like elevation of the retinal surface is accurately estimated.

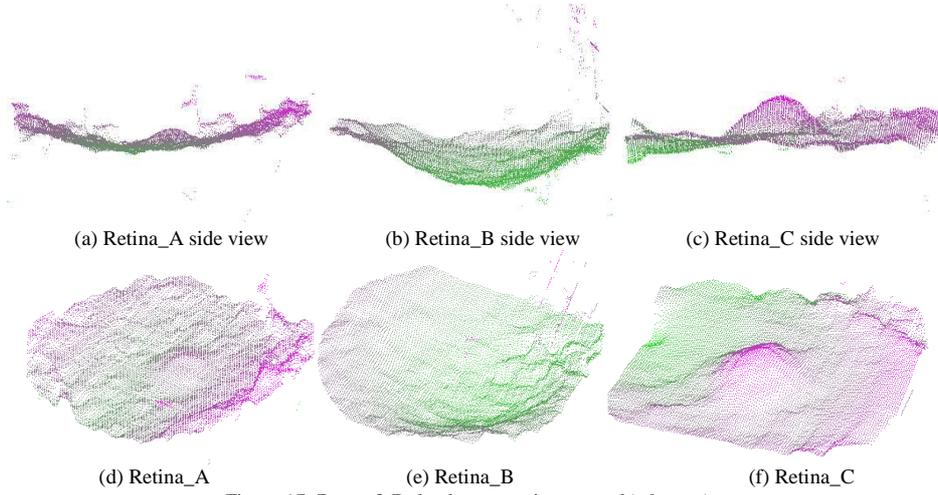


Figure 17. Dense 3-D depth maps using *mutual information*

4.3.2 Evaluation of the stereo matching

The evaluation of 3-D surfaces is a difficult process in medical imaging because of the difficulty in accessing ground truth measurements. One possible source of ground truth data is obtained in this study from OCT. However, because OCT data and fundus images are acquired by different machines, it is not easy to align the data to the 3-D shape data. Therefore, for evaluation, we matched several feature positions, which have high curvature values, from the OCT to the 3-D depth map to align them to obtain a 3-D transformation matrix T . Differences between the 3-D point in the transformed OCT data, $T(X)$, and its nearest neighbor point in the 3-D depth map, Y , are calculated as an evaluation criterion.

$$E(X, Y, T) = \frac{1}{N_X} \sum_{\mathbf{x}_i \in X} \left\| T(\mathbf{x}_i) - \arg \min_{\mathbf{y}_j \in Y} \|T(\mathbf{x}_i) - \mathbf{y}_j\| \right\| \quad (15)$$

where N_X is the number of the points in OCT data.

In the first experiment, we evaluate the performance of *mutual information* according to the window sizes. There are two windows in *mutual information*. One is \mathbf{z} in Equation (7), which is the area window enclosing the current pixel position. The other is W_p , which is the Parzen window to estimate the density

function. N_z and N_p are the sizes of each window. First, we fixed the $N_p = 3$ and measured the pixel errors, changing the area window size, N_z . Second, we fixed the area window size to $N_z = 14$ and monitored the errors, changing the Parzen window size, N_p . Three sets of stereo images of Retina_C, which has OCT ground truth, are tested for this evaluation. The error measures for these two scenarios are plotted in Figure 18-(a) and (b).

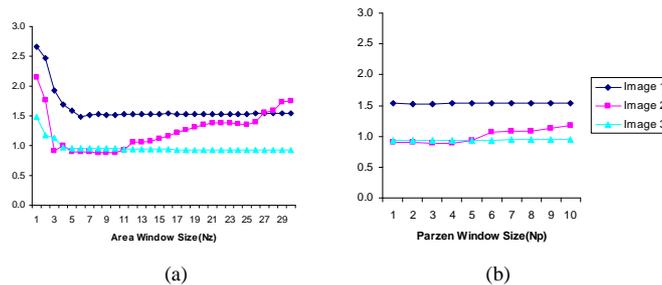


Figure 18. (a) Errors with varying area window size, N_z , of *mutual information*. (b) Errors with varying Parzen window size, N_p .

As shown in Figure 18-(a), as the area window size increases, the errors first decrease, then stabilize, and increase. The graph of Parzen window size also follows a similar pattern as shown in Figure 18-(b). Since *mutual information* is a statistical measure, more data would give better accuracy. However, when the window size is too large, the saliency of *mutual information* decreases. In addition, when a larger window is used, the computation is slower.

In the second experiment, we evaluated the performance of different stereo matching algorithms using OCT ground truth. We have used the evaluation codes from the Middlebury stereo database and source code (<http://www.middlebury.edu/stereo>) for SSD, DP (Dynamic Programming), SO (Scanline Optimization), and GC (Graph Cut) algorithms. For SSD, not only sum of squared differences, but AD (absolute differences) were also compared with different parameters, such as window size, shiftable window, and truncation. For global optimization algorithms, such as DP, SO, and GC, we compared the performances with two matching functions (SSD and AD) and with different smoothness terms, and the best parameters for each image were selected for each method. We implemented NCC (Normalized Cross Correlation) and *mutual information*. As shown in Table 2, *mutual information* outperforms all other

methods in terms of accuracy. *Mutual information* generated a low amount of noise and was highly accurate, even in areas with low texture.

TABLE 2
DISPARITY ERRORS BETWEEN OCT DATA AND 3-D SHAPE IN PIXEL METRIC FOR EACH STEREO ALGORITHM

STEREO MATCHING ALGORITHMS	IMAGE 1	IMAGE 2	IMAGE 3
SUM OF SQUARED DISTANCES (SSD)	0.282563	0.510051	1.352426
DYNAMIC PROGRAMMING (DP)	0.285503	1.142079	0.673890
SCANLINE OPTIMIZATION(SO)	0.332272	0.626138	1.777096
GRAPH CUT (GC)	0.290186	0.515139	1.708525
NORMALIZED CROSS CORRELATION(NCC)	0.159760	0.437529	0.258479
MUTUAL INFORMATION(MI)	0.152694	0.328937	0.245442

THE PARAMETERS ARE FINE-TUNED TO OBTAIN THE BEST PERFORMANCE. IN THIS TABLE, THE BEST RESULTS FOR EACH METHOD ARE SHOWN.

In Figure 19-(a)(b)(c)(d), the surfaces of the 3-D retinal fundus are shown. We reduced the boundary area by a small amount to remove outliers caused by occlusions. In Figure 19-(e)(f)(g)(h), the image texture is mapped on top of the 3-D depth map. The line, which matches with the intersection position of the surface images on the left, displays the 3-D depth of the texture. Except the boundary area, most of areas are well reconstructed. The 3-D shape of lesions and an optic disc are also accurately estimated.

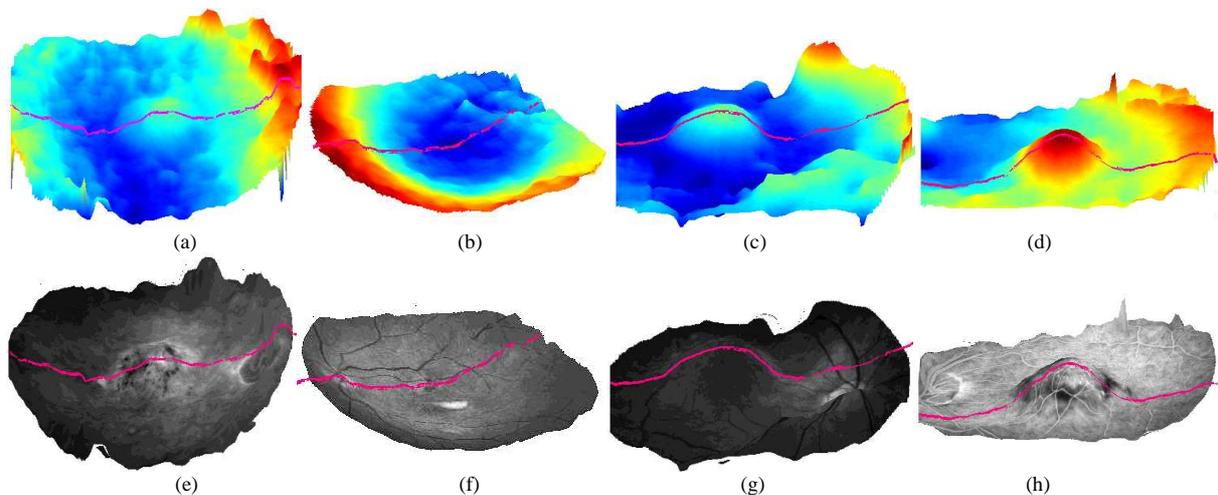


Figure 19. Reconstructed 3-D shape of the retinal fundus (a)(b)(c)(d) Surface maps. (e)(f)(g)(h) Texture maps. Each line matches with the line of the surface map on the left.

5 CONCLUSION

We have described methods for 2-D registration and 3-D shape reconstruction of the retinal fundus. Extracting Y-features with an articulated model provides robust, accurate, and fully automatic

registration of color and fluorescein retinal images. The PCA analysis of the directional filter responses generates good estimates of the initial position of Y-features. The fitting of the articulated Y-feature model provides accurate estimates of the Y-feature positions in the image and removes false positives. Our experiments show that the proposed global registration has good performance in registering color images and fluorescein angiograms of the retina, even in difficult cases where the images contrast is poor.

For 3-D shape reconstruction, we proposed a method that integrates reliable landmarks, defined by Y-features, a plane-and-parallax approach for robust epipolar geometry estimation, and the *mutual information* criteria to estimate the dense stereo map. The performance of 3-D reconstruction is validated by OCT data. The 3-D reconstruction is robust even though the retinal surface is relative flat.

All processes were completed without any user interaction, and they were extensively tested and evaluated on a number of retinal images. By registering the retina images and reconstructing the 3-D shape of the retinal fundus, the proposed method can provide a useful tool for ophthalmologists in the diagnosis of retinal disease. In future work, we intend to focus on the 3-D Euclidean reconstruction of a retinal fundus and registration of images using the reconstructed 3-D structure.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (NIH) under grant No. R21 EY015914-01 and the Doheny Eye Institute.

REFERENCES

- Bartoli A., Sturm P., 2004. Nonlinear Estimation of the Fundamental Matrix with Minimal Parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 25, No. 3, 426-432.
- Bemmel C.M.V., Spreuwers, L.J., Viergever, M.A., Niessen, W.J., 2003, Level-Set Based Artery-Vein Separation in Blood Pool Agent CE-MR Angiograms, *IEEE Transactions on Medical Imaging*, vol. 22, pp. 1224-1234.
- Bertsekas, D.P., 1999. *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific.
- Can, A., Shen, H., Turner, J.N., Tanenbaum, H.L., Roysam, B., 1999. Rapid automated tracing and feature extraction from live high-resolution retinal fundus images using direct exploratory algorithms. *IEEE Transaction on Information Technology for Biomedicine* 3:2, 125-138.
- Can, A., Stewart C.V., Roysam, B., 1999. Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 286-292.
- Can, A., Stewart C.V., Roysam, B., Tanenbaum, H.L., 2002. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on PAMI* 24, 347-364.
- Choe, T.E., Cohen, I., 2005. Registration of Multimodal Fluorescein Images Sequence of the Retina. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV) 2005*, 106-113.
- Choe, T.E., Cohen, I., Lee, M., Medioni G., 2006. Optimal Global Mosaic Generation from Retinal Images. In: *Proceedings of International Conference on Pattern Recognition (ICPR 2006)*, 681-684.
- Choe, T.E., Cohen, I., Medioni G., 2006. 3-D Shape Reconstruction of Retinal Fundus. In: *Proceedings of CVPR 2006*, 2277-2284.

- Choe, T.E., Cohen, I., Medioni G., Walsh, A., Sadda S., 2006. Evaluation of 3-D Shape Reconstruction of Retinal Fundus. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006) , 134-141.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Assoc. Comp. Mach.*, 24(6): 381-395.
- Floyd, R., 1962. Algorithm 97: Shortest Path. *Communications of the ACM*, 5:6 (1962), p. 345.
- Foracchia, M., Grisan, E., Ruggeri, A., 2004. Detection of Optic Disc in Retinal Images by Means of a Geometrical Model of Vessel Structure. *IEEE Transactions on Medical Imaging*, vol 23, no. 10, 1189-1195.
- Frangi, A., Niessen, W. J., Vincken, K. L., Viergever., M. A., 1998 Multiscale vessel enhancement filtering. *Medical Image Computing and Computer-Assisted Intervention (MICCAI 1998)*, 130-137, 1998
- Gluckman, J., Nayar, S.K., 2001. Rectifying transformations that minimize resampling effects. In: Proceedings of CVPR 2001 111-117.
- Hartley, R., Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Hartley, R., 1995. In Defence of the 8 point Algorithm. In: Proceedings of ICCV 1995, 1064-1070
- Hirschmüller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: Proceedings of CVPR 2005, 807–814
- Kai, Z., 2005. Stereo Matching and 3-D Reconstruction for Optic Disc Images. *Computer Vision for Biomedical Image Applications*
- Kim, J., Kolmogorov, V., Zabih, R., 2003. Visual correspondence using energy minimization and mutual information. In: Proceedings of ICCV 2003. Proceedings, 13-16 vol.2, Oct. 2003, 1033-1040
- Kolmogorov, V., Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. In: Proceedings of ICCV 2001.
- Kumar, R., Anandan, P., Hanna, K., 1994. Shape Recovery from Multiple Views: A Parallax Based Approach. In: Proceedings of ICPR 1994, 685-688.
- Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.
- Magarely, J., Dick, A., 1998. Multiresolution Stereo Image Matching using Complex Wavelet. In: Proceedings of ICPR 1998, Vol 1, 4-7.
- Malladi, R., Sethian, J. A., Vemuri, B. C., 1995, Shape modeling with front propagation: A level set approach, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 158-175, Feb. 1995.
- Matsopoulos, G.K., Mouravliansky, N.A., Delibasis, K.K., Nikita, K.S., 1999. Automatic registration of retinal images with global optimization techniques. *Transactions on Information Technology in Biomedicine*, vol. 3, 47–60.
- Moallem, P., Faez, K., 2001. Search space reduction in the edge based stereo correspondence. In: Proceedings of 6th International Fall Workshop on Vision, Modeling, and Visualization 2001 (VMV2001), 423-429.
- Pallawala, P.M.D.S, Hsu, W., Lee, M.L., Eong, K.-G. A., 2004. Automated Optic Disc Localization and Contour Detection Using Ellipse Fitting and Wavelet Transform. In: Proceedings of ECCV 2004, 139-151.
- Podoleanu, A.G., Rogers, J.A., Jackson, D.A., 2000. Three dimensional OCT images from retina and skin. Vol. 7, No. 9, *Optic Express*.
- Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, vol. 47, no. 1-3, 7–42.
- Stewart, C.V., Tsai, C.L., Roysam B., 2003. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Transactions on Medical Imaging*, Vol. 22, Issue 11, 1379-1394.
- Sun, J., Zheng, N.N., Shum, H., 2003 . Stereo Matching Using Belief Propagation. *IEEE Transactions on PAMI*, Vol. 25, No. 7.
- Tsai, C.L., Stewart, C.V., Tanenbaum, H.L., Roysam, B., 2004. Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images. *IEEE Transactions on Information Technology in Biomedicine*, Volume 8, Issue 2, 122 – 130
- Tsai, C.L., Majerovics, A., Stewart, C.V., Roysam, B., 2003. "Disease-Oriented Evaluation of Dual-Bootstrap Retinal Image Registration", In: Proceedings of MICCAI 2003, 754-761.
- Viola, P.A., Wells, W.M. III., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154.
- Walter T., Klein, J.-C., 2001. Segmentation of Color Fundus Images of the Human Retina: Detection of the Optic Disc and the Vascular Tree Using Morphological Techniques. In: Proceedings of ISMDA 2001, 282-287
- Wells, W.M. III, Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, vol. 1, no. 1, 35–51.
- Zana, F., Klein, J.C., 2001. Segmentation of Vessel-like Patterns Using Mathematical Morphology and Curvature Evaluation. *IEEE Transactions on Image Processing*, vol. 10, no. 7, 1010–1019.
- Zana, F., Klein, J.C., 1999. A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform. *IEEE Transactions on Medical Imaging*, vol. 18, issue 5, 419–428.
- Walsh, A.C., Updike, P.G., Sadda, S.R., 2005. Quantitative Fluorescein Angiography. *Retina*, 4th edition, Chapter 52, Mosby,
- Werner, C.T., Matas, J., 2005. Two-view Geometry Estimation Unaffected by a Dominant Plane. In: Proceedings of CVPR 2005, 772-779.
- Webpage of Middlebury stereo database and source code: www.middlebury.edu/stereo