

# Evaluation of 3-D Shape Reconstruction of Retinal Fundus

Tae Eun Choe<sup>1</sup>, Isaac Cohen<sup>1</sup>, Gerard Medioni<sup>1</sup>,  
Alexander C. Walsh<sup>2</sup>, and Srinivas R. Sadda<sup>2</sup>

<sup>1</sup> IRIS, University of Southern California  
3737 Watt way, Los Angeles, CA, 90248, USA  
{tchoe, icohen, medioni}@usc.edu

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

**Abstract.** We present a method for the 3-D shape reconstruction of the retinal fundus from stereo paired images. Detection of retinal elevation plays a critical role in the diagnosis and management of many retinal diseases. However, since the shape of ocular fundus is nearly planar, its 3-D depth range is very narrow. Therefore, we use the location of vascular bifurcations and a plane+parallax approach to provide a robust estimation of the epipolar geometry. Matching is then performed using a mutual information algorithm for accurate estimation of the disparity maps. To validate our results, in the absence of camera calibration, we compared the results with measurements from the current clinical gold standard, optical coherence tomography (OCT).

## 1 Introduction

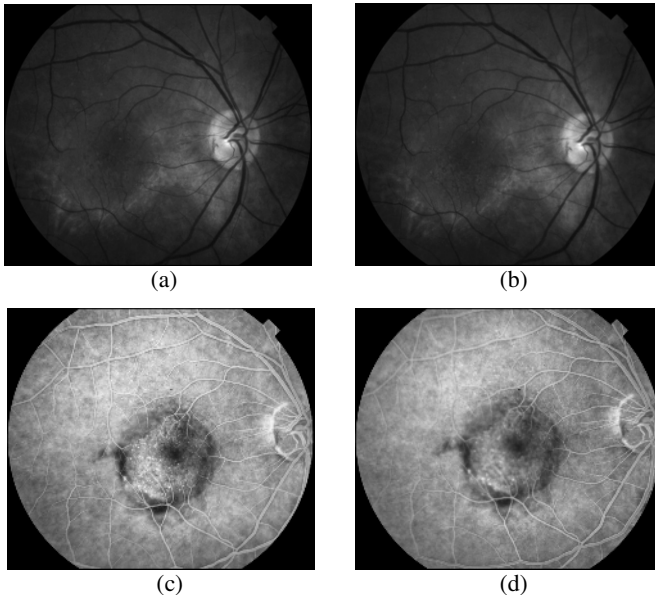
Retinal diseases are the most common causes of acquired blindness among adults in developed nations. Glaucoma is also a frequent cause of vision loss worldwide. One characteristic feature of all of these diseases is their distortion, either through excavation, elevation or thickening, of the ocular fundus topography. For this reason, 3-D visualization or imaging of the retina and optic nerve are necessary for the accurate diagnosis and evaluation of these disorders. Traditionally, this assessment has been performed by highly-trained eye-care professionals using examining microscopes. More recently, an imaging technique known as optical coherence tomography (OCT) has emerged for this purpose. 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  $3\mu\text{m}$  [11]. Unfortunately, the cost of OCT equipment and the expertise required for its interpretation has limited widespread adoption of this technology.

An alternative to subjective inspection and high-cost imaging consists of inferring the 3D shape of the retina using images acquired with a common, lower-cost fundus camera. This paradigm suggests a potentially cost-effective solution for developing nations faced with the emerging epidemic of diabetes mellitus. Since this analysis can

be carried out either on-site or remotely, it is also well-suited to a modern era of global telemedicine.

In the assessment of pathologic changes due to retinal disease or glaucoma, the 3-D reconstruction of two main areas of the ocular fundus is important: the optic disc and the macula. The numerous features provided by blood vessels merging at the optic disc make this area easier to reconstruct than the macula which is equally important in human vision. For example, the normal macula is devoid of large vessels and other consistent registration points. Therefore, stereoscopic reconstructions of this essential part of the eye are quite challenging. Retinal images also have unique features that prevent classical stereo algorithms from estimating the 3-D shape of the fundus from stereo paired images. Intensities and colors of the same physical positions may vary between consecutive images. In addition, the shape of ocular fundus is nearly planar and so, its 3-D depth is very narrow. This prevents accurate estimation of the epipolar geometry using traditional stereo algorithms. Finally, image blur may prevent accurate stereo reconstructions.

Scharstein and Szeliski present a comprehensive review of commonly used methods for stereo estimation [12]. Graph cut [7] and belief propagation [13] methods have shown good performance on test sets of stereo images which deal with scenes with large depth changes, similar intensities for identical pixels, and textured regions.



**Fig. 1.** Two pairs of stereo retinal images from a patient with age-related macular degeneration (a)(b) A characteristic 'blister' in the macula is harder to see in the red-free images (a)(b) than it is in the images taken after injection of fluorescein dye (c)(d) into the antecubital vein. Nevertheless, the image in (d) is blurred which complicates stereoscopic measurements.

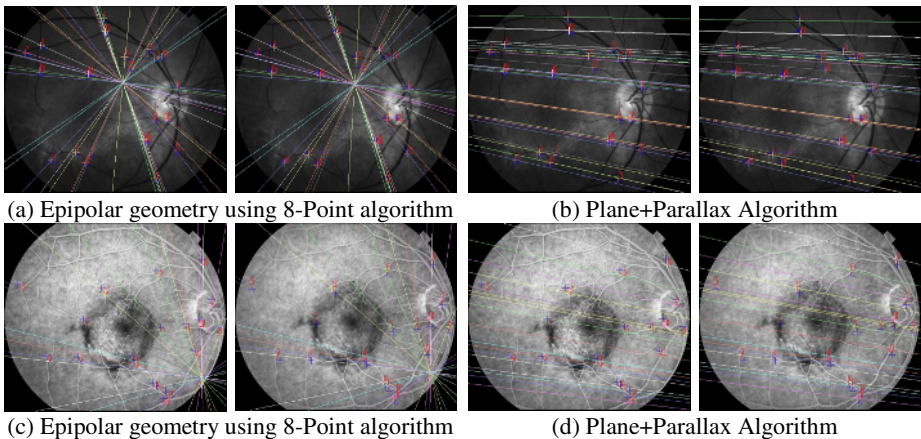
As discussed above, retinal images have challenging features such as inconsistent intensities, smooth and thin surfaces, and less texture in the image, that complicate stereoscopic estimations. Methods commonly used in the literature are not well adapted for stereo estimation of the retina from fundus images, as attested by our comparative study presented in the experimental results section. Figure 1 shows various sets of stereo images.

Our paper proposes a method for 3-D reconstruction from a pair of retinal images and its evaluation. The method consists of three steps. First, using matched Y-features, a plane+parallax approach is taken to estimate the epipolar geometry. After image rectification, the search space on the scanline for stereo matching is estimated based on the Y-feature correspondences. Second, our mutual information method matches points to produce a dense disparity map. Finally, the reconstructed 3-D shape is validated by OCT data.

## 2 Estimation of Epipolar Geometry

This section briefly describes the method for estimating the epipolar geometry from a pair of fluorescein images of the retina [1]. Many types of robust features can be used to estimate the epipolar geometry such as Harris corners [4] or SIFT features [10]. We utilized the locations of vessels' bifurcations, so-called Y-features, which are well distributed in the fundus and more robust to intensity changes within images. These features are used to estimate the epipolar geometry and search space in stereo matching. Extracted Y-features are matched by maximizing a mutual information criterion.

Typically, 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 [9]. Figure 2-(a)(c) shows the erroneous result of the



**Fig. 2.** (a)(c) Examples of inaccurately estimated epipolar geometry with 8-point algorithm in the case of a translation of the camera. (b)(d) The epipolar lines obtained from the plane+parallax based fundamental matrix estimation.

corresponding epipolar lines defined by the fundamental matrix estimated with an 8-point algorithm. In this example, the motion of the camera between the two acquisitions is close to a pure translation. However, a degenerate case occurs in the estimation of the fundamental matrix when all the points lie on a plane. This happens quite frequently in retinal imaging since the surface of the retina is relatively flat. Although the points are not on the same plane, the range of 3-D depth appears to be too narrow to get a satisfactory fundamental matrix from the 8-point algorithm.

To overcome this limitation of the 8-point algorithm on retinal images, we implemented the plane+parallax algorithm proposed by Kumar et al [8]. 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. We have implemented a RANSAC-based algorithm for the plane+parallax method. Fig. 2-(b)(d) shows the corresponding epipolar lines by the plane+parallax approach. After estimating the fundamental matrix, stereo images are rectified using Gluckman and Nayar’s method [3] for the estimation of the depth map.

### 3 Stereo Matching

After rectification, all matching pixels are located on the same scanlines. On the scanline, the lower and the upper bound of the search space is estimated by utilizing disparities of the matched Y-features [2]. This method produces a 1-D narrow search space, which enables a more accurate computation of the dense disparity map. In the case of retinal imaging where the intensities of the matching areas vary across the stereo images, a general cross-correlation based algorithm did not seem to provide satisfactory results. Instead, we implemented a mutual information algorithm for point matching along the scanlines and estimation of the depth map.

Mutual information for a pair of windows  $\mathbf{z}_A$  and  $\mathbf{z}_B$  is defined by [14]:

$$MI(\mathbf{z}_A, \mathbf{z}_B) = H(\mathbf{z}_A) + H(\mathbf{z}_B) - H(\mathbf{z}_A, \mathbf{z}_B) \tag{1}$$

where  $H(\mathbf{z}) = -\int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z}$  is Shannon entropy of the image window  $\mathbf{z}$ , and  $p$  is the distribution of the grey levels in the considered window. In [6], the authors 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. We consider the following approximation of the entropy [14]:

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

where  $N_z$  is the size of the window  $\mathbf{z}$ , and the density function  $p(z)$  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 sample in the Parzen window  $W_p$ , and  $g_\psi(z)$  is the uni- or bi-variate Gaussian density function with diagonal covariance matrix  $\psi$  [14]. In [5], the author estimated the discretized density function  $p(z)$  in order to reduce the time complexity.  $p(z)$  is calculated at the given point and it is convoluted by 2-D Gaussian to simulate the density function. However this simplification reduces the accuracy of the mutual information. The entropy function is then rewritten as:

$$H(\mathbf{z}) \approx -\frac{1}{N_z} \sum_{z_i \in \mathbf{z}} \ln \frac{1}{N_p} \sum_{z_j \in W_p} g_\psi(z_i - z_j) \quad (2)$$

We utilize Equation (2) to implement the mutual information. In mutual information, the marginal entropies  $H(\mathbf{z}_A)$  and  $H(\mathbf{z}_B)$  are included and the density function  $p(z)$  is calculated individually using Parzen window estimation with Gaussian assumption. The disparity of each pixel is determined by following equation.

$$disparity = \arg \max_{d \in S} MI(d) = \arg \max_{d \in S} MI(\mathbf{z}_A, \mathbf{z}_{B,d}) \quad (3)$$

where  $MI(\mathbf{z}_A, \mathbf{z}_{B,d})$  is the mutual information from Equation (1) 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 then is compared with a window  $\mathbf{z}_{B,d}$  in the other image within the search space  $S$ . The pair of windows that has the maximum mutual information determines the disparity of each pixel position.

To speed up the calculation of MI, we constructed a look-up table of Gaussian function in advance. Since the heaviest calculation of MI is the Gaussian function and the variable is an integer that ranges from 0 to 255, for each input value, the Gaussian function was pre-calculated and saved in the look-up table. The same method was applied to the 2-D Gaussian function, which requires two integer input variables. This implementation made the calculation time of MI 8 times faster than the regular MI implementation.

Since the range of disparity values is too narrow, 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 using subpixel accuracy by using a quadric interpolation relying on neighboring mutual information values. Every disparity is determined individually by mutual information matching criteria. We did not apply a smoothness constraint to calculate disparities. Based on the taxonomy of stereo algorithms [12], only the matching function is applied. Neither aggregation nor optimization methods are used. In our experiments we have noted that these methods degraded the accuracy of the matching.

## 4 Evaluation of 3-D Surface

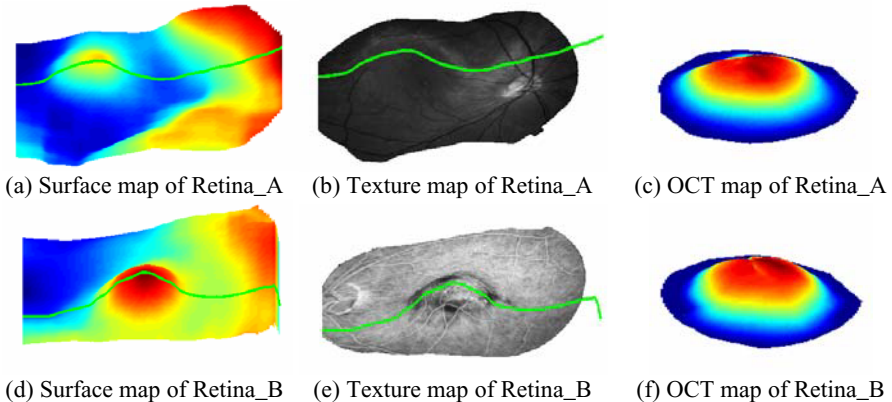
The reconstructed 3-D shape is evaluated by comparing our results to aligned OCT data from a commercial instrument (Carl Zeiss Meditec, Dublin, CA). Our data differs slightly from OCT data since we have a disparity map – not a true depth map. A more

accurate depth map requires knowledge of internal camera parameters that are currently unavailable.

#### 4.1 3-D Shape of Retinal Fundus

We conducted experiments on numerous stereo pairs and selected three pairs, named Retina\_A, B and C, for evaluation. Fig. 1 presents the two selected pairs of stereo images that had corresponding OCT data. 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 right stereo image of Retina\_B is unfocused, which makes the image reconstruction difficult.

After Y-features were extracted and matched from each image, the epipolar geometry was estimated with the plane+parallax algorithm (Fig. 2). The stereo images were rectified based on the fundamental matrix, and the search space was estimated from a set of matching Y-features. Using mutual information, subpixel resolution dense disparity maps were estimated (Fig. 3).



**Fig. 3.** Reconstructed 3-D shape of the retinal fundus (a)(d) Surface maps with an intersection line. (b)(e) Texture maps with intersection line in the same position. (c)(f) The enlarged 3-D OCT map of a blister area.

Fig. 3-(a)(d) shows different surfaces of 3-D retinal fundus that were obtained from several sequences of stereo images. Displayed depth values are scaled (by a factor of 30) to magnify the 3-D depth. Since the camera calibration is not performed, radial distortion and perspective effect still exist in the 3-D shape. The center of the estimated 3-D depth maps are very smooth with little noise and did not require a post processing step such as smoothing and regularization. The calculated depth values near the boundary are incorrect due to the occlusions.

In Fig. 3-(b)(e), the image texture is mapped over the 3-D depth map. The green lines, which respectively correspond to the line of the surface images on the left, display the 3-D depth of the texture. The 3-D shape of the blister and the optic disc are accurately estimated.

Fig. 3-(c)(f) presents OCT maps of the ‘blisters’. 3-D OCT maps are extracted from data exported by the OCT machine using custom software (OCTOR, Doheny Image Reading Center). The size and degree of the ‘blister’ areas closely match the OCT data.

## 4.2 Combination of OCT and 3-D Shape Model

The reconstruction of 3-D surfaces is one of the most difficult processes 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 hard to align the data to the 3-D shape data. Therefore, for evaluation, we manually matched several feature positions from the OCT to the 3-D depth map to align them to obtain a 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,  $Y$ , were calculated as an evaluation criterion.

$$E(X, Y, T) = \frac{1}{N_X} \sum_{\mathbf{x}_i \in X} \left\| T(\mathbf{x}_i) - \arg \min_{\mathbf{y}_j \in Y} \left\| T(\mathbf{x}_i) - \mathbf{y}_j \right\| \right\| \quad (4)$$

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

We evaluated the performance of different stereo matching algorithms using the evaluation codes from the Middlebury database [15] for SSD, dynamic programming, scanline optimization, and graph cut algorithms. The best parameters for each image were selected for each method. We implemented normalized cross-correlation and mutual information. Table 1 demonstrates that, in our study, mutual information outperformed other methods in terms of accuracy. Mutual information generated little noise and is highly accurate, even in areas with low textures.

**Table 1.** Difference between OCT data and 3-D shape in pixel metric. For each method, parameters with the best performance for each image are selected.

|                                    | Retina_A        | Retina_B        | Retina_C        |
|------------------------------------|-----------------|-----------------|-----------------|
| Sum of Squared Difference (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)            | <b>0.152694</b> | <b>0.328937</b> | <b>0.245442</b> |

## 5 Conclusion

Reconstructing the 3-D shape of the human retina from a stereo pair of fundus images is a challenging task because of intensity variation, the paucity of features in the macular area, and the plane-like surface of the retina. In this paper, we proposed a method for 3-D reconstruction of the retinal surface in an attempt to overcome such difficult cases. Specifically, our method employs a plane+parallax approach for robust epipolar geometry estimation and utilizes the mutual information criteria for

estimating a dense stereo map. The proposed method has been tested on numerous stereo sequences and validated by aligning OCT data with reconstructed 3-D depth data. Techniques such as this can be implemented easily in a telemedicine system which may have great utility in addressing the epidemics of eye disease that are emerging around the world due to diabetes, macular degeneration and glaucoma.

## Acknowledgements

This research was partially supported by the National Institute of Health (NIH) under grant No. R21 EY015914-01, the Doheny Eye Institute, and the James H. Zumberge Faculty Research and Innovation Fund.

## References

1. T.E. Choe, I. Cohen, "Registration of Multimodal Fluorescein Images Sequence of the Retina," ICCV 2005, pp. 106-113, 2005
2. T.E. Choe, I. Cohen, G. Medioni, "3-D Shape Reconstruction of Retinal Fundus," CVPR 2006, 2006
3. J. Gluckman and S. K. Nayar. Rectifying transformations that minimize resampling effects. In IEEE Proceedings of Computer Vision and Pattern Recognition, Kauai, December 2001.
4. C. Harris and M. Stephens, "A combined corner and edge detector", Proc. Alvey Vision Conf., Univ. Manchester, pp. 147-151, 1988.
5. H. Hirschmüller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," CVPR 2005, Volume 2, 20-26 June 2005 Page(s):807 – 814
6. J. Kim, V. Kolmogorov, R. Zabih, "Visual correspondence using energy minimization and mutual information", Ninth IEEE International Conference on Computer Vision, 2003. Proceedings, 13-16 Page(s):1033 - 1040 vol.2, Oct. 2003
7. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in International Conference on Computer Vision, 2001.
8. R. Kumar, P. Anandan and K. Hanna, "Shape Recovery From Multiple Views: A Parallax Based Approach," DARPA IU Workshop, Monterey, Calif., Nov. 1994
9. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. Nature, 293:133–135, Sept 1981.
10. D. Lowe, "Distinctive image features from scale invariant keypoints," IJCV, Vol. 60, No. 2, pp. 91-110, 2004
11. A.G. Podoleanu, J.A. Rogers, D.A. Jackson, "Three dimensional OCT images from retina and skin", Vol. 7, No. 9, Optic Express, Oct 2000
12. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV, vol. 47, no. 1-3, pp. 7–42, April 2002.
13. J. Sun, N.N. Zheng, H. Shum, "Stereo Matching Using Belief Propagation," IEEE Trans. on PAMI, Vol. 25, No. 7, July 2003
14. P.A. Viola and W.M. Wells, III. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137–154, September 1997.
15. Webpage of Middlebury stereo database and source code: [www.middlebury.edu/stereo](http://www.middlebury.edu/stereo)