Proceedings of the 29th Annual International Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Three-Dimensional Reconstruction of Optic Nerve Head from Stereo Fundus Images and Its Quantitative Estimation

Toshiaki Nakagawa, Yoshinori Hayashi, Yuji Hatanaka, Akira Aoyama, Takeshi Hara,
Akihiro Fujita, Masakatsu Kakogawa, Hiroshi Fujita, and Tetsuya Yamamoto

Abstract— **It is important for diagnosis of glaucoma to grasp 3-D structure of an optic nerve head (ONH). The quantitative 3-D reconstruction of the ONH is required for the diagnosis. We propose a technique to obtain the depth value from stereo image pair of a retinal fundus for the 3-D reconstruction of the ONH. Our technique mainly consists of four steps: (1) cutout of the ONH region from the fundus images, (2) registration of the stereo pair, (3) disparity detection, and (4) depth calculation. For quantitative estimation of the depth value measured by using this method, the depth value was compared with the measurement results determined from the Heidelberg Retina Tomograph (HRT), which is a confocal laser-scanning microscope. As a result, the depth value of the ONH obtained using the stereo retinal image pair was in accordance with that obtained using the HRT (*r*=0.91). These results indicate that the stereo fundus images could be useful for assessing the depth value of the ONH for the diagnosis of glaucoma.**

I. INTRODUCTION

The cup/disc (C/D) ratio, which is the ratio of the diameter of the depression (cup) to that of the optic nerve head (ONH, disc), is one of the important parameters for an early diagnosis of glaucoma. The C/D ratio is generally used in clinical practice because its value is greater in the case of glaucoma. The interpretation of the ONH, which actually has a 3-D structure, by using a 2-D image is subjective and there is a wide variation between the examination of the ONH by different observers and even between the examinations by the same observer [1]. A more quantitative alternative is to use the Heidelberg Retina Tomograph (HRT), which is a confocal laser-scanning microscope, for the acquisition and analysis of the 3-D measures of the ONH [2, 3]. It has been revealed that an HRT is capable of ONH imaging, and it is an established technique for detecting glaucomatous structural changes.

A computerized technique for the qualitative estimation of the depth of the ONH from the stereoscopic pairs of retinal-fundus images has been suggested for the 3-D analysis of the depression of the ONH [4, 5]. It has been shown that this technique is useful for the investigation of the 3-D measures of the ONH. However, the experimental results regarding the quantitative depth value calculated from the stereo image pair of the ONH have not been reported. Moreover, there have been no studies in which the depth value calculated from the stereo image pair has been compared with the HRT outputs.

In this study, an automatic method for reconstructing the 3-D structure of the ONH by using the stereo fundus images is proposed. In order to evaluate the accuracy of our method, the depth values of the ONH obtained from the stereo fundus image pairs are compared with the HRT measurement results.

II. METHODOLOGY

In our technique, the depth value is obtained from the stereo fundus image pair; it mainly consists of four processes. A stereo image pair consists of a "left image" and a "right image" captured from different perspectives. The stereo image pair can be generated by taking two shots with a parallel shift using a single-lens fundus camera or by taking a single shot using a stereo-fundus camera. In the first step, the images of the ONH region are cut out from the original fundus images in the first step. In the second step, the registration process of the stereo ONH image pair is performed in order to remove any displacements. In the third step, the "corresponding points" in each stereo ONH image are detected. In the fourth step, the depth values of the 3-D structures are calculated from the results of the disparities detected in the configuration of the corresponding points.

A. Cutout of the ONH region

The images of the ONH region were cut out from the original stereo image pair in order to reduce the processing area to expedite the subsequent steps. In this processing step, the fundus images were cropped to form quadrates at the position of the ONH region that was extracted automatically. The ONH region has relatively high pixel values in three channels (R, G, and B components) in the color stereo fundus image pairs. P-tile thresholding [6] can be applied to define a threshold for an approximate extraction of the ONH region because individual variations of the ONH do not vary significantly.

The blood vessels (BVs) running on the surface of the ONH interfered with the correct extraction of the ONH region

---

Manuscript received April 2, 2007. This work was supported in part by a grant for the "Knowledge Cluster Creation Project" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

T. Nakagawa, Y. Hayashi, A. Aoyama, T. Hara, H. Fujita and T. Yamamoto are with the Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan (e-mail: nakagawa@fjt.info.gifu-u.ac.jp).

Y. Hatanaka is with the Department of Electric Control Engineering, Gifu National College of Technology, 2236-2 Kamimakuwa, Motosu, Gifu 501-0495, Japan.

A. Fujita is with the Kowa Company, Ltd., 1-3-1 Shinmiyakoda, Hamamatsu, Shizuoka 431-2103, Japan.

M. Kakogawa is with the TAK Co., Ltd., 4-35-12 Kono, Ogaki, Gifu 503-0803, Japan.

in the P-tile thresholding operation. In order to solve this problem, the extraction of the ONH region was performed by using the images in which the BVs were erased. These erased pixels were then interpolated by using the RGB values of the pixels in the surrounding region. The pixel value P used in the interpolation was calculated as

$$P=\frac{\sum_{k=1}^{n}\frac{p_k}{l_k}}{\sum_{k=1}^{n}\frac{1}{l_k}}, \tag{1}$$

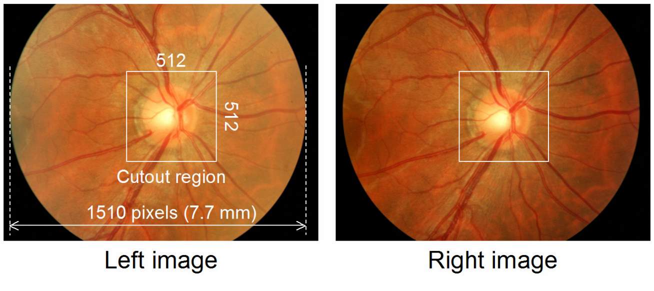
where p_k denotes the value of the pixels in the surrounding region, n is the number of surrounding pixels, and l_k is the distance between the interpolated pixel and each surrounding pixel.

The BVs were extracted by using the black-top-hat transformation, which is a type of grayscale morphological operation, from the G-component of the color fundus images. The structure element used in this transformation was a disc whose diameter was set to the same level as the thickness of the BVs in the ONH region. The regions containing BVs were extracted after applying the Otsu thresholding technique [7] to the black-top-hat-transformed image.

The center of the square of the cutout of the ONH region was the gravity point of the ONH region extracted from the images in which the BVs were erased. In this study, the size of the original fundus image was 1600 × 1200 pixels, the angle of view was 27°, and the size of the cutout region was 512 × 512 pixels, as shown in Fig. 1.

B. *Registration of the stereo fundus image pair*

The disparity, which is defined as the difference in the position of the corresponding points in the stereo image pair, depends on the change in the position of not only the camera but also the subject (subject's motion). The disparity due to the subject's motion affects the calculated result of the depth value. In order to accurately measure the depth value, it is necessary to rectify the disparity due to the subject's motion. However, it is difficult to determine the motion that induces the disparity only on the basis of observations.



Left image Right image

Fig. 1. Example of an original stereo fundus image pair and cutout region from the left and right images.

In the retinal fundus, the bell-shaped ONH has a dent on the opposite side, that is, the side facing the camera. Therefore, the cup region of the ONH exists at a distant position from the camera and the disparity is small. Theoretically, it is possible to use the pixels in the ONH region, which have small disparities, for image registration. However, the region of the retina around the ONH is more suitable for the image registration task because the blood vessels in the retina exist even on the curved surface; in this region, the right and left images exhibit a parallel shift. Moreover, the 3-D structure of the retinal region is simpler than that of the ONH. From the abovementioned description, the image registration for rectifying the disparity due to a subject's motion was performed by using the pixels from regions other than the ONH region.

In order to exclude the ONH region from the registration process, the pixels of a fundus image were allocated to two regions: the retina and the ONH region. The pixels in the retina region were used for registering the stereo image pair. The boundary between the two regions was obtained by automatically extracting the ONH region.

In the first step, a contour of the ONH region was extracted from two stereo fundus images. The ONH region has a tendency to have a higher pixel value than the other regions. Furthermore, the contour of the ONH region can be expressed with a smooth closed curve in many cases. Therefore, the contour that exhibits high edge intensity, which is defined as a change in the brightness, was extracted as the smooth closed curve by using an active-contour model [8].

In the next step, the registration of the stereo image pair was performed by using all the pixels of the images from regions other than the ONH region. If the positional error is minimum, the sum of all the differences between the pixel values of the two images will be minimum. Therefore, the right image was translated and rotated until the sum of all the differences was minimum. This registration procedure used the cross-correlation r between the two images, which is calculated as

$$r=\frac{\sum_{i=0}^{W}\sum_{j=0}^{H}\{L(i,j)-\overline{L}\}\{R(i,j)-\overline{R}\}}{\sqrt{\sum_{i=0}^{W}\sum_{j=0}^{H}\{L(i,y)-\overline{L}\}^2}\sqrt{\sum_{i=0}^{W}\sum_{j=0}^{H}\{R(i,j)-\overline{R}\}^2}}, \tag{2}$$

where L and R are the feature values in the coordinate system (i, j) of the left and right images, respectively; W and H, the width and height of the image, respectively; and $\overline{L}$ and $\overline{R}$, the average pixel values in the left and right images, respectively. The features used in the registration were the pixel values of the RGB components and its edge images created by the Sobel filter using the RGB values.

C. *Disparity detection*

The required disparity for obtaining the depth value was calculated from the location differences between the corresponding points. The detection of the corresponding points was performed using the pixels of the area within the registered ONH image pair. The detection of the corresponding points comprised the search for a point on the right image that corresponded to the reference point on the left image. The search was performed by setting up regions of interest (ROIs) including the pixels around the reference point and the candidate point separately. Two points on the left and right images having a similar texture in their respective ROIs were regarded as the corresponding points. The similarity was measured by the cross-correlation *r* defined as

$$r = \frac{\sum_{i=-W/2}^{W/2} \sum_{j=-H/2}^{H/2} \{L(x_L+i, y_L+j) - \overline{L}\}\{R(x_R+i, y_R+j) - \overline{R}\}}{\sqrt{\sum_{i=-W/2}^{W/2} \sum_{j=-H/2}^{H/2} \{L(x_L+i, y_L+j) - \overline{L}\}^2} \sqrt{\sum_{i=-W/2}^{W/2} \sum_{j=-H/2}^{H/2} \{R(x_R+i, y_R+j) - \overline{R}\}^2}}, \quad (3)$$

where *L* and *R* are the feature values of the pixels in the ROIs set in the ONH image pair, $\overline{L}$ and $\overline{R}$ are the average feature values of the ROIs, (x_L, y_L) is the coordinate of the reference point in the left image, (x_R, y_R) is the coordinate of the candidate point in the right image, and *W* and *H* are the width and height of the ROI, respectively.

The features used in the detection of the corresponding points were the pixel values of the RGB components and its edge images created by the Sobel filter using the RGB value. The parameters of this process are shown in Fig. 2. The size of the ROI was set to 21 × 21 pixels and the searching range was set to 41 × 23 pixels ($[x_L - 5$ pixels, $x_L + 15$ pixels$]$). The reference points arranged in the equally spaced positions and the interval were set to 4 pixels. The point having the maximum cross-correlation coefficient was considered to be the corresponding point. When the maximum value of the cross-correlation coefficient was smaller than a preset threshold value, it was assumed that the corresponding point of the reference point was not found. The disparity of the point that did not have a corresponding point was interpolated by the average of the disparities of the surrounding reference points.

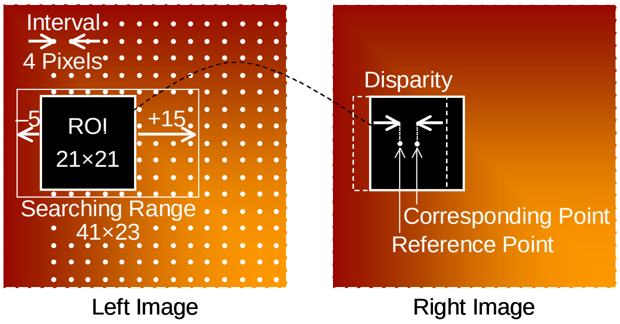


Fig. 2. Parameters in detecting the corresponding points in the left and right images.

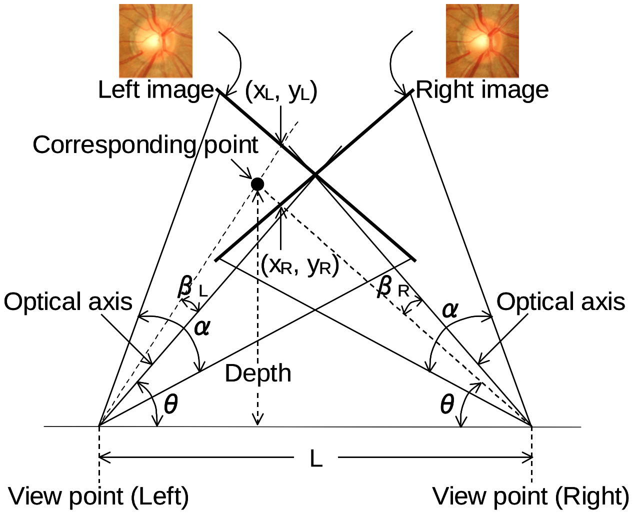


Fig. 3. Convergent visual system for depth calculation of stereo image pair.

D. *Depth calculation*

The depth value of the 3-D position was determined according to the value of the disparity in each location of the reference point. The depth value was calculated as

$$Depth = \frac{L \times \tan(\theta - \beta_L) \times \tan(\theta + \beta_R)}{\tan(\theta - \beta_L) + \tan(\theta + \beta_R)}, \quad (4)$$

$$\beta_L = \tan^{-1}\{x_L \times \tan(\frac{\alpha}{2} \times \frac{\pi}{180}) \times \frac{2}{W}\}, \quad (5)$$

$$\beta_R = \tan^{-1}\{x_R \times \tan(\frac{\alpha}{2} \times \frac{\pi}{180}) \times \frac{2}{W}\}, \quad (6)$$

and

$$x_R = x_L + disparity, \quad (7)$$

where x_L and x_R are the horizontal coordinates of the corresponding points in the left and right images, respectively. The original points in the coordinate system were arranged on the optical axis of the left and right viewpoints. α is the angle of view of the images; β, the angle between the position of the corresponding point and the optical axis; W, the width of the images; and L, the length of the baseline, which is the distance between the optical centers of the camera.

III. Results and Discussions

The proposed technique was evaluated using 12 normal fundus stereo image pairs. Figure 4 shows the depth values of the same ONH measured using the stereo image pairs and the HRT. The results of the depth measurements using the stereo image pairs agreed well with the results of the HRT at

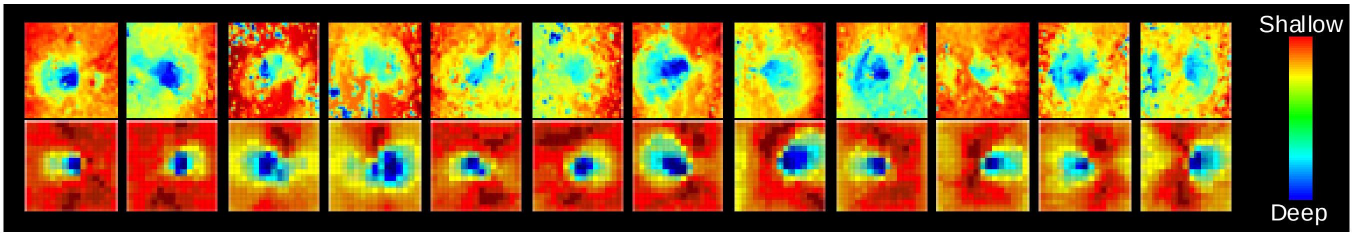


Fig. 4. The upper row shows the depth maps of the OHN from stereo fundus image pairs. The lower row shows the depth maps obtained using the HRT. The depth of the ONH obtained using the stereo images agreed with the results obtained using the HRT.

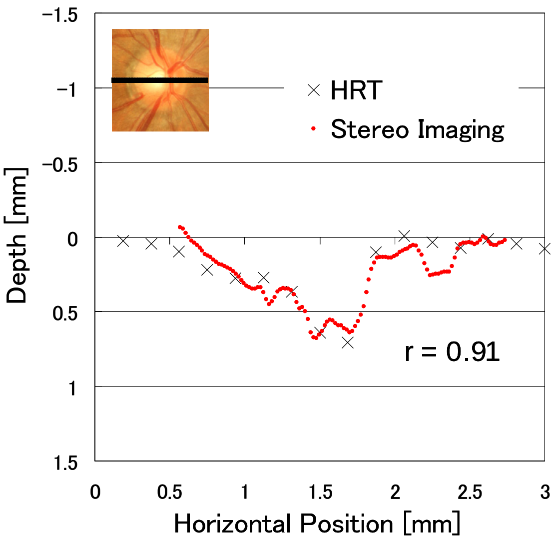


Fig. 5. Example of depth information obtained from the stereo fundus image pair and HRT.

corresponding levels. The depth values measured using the stereo image pairs fluctuated. This is due to the slight uncertainty in the disparity detection. It is presumed that the median filter can effectively eliminate this incorrect disparity. Figure 5 shows the profile trace along the center of the OHN. The correlation coefficient *r* between the two results obtained by the two methods was 0.91. This result indicates the validity of proposed method to obtain quantitative depth values.

IV. CONCLUSION

In this study, we conducted quantitative measurements of the depth value of an ONH region from the stereo fundus images. The depth values obtained from the stereo image pairs were in accordance with the results of the HRT. These depth values may be useful as assisting parameters for ophthalmologists in the diagnosis of the degree of glaucoma.

ACKNOWLEDGMENT

This work was partly supported by a grant for the Knowledge Cluster Creation Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The authors would like to acknowledge the contribution of Dr. R. Shiraki of the Shiraki Eye Clinic for the acquisition of the HRT data.

REFERENCES

[1] J. M. Tielsch, J. Katz, H. A. Quigley, A. Sommer, “Intraobserver and interobserver agreement in measurement of optic disc characteristics,” Opthalmol., vol.95, no.3, pp.350–356, 1988.

[2] F. S. Mikelberg, C. M. Parfitt, N. V. Swindale, S. L. Graham, S. M. Drance, R. Gosine, “Ability of the Heidelberg Retina Tomograph to detect early glaucomatous visual field loss,” J. Glaucoma, vol.4, no.4, pp.242–247, 1995.

[3] M. Iester, F. S. Mikelberg, S. M. Drance, “The effect of optic disc size on diagnostic precision with the Heidelberg Retina Tomograph,” Ophthalmol., vol.104, no.3, pp.545–548, 1997.

[4] E. Corona, S. Mitra, M. Wilson, T. Krile, Y. H. Kwon, P. Soliz, “Digital stereo image analyzer for generating automated 3-D measures of optic disc deformatin in glaucoma, IEEE Trans. Med. Imaging, vol.21, no.10, pp.1244–1253, 2002.

[5] J. Xu, O. Chutatape, “Auto-adjusted 3-D optic disk viewing from low-resolution stereo fundus image,” Computers in Biology and Medicine, vol.36, pp.921–940, 2006.

[6] J. R. Parker, “Algorithms for Image Processing and Computer Vision,” Wiley Computer Publishing, 1997.

[7] N. Otsu, “A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man. Cybern. vol.SMC-9, pp.62–66, 1979.

[8] M. Kass, A. Witikin, D. Terzopoulos, “Snakes: Active contour models,” Int. J. Computer Vision, vol.1, pp.321–331, 1987.