Stereoscopic Reconstruction Of Retina Images Based On Angiographic And Laser-Scan Ophthalmoscopic Images

Helder J. Araujo
Dept. of Electrical Engineering
University of Coimbra
3000 Coimbra
Portugal

Rufino Silva
Dept. of Ophthalmology
University Hospital
3000 Coimbra
Portugal

Jose F. Abreu
Dept. of Ophthalmology
University Hospital
3000 Coimbra
Portugal

Jose G. Cunha-Vaz
Dept. of Ophthalmology
University Hospital
3000 Coimbra
Portugal

Introduction

here are several imaging devices that enable the acquiition of images from the human retina. In this paper we tart by considering the images obtained by means of ngiography. Stereoscopic (3D) reconstruction is first erformed by using only angiographic images. Next ie combination and fusion of information from both ngiographic and laser-scan ophthalmoscopic images is onsidered in order to overcome a number of difficul-es associated with the processing of the angiographic nages.

From the medical point of view the 3D reconstruc-on and quantification of anatomical structures in the uman retina is important for the evaluation of sev-al pathologies. In particular it is very useful to have ie 3D quantification and reconstruction of the macu-r edema or thickening of the retina. Also useful is the ieasurement of the lesions' thicknesses. This type of econstruction is not only useful for diagnosing patholo-es but also to evaluate the effects of certain type of eatments. For example it would be helpful to have ich a system to evaluate the reversibility of the laser eatment in small tumors.

Angiographic images are acquired by two cameras hose optical axes can be considered parallel. The dif-culties associated with the angiographic images of the tina are a consequence of both the retina geometry id composition. The retina has basically a spheri-l shape and is made up of transparent layers of cells. o deal with these problems we are using and compar-g several types of algorithms for determining stereo rrespondence. The angiographic images have a large epth of field and therefore several anatomical struc-ires are simultaneously in focus. Several matching gorithms are being tested, the main difficulty being ie ill-defined boundaries in shape and reflectance of ie anatomical structures. To overcome some of these fficulties we decided to use also laser-scan ophthalmoscopic (L.S.O.) images. L.S.O. images are obtained with extremely shallow depths of field. Furthermore the difference in depth between the several images is also known. However the combination and fusion of data from the reconstructed images from the stereo pairs and the L.S.O. images raises the problem of calibration and registration.

Stereoscopic Reconstruction from Angiographic Images

The pairs of the retina stereo images were obtained without the injection of a dye (which is done in fluorescein angiography). Figure 1 shows a stereo pair of a right eye retina, as they are obtained, i. e. without any pre-processing. We are interested in reconstructing not only the vessels but *all* the anatomical structures present in the retina. The retina is a delicate multilayer light-sensitive and transparent membrane. It is permeated with blood vessels, which are derived from the central retinal artery and vein. The thickness of the retina is maximum (0.5 mm) near the papilla, which is the region where the optic nerve leaves the eye. At the equator the retina is 0.2 mm thick. Usually the retina is subdivided into ten more or less well-defined layers, each of which has its own functional significance.

It is the transparency of the anatomical structures present in the retina that makes the stereo correspondence problem specially difficult. As a consequence of the transparency, the boundaries of the structures are ill-defined both in terms of shape and reflectance. The problem of stereo correspondence has been an extremely important research topic in computer vision. A wide variety of algorithms has been proposed to solve the correspondence problem [Marr and Poggio, 1979]. Some of the most popular classes of methods are:

- Intensity-based area correlation techniques.
- Relaxation techniques.

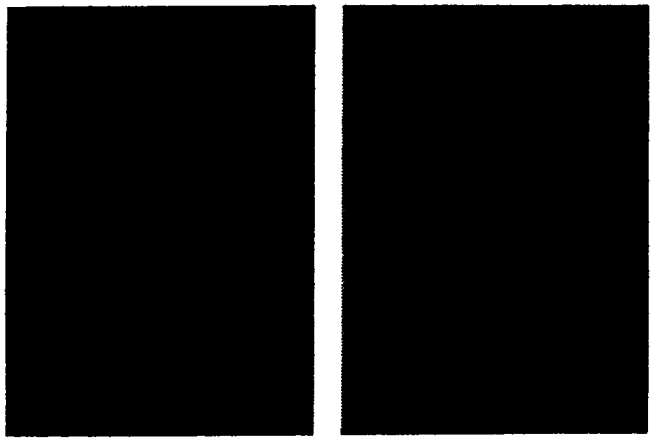

Figure 1: Left and right images of a right retina

- Dynamic programming [Ohta and Kanade, 1985].
- Prediction and verification techniques.

These classes of methods can also be classified into intensity-based or feature-based algorithms according to the type of primitives they use in trying to find correspondences. Many of the feature-based stereo algorithms use edges as primitives. These matching primitives should be specific, so that corresponding points may be accurately localized,and they should be stable under photometric and geometric distortions caused by the difference in viewing position. In the case of the retina images the anatomical structures have ill-defined boundaries in shape and reflectance and exhibit non-step edge profiles. For these reasons we decided not to use edge-based stereo algorithms. We have already tested several intensity-based algorithms. One of them was the algorithm described in [Kim *et al.*, 1990]. This algorithm establishes correspondences between points by examining whether two points have similar intensity gradients within a small neighborhood around those two points. Even though the global performance of the algorithm was satisfactory, its computational cost prevents its practical use in a medical application. Due to the particular nature of our images we decided to test and use an algorithm based on the one described in [Cox *et al.*, 1992].

The Stereo Correspondence Algorithm

The central assumption of this stereo matching algorithm is that corresponding pixels in the left and right images have gray level values that are normally distributed about a common value. The problem of establishing correspondence is then formulated as one of optimizing a Bayesian maximum likelihood cost function. The problem of stereo matching is considered as a special case of a sensor fusion problem. The cost function used is based on the approach proposed by [Pattipati *et al.*, 1990] for multisensor data association. Within the framework of this approach the problem of stereo correspondence is considered a special case of a data association problem, in which we try to associate the measurements, i. e. the images, obtained by the two cameras. Let $s = 1, 2$ denote the two cameras. Suppose that the number of measurements of sensor s is n_s. These measurements represent the pixels' gray-level values obtained by each camera along corresponding epipolar lines. Assume that there are T 3D points (T unknown) on the epipolar plane. Each camera s measures only a gray-level value $g_{s,j}$ of each of the potential 3D points j. The noisy scalar measurement $z_{s,i}$ is given by

$$z_{s,i} = (1 - \alpha_{s,i})[g_{s,j} + v_{s,i}] + \alpha_{s,i}\omega_{s,i} \qquad (1)$$

where $\alpha_{s,i}$ is a binary random variable such that $\alpha_{s,i} = 0$ when the measurement is from a true 3D point, and $\alpha_{s,i} = 1$ if it is a spurious measurement. The statistical errors associated with the measurements of a true 3D point $v_{s,i}$ are assumed to be $N(0, \sigma_s^2)$. The measurement noises are assumed to be independent across cameras. The density of spurious measurements $\omega_{s,i}$ is assumed to be uniform. Let Z_s represent the set of measurements obtained by each camera along corresponding epipolar lines. Let also P_{D_s} denote the detection probability of camera s. To account for the incomplete pixel-3D point associations caused by missed detections, a dummy measurement $z_{s,0}$ is added to each of the measurements of camera s. The matching to $z_{s,0}$ indicates no corresponding point. The set of measurements from camera s (including the dummy measurement $z_{s,0}$) is denoted by $Z_s = \{z_{s,i_s}\}_{i_s=0}^{n_s}$ where n_s is the number of measurements from camera s. For epipolar alignment of the scanlines, Z_s is the set of measurements along a scanline of camera s. These measurements z_{s,i_s} are the gray-level intensity values. Each measurement z_{s,i_s} is assumed to be corrupted by additive white noise.

The condition that intensity value z_{1,i_1} from camera 1, and measurement z_{2,i_2} from camera 2 originate from the same location x_p in space, i. e. that z_{1,i_1} and z_{2,i_2} correspond to each other is denoted by Z_{i_1,i_2}. That is Z_{i_1,i_2} is a 2-tuple of measurements. For example, $Z_{i_1,0}$ denotes occlusion of feature z_{1,i_1} in camera 2.

The likelihood function of this 2-tuple being the set of gray-level values that originated from the same 3D point is the mixed probability density function:

$$\begin{aligned}\Lambda(Z_{i_1,i_2} \mid x_p) = \\ [P_{D_1} p(z_{1,i_s} \mid x_p)]^{1-\delta_{0i_1}} [1 - P_{D_1}]^{\delta_{0i_1}} \cdot \\ \cdot [P_{D_2} p(z_{2,i_s} \mid x_p)]^{1-\delta_{0i_2}} [1 - P_{D_2}]^{\delta_{0i_2}} \qquad (2)\end{aligned}$$

where δ_{0,i_s} is the Kronecker function defined by

$$\delta_{0,i_s} = \begin{cases} 1 & \text{if } i_s = 0 \text{ denoting a missed detection,} \\ & \text{i.e. a measurement that is not assigned} \\ & \text{a corresponding point (it is occluded)} \\ 0 & \text{otherwise} \end{cases}$$

Therefore function $\Lambda(Z_{i_1,i_2} \mid x_p)$ denotes the likelihood that the measurement pair Z_{i_1,i_2} orginated from the same 3D point x_p and also represents the local cost of matching two points z_{1,i_1} and z_{2,i_2}. The term $p(z \mid x_p)$ is a probability density distribution that represents the likelihood of measurement z, assuming it originated from a point $x_p = (x, y, z)$ in the scene. The parameter P_{D_s}, as already mentioned, represents the probability of detecting a measurement originating from x_p at sensor s. This parameter is a function of the noise, number of occlusions, etc. It is assumed that the measurement vectors z_{s,i_s} are normally distributed about their ideal value z. Once established the cost of the individual pairings Z_{i_1,i_2}, it is necessary to determine the cost of all pairs. Denote by

$$\gamma = \{\{Z_{i_1,i_2}, i_1 = 0, 1, \ldots, n_1; i_2 = 0, 1, \ldots, n_2\}, Z_f\}$$

a feasible pairing of the set Z into measurement-3D point associations (2-tuples) Z_{i_1,i_2} except the case when all $i_s = 0$. The set Z_f contains the measurements from all sensors deemed spurious, that is, not associated with any 3D point in the pairing γ. Denote the set of all feasible partitions as $\Gamma = \{\gamma\}$. We wish to find the the most likely pairings or partitions γ of the measurement set Z. The most likely pairings are obtained by maximizing, over the set of all feasible partitions, the ratio of the joint likelihood function of all the measurements to the likelihood function when all the measurements are declared false. Let us denote the latter partition as γ_0. Then γ_0 denotes the case where all measurements are unmatched, i.e. the case in which there are no corresponding points in the left and right images. That is, we want to determine

$$\max_{\gamma \in \Gamma} \frac{L(\gamma)}{L(\gamma_0)}$$

where the likelihood $L(\gamma)$ of a pairing is defined as

$$L(\gamma) = p(Z_1, Z_2 \mid \gamma) = \prod_{Z_{i_1,i_2} \in \gamma} \Lambda(Z_{i_1,i_2} \mid x) \left(\frac{1}{\phi_1}\right)^{m_1} \left(\frac{1}{\phi_2}\right)^{m_2} \quad (3)$$

where ϕ_s is the field of view of camera s and m_s is the number of unmatched measurements from camera s in partition γ. Basically we assume that the density of the spurious measurements is uniform and is given by $1/\phi_s$. The likelihood of no matches $L(\gamma)$ is given by $L(\gamma_0) = 1/(\phi_1^N \phi_2^M)$, where N and M are the total number of points in the left and right images respectively. The maximization of the ratio of likelihood functions is performed using a dynamic programming algorithm subject to uniqueness and ordering constraints. These constraints imply the following assumptions:

a feature in the left image can match to no more than one feature in the right image and vice-versa (uniqueness).

Figure 2: Image with disparities coded as gray levels

- if z_{i_1} is matched to z_{i_2} then the subsequent image point z_{i_1+1} may only match points z_{i_2+j} for which $j > 0$.

The result from the application of this algorithm to the images of Figure 1 is displayed in the image of Figure 2. This results was obtained with the following parameter values: $P_{D_1} = P_{D_2} = 0.99, \sigma = 2.0$ and a maximum disparity of 25 pixels. In this image the disparities are coded as gray levels. Black (gray-level 0) corresponds to occlusion.

Reconstruction from Laser-Scan Ophthalmoscopic Images and Stereo

L.S.O. images are obtained with extremely shallow depths of field (in the case of the equipment we are using the depth of field is less than 0.1 μm). One of these images is displayed in Figure 3. Our equipment permits the acquisition of a sequence of nine images spaced 150 μm apart. We do not know the exact depth at which each image is acquired but we do know their relative depth ordering. Due to the thickness of the retina not all of the nine images are useful. Usually only 4 or 5 of them are images of the retina; the remaining are images of the vitreous body. Due to the nature of these images some of the retina structures are in focus in no more than one or two images of the sequence. Other structures, like the papilla may be visible in several images but with slightly different sizes. To 3D reconstruct the retina we propose to combine this sequence of images with the stereo reconstructed image. The stereo reconstructed image is obtained from the computed disparities, using the results from the camera calibration. For that purpose we are currently using a procedure [Batista *et al.*, 1986] that improves on Tsai's method [Tsai, 1987], adequately adapted to the ophthalmoscope we are using. By using the L.S.O. images we aim at reconstructing the retina with higher accuracy than it would be possible by using only the stereo images. Furthermore it is also possible, by combining

Figure 3: Laser-scan ophthalmoscopic image

both types of images, to detect and characterize small abnormalities of the anatomical structures that are not completely visible in just one of these types of images.

To combine both types of images the problem of registration has to be solved. Therefore calibration of L.S.O. images is also performed. Currently we are using a simple and heuristic procedure that combines a calibration with a known pattern with a calibration based on the knowledge of the physical dimensions of the anatomical structures. The latter type of calibration is performed by using retina images of a healthy individual (where it is expected that the anatomical structures have standard dimensions). As the L.S.O. images are acquired with much higher accuracy than the angiographic images we use the former to correct and improve the angiographic based stereo reconstruction. Consequently anatomical structures have to be matched in both types of images. For that purpose the L.S.O. images are considered to make up a "stack" and a focus measuring operator is applied to all the images [Dias *et al.*, 1991]. The focus measuring operator we are using is based on the computation of the image gradient. This operator is applied in all the images of the sequence. Next each image of the sequence is divided into smaller regions or windows which are analyzed by considering each region in all the images. This way it is possible to determine, for each region, the image of the sequence where that region is more on focus (we try to a maximum in the response of the operator measuring the sharpness of focus). Matching can then be performed based on domain-specific knowledge. It is extremely difficult and complex to implement a fully automated matching procedure, due to the specific nature of these images. For the moment we propose to implement an operator-assisted matching procedure. After the anatomical structures having been located in an image of the L.S.O. sequence (by using a sharpness of focus measure) the operator will locate the same structures in the stereo image. Based on the depth information from the L.S.O. images the stereo image can then be recomputed so that a more accurate 3D reconstruction of the retina can be obtained.

Conclusions

In this paper we described work currently under progress that aims at the 3D reconstruction of human retinas. Two modalities of images are employed: angiographic and laser-scan ophthalmoscopic. The specific nature of the images raises a number of difficulties because of the almost non-existent sharp boundaries. We show that some of these difficulties can be overcome by using algorithms that do not make explicit use of edges or other geometrical features.

Acknowledgements

The authors would like to thank JNICT for its financial support through Project PBIC/C/SAU/1576/92

References

[Batista *et al.*, 1986] J. Batista, J. Dias, H. Araujo, and A. T. Almeida. Monoplane Camera Calibration: An Iterative Multistep Approach. In J. Illingworth, editor, *British Machine Vision Conference 1993*, pages 479–488. BMVA, May 1986.

[Cox *et al.*, 1992] I.J. Cox, S. Hingorani, B.M. Maggs, and S.B. Rao. Stereo Without Regularization. Internal NEC Report, October 1992.

[Dias *et al.*, 1991] J. Dias, A. Almeida, and H. Araujo. Depth Recovery Using Active Focus In Robotics. In *International Workshop on Intelligent Robots and Systems 1991*, pages 249–255, 1991.

[Kim *et al.*, 1990] N. H. Kim, A.C. Bovik, and S.J. Aggarwal. Shape Description of Biological Objects via Stereo Light Microscopy. *IEEE Trans. on Systems, Man and Cybernetics*, 20(2):475–489, March/April 1990.

[Marr and Poggio, 1979] D. Marr and T. Poggio. A Computational Theory of Human Stereo Vision. *Proceedings of the Royal Society of London*, 204(Series B):301–328, 1979.

[Ohta and Kanade, 1985] Y. Ohta and T. Kanade. Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. *IEEE Trans. on Patt. Analysis and Man. Intelligence*, PAMI-7(2):139–154, March 1985.

[Pattipati *et al.*, 1990] K.R. Pattipati, S. Deb, Y. Bar-Shalom, and R. B. Wasburn. Passive Multisensor Data Association Using A New Relaxation Algorithm. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*, pages 219–246. Artech House, 1990.

[Tsai, 1987] R. Tsai. A Versatile Camera Calibration Technique For High-Accuracy 3d Machine Vision Metrology Using Off-the-Shelf TV Cameras And Lenses. *IEEE Journal of Robotics and Automation*, RA-3(4), August 1987.