DISPARITY ESTIMATION IN UNCALIBRATED STEREO RETINAL IMAGES

R. Kolář*, L. Kubečka*, J. Jan*, R. Chrástek**

*Department of Biomedical Engineering, FEEC, Brno University of Technology, Brno, Czech Republic

**Department of Computer Science, Friedrich-Alexander University of Erlangen-Nürnberg, Germany

kolarr@feec.vutbr.cz

Abstract: We present a new method for estimation of disparity in retinal stereo-photograph (topography of the retinal surface). It is based on the edge detection at various levels of resolution and combination of disparity map estimation. Some preprocessing and post postprocessing steps are also presented to improve the results. The images are compared with topographic images obtained from confocal scanning laser ophthalmoscope (CSLO).

Introduction

The 3D visualization of the retina, particularly the optical disc (OD), became very important diagnostic tool in glaucoma disease and its progression. The OD shape and some derived parameters are taken into consideration during the medical diagnosis process [9].

Nowadays, there are two main methods for obtaining the OD surface. One is based on a mono or stereo ophthalmoscope and the second method utilizes the confocal scanning laser ophthalmoscope (CSLO). In the next paragraphs, we consider only photographs taken by digital opthalmoscope (fundus camera) in *stereo mode*. The main advantage of this method is its low cost in comparison with CSLO (HRT – Heidelberg Retina Tomograph).

We present a parametric and multiresolution-based method for 3D optical disc reconstruction from uncalibrated stereo retinal images. The visual comparison with CSLO is made in the last paragraph.

Method

The photographs of the retinal surface were obtained by the digital ophthalmoscope (Nonmydriatic Fundus camera, KOWA, Nonmyd α-D [10]). This camera is not 'true' stereo camera, because it doesn't have a binocular head. But there is a 'stereo-mode' where stereo images may be taken. The procedure for obtaining these images requires cooperation with the patient:

- the patient is looking at point *P* situated at the left side - the first image is taken;
- point *P* is moved in horizontal direction to the right;
- the patient is looking at point *P* in new location - the second image is taken;

As a result, there are two stereo *color* images. It has been shown [1,2] that the green channel is sufficient when analysing these images. Another possibility, discussed in some articles, is to convert RGB image to the grayscale using all three channels. But we didn't find any noticeable differences in resulting retinal surface. We therefore omit the R and B channel in the next processing and use only green channel.

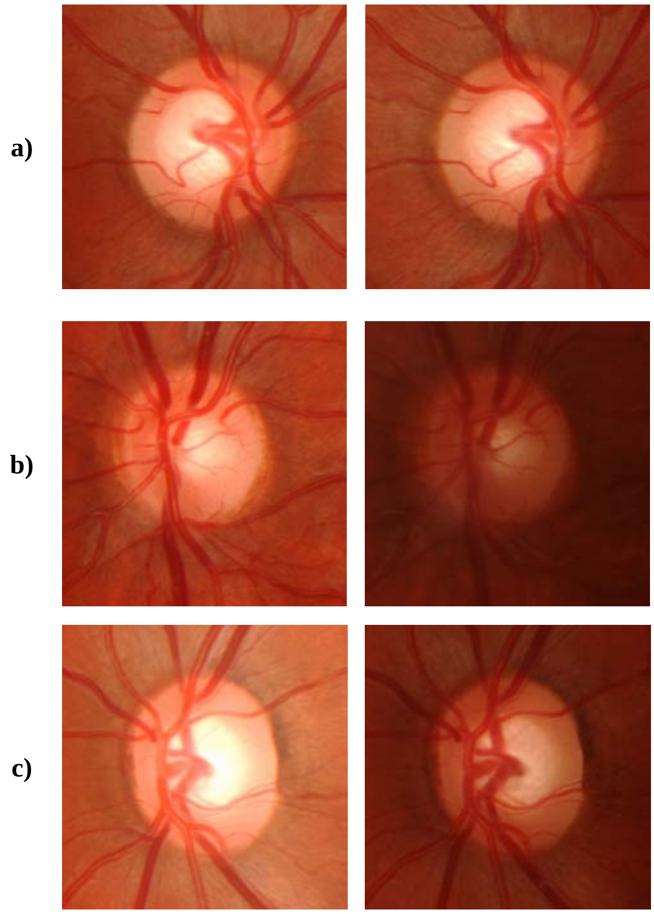


Figure 1: Samples of stereo-image pairs; a) good quality and b), c) are images of poor quality – different left-right image contrast, poor or over-illuminated image

As a good stereo pair is considered an image pair having a long horizontal displacement and a small vertical displacement [2] (see Figure 1 for examples of stereo-image pairs). It is textured enough and has a high contrast. But it is clear that during the acquisition process the patient can move his head and consequently the eye-shift is not only horizontal but also vertical and possibly rotational. To cope with this problem, *left – right* image registration method based on the mutual information is used [3].

To *register* two images means to find the spatial transformation aligning the contents of the images. There are many ways to do that, one of them is to maximize a similarity criterion using optimization techniques [5]. It can be formalized as

$$\alpha_0 = \arg\min_\alpha C(R, T_\alpha(F)), \qquad (1)$$

where R is the reference image and F is the floating image, which is transformed by spatial transformation T_α to the coordinates of the reference image. The registration quality, corresponding to the transform T, is evaluated by the criterion function C. $T_{\alpha 0}$ is the optimal registering transform with respect to the criterion. Affine transformation T, encompassing shift and rotation was found sufficient to compensate the misaligning among both images. The Powell's optimization method using series of line minimization was incorporated into this registration algorithm. For more details see [3, 4].

We have several stereo-image pairs taken by KOWA camera and many single images in our stereo database. There are big differences in image quality. Three samples of stereo-images are shown on Figure 1. Unequal retina illumination, different contrast within the images and over illuminated regions are the main problem that must be solved. These properties make the disparity map estimation complicated, because they abase the texture that is important in estimation process. Therefore, there is a need for preprocessing.

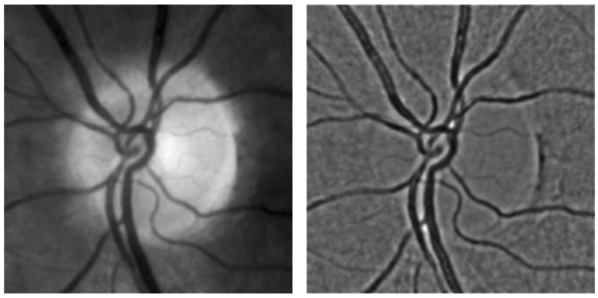
As a *preprocessing* operation, we've used following steps:
- defining the region around the optical disc;
- decimation;
- filtering by a lowpass filter;
- contrast correction;

The original fundus images from KOWA camera [10] have a high resolution (1200x1600px, resolution about 4.3μm/px) and a large field of view in comparison with CSLO. A sample of an original image is shown on Figure 2. The square regions of interest (680x680 pixels) were manually determined for each stereo-image pair. This subimage was then decimated with factor 2 to the size 340x340 pixels. This operation doesn't affect the result and speeds up the computation time while the disparity map is still dense. The resolution is decreased (about 9 μm/px), but it is comparable with the images from CSLO (10μm/px).

Figure 2: One original image

Lowpass filtering of these subimages was employed to reduce the noise artefacts and decrease the noise sensitivity during estimation process. Simple averaging filter of size *[5x5]* pixels was used.

The last preprocessing operation is contrast correction. We use the approach based on a model of nonuniform illumination (background model), which is generated by low-pass filtering of the image with filter mask *31×31* pixels to remove the vessels in the background. The background is to be subtracted from the original image. Since there is different contrast in lighter and darker areas of image, contrast correction based on the background model should be applied [4].


a) b)
Figure 3: a) image after lowpass filtering; b) image after contrast enhancement

The whole algorithm of the proposed approach is depicted on Figure 4. After preprocessing steps a SSD (sum of squared differences) based approach for disparity estimation is performed. The SSD is a measure of similarity between two regions:

$$ssd(d) = \sum_{(i,j)\in w} \left[I_{left}(i,j) - I_{right}(i,j+d) \right]^2, \qquad (1)$$

where d is a horizontal displacement between the square regions in left image I_{left} and right image I_{right}. We have to search for the smallest *ssd* value (d_{min}) in a predefined range ($-d_{max}$, d_{max}), where d_{max} denotes the maximum possible shift between regions.

Because the retina stereo fundus images have a texture of low contrast they are not convenient for direct processing by SSD method. Hence, we use *parametric*

ISSN: 1727-1983

images as the input for SSD disparity estimation. These are created by the edge detection at various resolution levels:

$$I_{edge} = \left(h_{LP} * I\right) * h_{edge} \, , \qquad (2)$$

where h_{LP}, h_{edge} are impulse responses of the low pass (LP) filter or the edge detector, respectively; * denotes a 2D convolution.

The *Gaussian pyramid-like* structure was used for the resolution setting. The highest resolution is for $h_{LP}=1$. Lower resolution levels are achieved by iterative convolution with LP filter of *5x5* pixels. We have used the strategy inspired by [8], but no subsampling was performed. The filter h_{LP} used for the convolution (2) is:

$$h_{LP} = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} / 256 \, ,$$

which is *broader* than Gaussian. This corresponds to the coefficient $a=0.375$ according to [8].

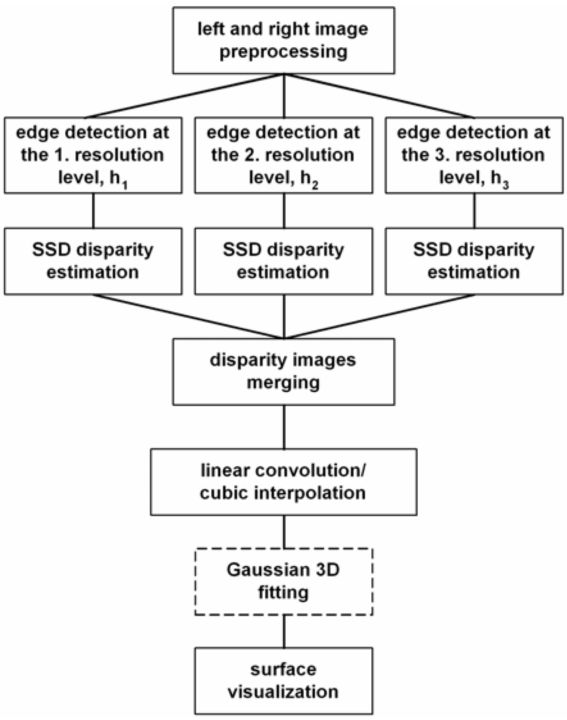


Figure 4: Flow chart of the proposed approach.

The Prewitt operators are used as the edge filters:

$$h^{hor}{}_{edge} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad h^{ver}{}_{edge} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

Because the convolution is a linear operation, we first apply the convolution between the impulse responses $h_{edge} * h_{LP}$ and resultant impulse response is applied on the image *I*. Thus we get three impulse responses h_1, h_2, h_3 each for vertical and horizontal direction:

$$h_1^{ver} = h^{ver}{}_{edge} \qquad h_2^{ver} = h^{ver}{}_{edge} * h_{LP}$$

$$h_1^{hor} = h^{hor}{}_{edge} \qquad h_2^{hor} = h^{hor}{}_{edge} * h_{LP}$$

$$h_3^{ver} = h^{ver}{}_{edge} * h_{LP} * h_{LP}$$

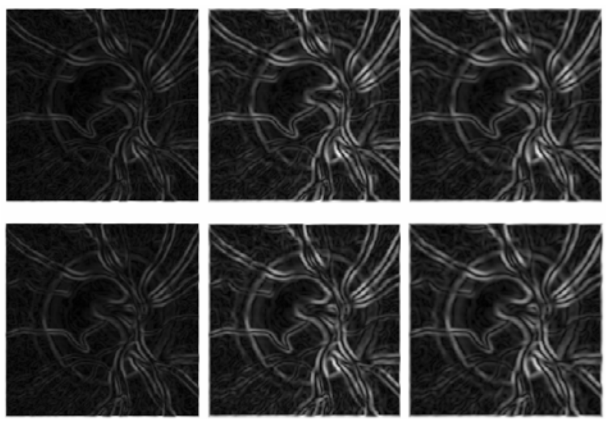
$$h_3^{hor} = h^{hor}{}_{edge} * h_{LP} * h_{LP}$$

The square root from squared horizontal and vertical filters is computed and used for the next processing.

We can see a similarity with the undecimated discrete wavelet transform (UDWT), defined by Mallat [7]. This UDWT was also tested with various kinds of filters and we have obtained similar results only for Haar filter/wavelet, but no improving in estimated disparity was observed, even with more complicated wavelets.

The edges detected at these three levels are shown on Figure 5. The strong edges, from blood vessel, as well as weak edges, arising from texture, are visible. These are important in SSD disparity estimation.

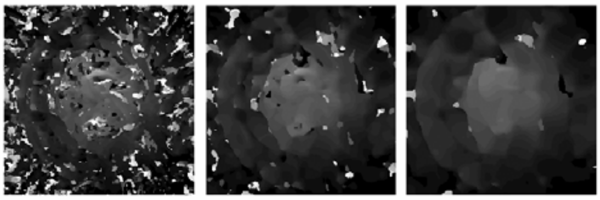
Once these images are obtained, the SSD matching can be applied. The corresponding disparity maps are shown on Figure. 6. The white colour corresponds to the maximum displacement/disparity ($d=d_{max}=40$ pixels) and black colour to $d=0$. These maps are used for creating the result disparity map. We have observed that satisfactory result can be obtained with two merging techniques (Figure 4). The first technique computes the mean value from these three disparities. The second technique takes only the minimum value from these three disparities, which is perhaps more straightforward because larger disparities are probably wrong estimates. We don't compare the quality of these two techniques here and we will only use the second technique (see Figure. 7a)



1. level 2. level 3. level

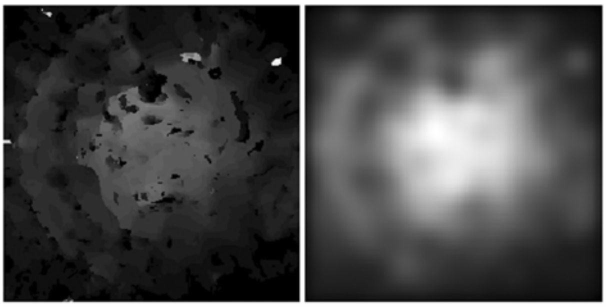
Figure 5: Edges detected at three resolution level for left image (first row) and right image (second row). The original stereo image pair is shown on Figure 1a

After this disparity estimation, it is necessary to apply some postprocessing operations to eliminate disparity errors. We suppose that there are no partially occluded points [6], because of the smooth surface and small angle shift between the stereo views [2]. Thus, sharp changes and edges in estimated disparity can be considered errors in disparity estimation process. To eliminate these changes, we employ a simple averaging filter with equal weights, size *[15x15]* pixels. This filter is applied three times, which is identical to cubic interpolation [2]. The result is shown on Figure 7b.



1. level 2. level 3. level

Figure 6: Disparity images for the three resolution levels.



a) b)

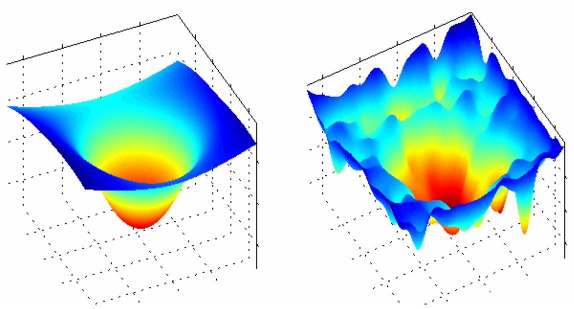
Figure 7: a) Disparity map obtained as a minimum of disparities on Figure 5; b) Smoothed disparity by linear convolution filter (cubic interpolation)

Although the preprocessing should eliminate the dependence of the result on stereo image quality, there is still quite strong dependence. Therefore, we should employ another technique to increase the result quality. It is based on *a priori* modelling of the shape of the OD. This shape can be modelled with the help of 2D Gaussian function *G(x,y)*. More information about can be found in [9]. Here we apply the Gaussian shape fitting algorithm to the resulted disparity map using Powell optimization algorithm. The criterion for minimization is a *standard* sum of squared differences, where the differences between model and disparity map are used. The fitted Gaussian is shown on Figure 8a while corresponding 3D representation of the disparity map is on Figure 8b. It has been shown that some parameters from fitted Gaussian shape can be used for glaucoma diagnosis support [9].

A heuristic approach was used to smooth the disparity map. This modification lies in simple, *pixel-by-pixel*, multiplication:

$$D_{result}(x,y) = D(x,y) \cdot G_{norm}(x,y), \qquad (3)$$

where $G_{norm}(x,y)$ is fitted Gaussian function and $D(x,y)$ is estimated disparity image. This operation can be considered as a kind of filtering emphasizing the OD shape and decreasing the small changes in surroundings.



a) b)

Figure 8: a) Result of 2D Gaussian fitting; b) Estimated disparity from Figure 7b, displayed in 3D space

Results and Conclusion

It is not easy to evaluate the resulting disparity maps. Because our database has corresponding images from CSLO, we can present visual comparison with the topographic images. Figure 9 compares the disparity image (a) and topographic image obtained by CSLO (b). Each image has another resolution, but strong correlation in the shapes can be seen. Figure 10 shows the same content, but for another patient.

It has to be emphasized, that in spite of the preprocessing and postprocessing operations, the result greatly depends on the quality of stereo images and on the cooperation between the patient and the physician. A more advance technique should be used to eliminate this dependence (phase-based or maximum flow techniques are promising).

Acknowledgment

This work has been supported by the grant no. 102/03/P153 of Czech Science Foundation and by the Research center no. 1M6798555601 under Ministry of Education (Czech Republic).

The authors would like to thank to R. Chrástek from Department of Computer Science and G. Michelson from Department of Ophthalmology, Friedrich-Alexander University (Germany) for their valuable comments and for measuring and providing the image data.

References

[1] AWAWDEH A., FAN G. (2003): "Pseudo Cepstrum for Assessing Stereo Quality of Retinal Images", Proc. of the 37th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, pp.1953 - 1957.

[2] XU J., CHUTATAPE O. (2004): "3D Optic Disc Reconstruction from Low-Resolution Stereo Images Based on Combined Registration and Sub-

a)

b)

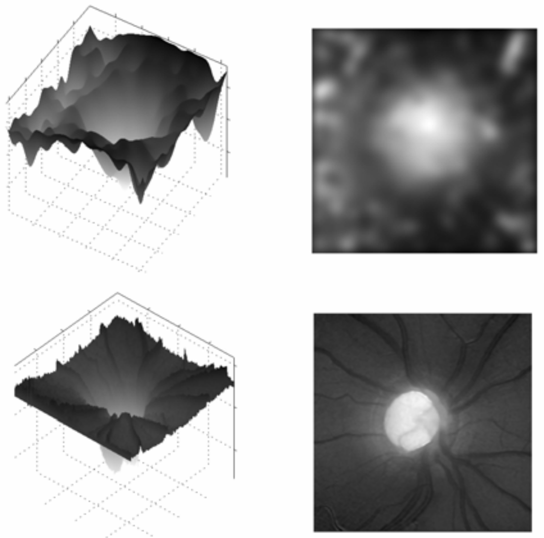


Figure 9: Patient no.1: a) Resulted disparity image; b) Topographic image obtained by CSLO

a)

b)

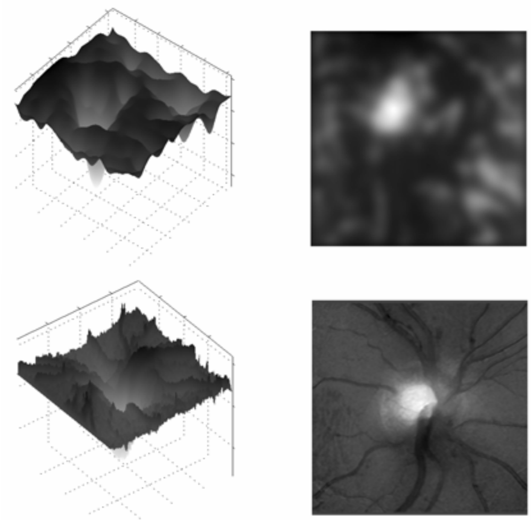


Figure 10: Patient no.2: a) Resulted disparity image; b) Topographic image obtained by CSLO

Pixel Matching”, Proc. of the 26th the IEEE EMBS, USA, pp.1825-1828

[3] KUBEČKA L., JAN,J. (2004): “Method for Registration of Multimodal Images of Human Retina”, Proc. of the 17th EURASIP Conf. Biosignal 2004 pp.255–257

[4] KUBEČKA L., SKOKAN M., JAN J. (2003): “Optimization Methods for Registration of Multimodal Images of Retina”, Proc. IEEE-EMBS, Cancún, Mexico, pp. 599 - 602

[5] PLUIM J.P., MAINTZ J.B.A., VIERGEVER M.A. (2003): "Mutual information-based registration of medical images: A survey", IEEE Trans. Med. Imag., vol.**22**, no.**8**, pp.986-1004

[6] SCHARSTEIN D. (1999): “View Synthesis Using Stereo Vision”, Lecture Notes in Computer Science, Springer

[7] MALLAT, S. (1999): ”A Wavelet Tour on Signal Processing”, Academic Press, 1999

[8] BURT, J.P., ADELSON, E.H. (1983): “The Laplacian Pyramid as a Compact Image Code”, IEEE Trans. Communication, COM-3l, No. **4**, pp. 532 - 540

[9] SWINDALE, N.V., STJEPANOVIC, G. CHIN, A., MIKELBERG, F.S. (2000): “Automated Analysis of Normal and Glaucomatous Optic Nerve Head Topography Images”, Onvestigative Ophthalmology and Visual Science, No. **7**, Vol. **41**, pp. 1730 - 1742

[10] KOWA, Internet Site address: http://www.kowa-usa.com/medical

ISSN: 1727-1983 © 2005 IFMBE