All-in-Focus Imaging Technique Used to Improve 3D Retinal Fundus Image Reconstruction

Danilo Motta
R&D - Wavetek
Universidade de São Paulo
São Carlos, SP - Brazil
daniloam@icmc.usp.br

Luciana de Matos
R&D - Wavetek
Universidade de São Paulo
São Carlos, SP - Brazil
lucianadematos-@gmail.com

Amanda Caniatto de Souza
R&D - Wavetek
São Carlos, SP - Brazil
amanda@wavetek.com.br

Rafael Marcato
R&D - Wavetek
São Carlos, SP - Brazil
rafaelmarcato-@wavetek.com.br

Afonso Paiva
Universidade de São Paulo
São Carlos, SP - Brazil
apneto@icmc.usp.br

Luis Alberto Vieira de Carvalho
R&D - Wavetek
São Carlos, SP - Brazil
luisalberto@wavetek.com.br

ABSTRACT

In this paper we have applied the stacking technique on images from a medical device. There is an urgent need in ophthalmology for more cost effective instrumentation for the early diagnosis of glaucoma, one of the leading diseases in the cause of blindness worldwide. The current techniques involve expensive optical equipments generally called fundus camera, which most of the time capture a single high resolution frame for one eye at a time. In this research we have used stereoscopic videos of a state-of-the-art 3D retinal camera, which has simpler optics and electronics when compared to current monoscopic models. Nevertheless, the cross correlation algorithms for depth computation are very sensible to image noise and out of focus regions. We demonstrate the efficiency of our technique on experiments involving a sequence of images extracted from videos in simulations of optic nerves using artificial objects.

Categories and Subject Descriptors

D.2.8 [**Computer Applications**]: Life and Medical Sciences—*Medical information systems*

Keywords

Optic Nerve, Stacking technique, image processing

1. INTRODUCTION

According to the World Health Organization there are approximately 285 million people in the world with some kind of vision problem [21]. The estimate is that by the year 2020 this number will double [24]. There are many factors which influence this estimate, among them the fact that world population is growing older, the lack of early diagnosis and treatment of many eye diseases, such as cataract and glaucoma, among others. Among the diseases related to the posterior segment, glaucoma is one of the leading causes of blindness in the elderly population in the world. There are many factors that may be used to explain why glaucoma is such an aggressive pathology, such as: far and foremost the fact that glaucoma in known to be a very hard pathology to detect in early stage [12], lack of experience in diagnosis by the eye care professional, lack of good and precise instrumentation. The Gold Standard procedure for glaucoma diagnosis today is based on a series of exams. The current diagnosis machines can be extremely expensive and involve a multitude of techniques such as HD fundus imaging [4], visual perimetry tests [16], eye pressure measurement using tonometers [3], among others. One of the most sensitive techniques among all these diagnostic tools is the analysis of the changes in the retinal disc (also known as optic nerve). The optic nerve of a glaucoma patient undergoes changes that most times are not perceivable just by looking at retinal images or the other exam reports mentioned above. The ideal technique for detecting these changes are those that may render accurate three dimensional information of the optic nerve head, such as OCT (Optical Coherence Tomography) [8, 28] and flicker techniques applied to HD retinal images. Nevertheless both these state-of-the-art techniques, although reasonably available in developed countries, are still considered extremely expensive in developing countries. In order to make other techniques popular, there is a need for more quantitative techniques and also for instrumentation that are not considered expensive in developing countries. With these motivations in mind, in this research we pursue to create a 3D retinal camera based method that is both suitable for quantitative glaucoma analysis and at the same time much less expensive than OCT or other systems currently available. Nevertheless the images obtained by this new instrument, although in high quality, have a short depth of focus. This fact alone is causing imprecision in the cross correlation algorithm for depth computation [5] and therefore rendering the technique quite inefficient since the device should have sub-pixel precision in order to be sensible for optic nerve depth diffe-

SAC'15 April 13-17, 2015, Salamanca, Spain.

http://dx.doi.org/10.1145/2695664.2695845

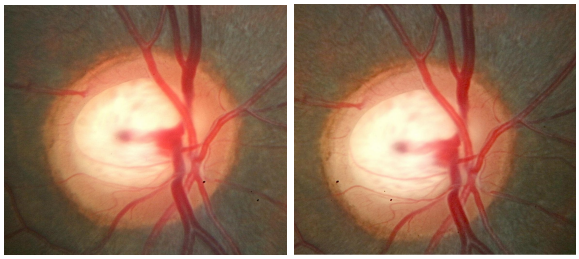

Figure 1: Fundus image pair.

rences for the same patient. In order to solve this problem in this paper we show a technique known generally as "focal stacking" to produce all-in-focus images that was developed and applied to the stereo retinal images of artificial models. This technique is a tool that has gained much interest for industrial and biological applications, such as photography for digital cameras [32, 14, 22, 30, 11, 15], and microscopic images [9, 1]. To our knowledge the application of all-in-focus imaging techniques in retinal images is also new and pioneered by our group. We compare the results obtained with and without our stacking technique and show the improvements obtained. We believe that, from the results obtained here, that the current application is more efficient and may improve considerably the precision of depth computation in retinal nerve images.

2. FUNDUS 3D RECONSTRUCTION

Although several techniques have been proposed in the computational stereo for extraction of three-dimensional structures (see [5, 26]), few studies have focused on 3D optic disc reconstruction. Algazi et al. [2] described two techniques to assess quantitatively the optic disc: in the first one, they used a sequential monocular display in which the two images are sequentially flickered, and in the second one they used stereophotogrammetric digital image processing. However, using computerized techniques to gauge subtle changes on Optical Nerve Head (ONH) are more efficient. In this sense, some works have been carried out [6, 31, 29, 18]. Nakagawa et al. [18] realized an automatic method for the reconstruction of the 3D structure of the ONH using the stereo retinal fundus image pair. The technique applied in our work is based on Nakagawa's method.

The first step of this technique is the acquisition of the stereoscopic fundus images. An example of pair of fundus images can be seen in Figure 1. Our camera system obtains synchronous stereoscopic images of the papillary region as explained in [17]. The fact that image capture is implemented by two different CCDs at the same time avoids other problems such as patient or camera movements, which may compromise the quality of the stereo pair.

These images are the result of 2D image mapped from a 3D scene, and in this mapping process several points of the scene are related to the same pixel of the image. Stereoscopic image techniques can be used to recover the information of the depth lost during the mapping process of the image.

2.1 Movement stabilization

In the general reconstruction processes, it is necessary firstly to implement a pre-processing step to adjust alignment differences between the two images pair [25]. In the current work fortunately this process is not necessary due to our double CCD and optical system. The one concern must be the time consistency between the image pairs with varying focus. We must then align the image pairs to avoid noise from motor or patience movements over time. To do this we apply a simple technique developed for this research. First, we convert one of the image pairs to the CIELAB color space [13]. With this we accomplish to represent the image in a space close to human vision, and retrieve information about the lightness a clear feature of the ONH. We simply search for a circular region where the lightness is greater then it's Otsu threshold [20]. With the bounding box of the ONH found, they all are simply stabilized by shifting the images to bring the center of it's ONH bounding box to the mean center of all found bounding boxes.

Our experiments in Section 4.1 show that this alignment process is not expensive computationally and reliable.

2.2 All-in-focus

By combining images captured with different camera settings, one can create a higher quality all-in-focus image. This is what all-in focus techniques accomplish, and for this propose, we used the study made by Pertuz et. al [23] as it better adapts to our problem. In Pertuz study is defined a metric to find which pixels of the image had a better focus based on the difference of the neighborhood of the pixel and its mean. This measure is associated with a selectivity measure that compute the estimated relevance of the image features. Then the composition of these two measures is used to select the image regions with better focus. The last step is a selective fusion of the images based on the gathered information. For more details we refer to [23].

The general process to perform an exam with retinography equipments involve three steps: (1) image acquisition, (2) image analyses and (3) diagnosis. In traditional retinal cameras both steps (1) and (2) is entirely done in a subjective manner and depends on issues such as physician's experience, type of disease, among others. Even in recent instruments which have almost automatic image capture and analyses modules, the physician still needs to subjectively choose different retinal regions manually. With the current approach presented here these 2 steps are automatic and therefore remove the interaction in choosing best focus and image analysis. The technique present here may reduce human errors in retinal exams and improve ophthalmological diagnosis of the retina. The physician only needs to centralize the optical disc in the monitor and the equipment can take a film of the eye fundus with varying focus, create the all-in-focus image and proceed as usual through the other steps of the exam.

2.3 Stereoscopic correspondence and Reconstruction

Given two images of the same scene from two different points of view, we must find correspondent features in both. For each pixel in an image the process of stereoscopic correspondence will try to find out the correspondent pixel in the stereo pair. For this task and the reconstruction we use the similarity metric and depth calculation presented in [18] but with changes in the image sampling to improve performance as explained in Section 2.5. This reconstruction method is advisable because it does not depends of the several parameters, solving the same problem stylishly.

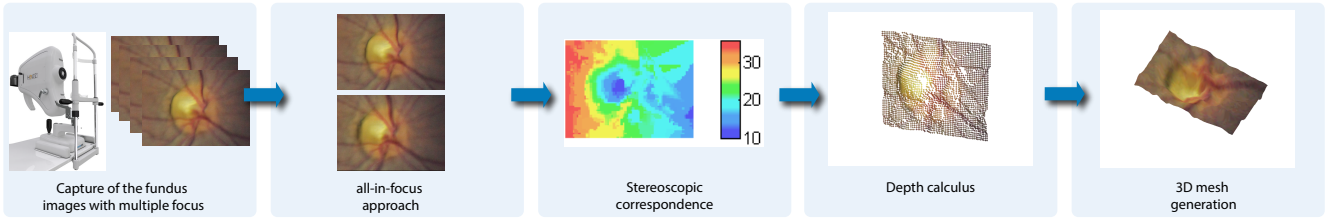

Figure 2: Main steps of the pipeline.

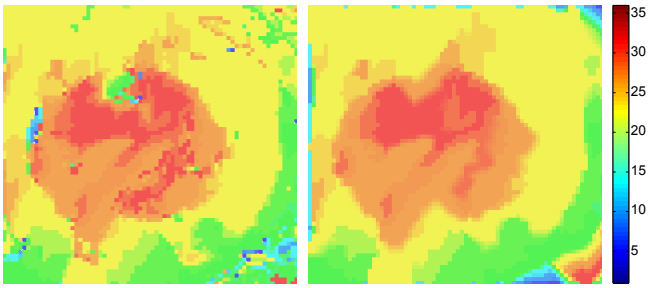

Figure 3: Denoising process: original data (left) and after filtering step (right). The colors code the disparity map.

2.4 Noise removal

Noise is ubiquitous in the process of cross-correlation in stereoscopic pairs [27, 6]. In our framework, we use the same density-based denoising method proposed by Motta et al. [17]. This method detects outliers in the disparity map using DBSCAN clustering algorithm [7].

However, the denoising step can produce unwated holes in the disparity map. In order to avoid this problem, we smooth the data using the technique proposed by Garcia [10], which deals gracefully with the missing data while penalizing any remaining noise. Figure 3 shows our denoising process in action.

2.5 Sampling

The traditional framework of Nakagawa et al. [18] increases considerably the resolution of the images by oversampling, consequently increasing the number of depth points and the computational cost. In our framework, we do not perform this image oversampling in order to obtain a viable computational method to attend the ophthalmology protocol. This improvement allows our framework to execute in interactive time.

3. OUR PIPELINE

Our pipeline is simple, as seen in Figure 2, a sequence of fundus images are captured by the equipment varying the focus and avoiding image displacements, however we apply an movement stabilization step to ensure time consistency between the pairs, as explained in Section 2.1. Then the all-in-focus technique is applied to compose a single pair of fundus images. The disparity map is created and we remove the noise as explained in Section 2.4. From the new map, depth is calculated by triangulation and interpolated to create the mesh. Finally, texture is added to the mesh to form our 3D reconstruction.

4. EXPERIMENTS AND RESULTS

With the next experiments we demonstrate the robustness of the steps of our method. To create a dataset with a measured ground truth we modeled three objects in plastic material. Each of this objects have different excavations size, they can be seen in descending order of depth in Figure 5. We refer to the objects in this order by model 1, 2 and 3 respectively. On the objects is hand-painted ONH representations so we can use our reconstruction algorithm. From these objects we used a caliper rule to measure a horizontal cut in each one and we compare this curve with a horizontal cut of our reconstruction result in Section 4.3. These models do not perfectly represent a retinal disk, however we will try to use then to prove that the reconstruction can detect the changes in the objects excavations. If the excavation change is clear, then is expected that this behavior will be the same for the retinal disk. We tried to make images as close as possible with the images taken during actual practice, at least in the region of interest, removing illumination problems as light spots.

In our experiments, we use images with resolution of 640×480 pixels, the cameras are separated by 120 millimeters with angle of $87°$.

4.1 Movement stabilization

In order to test our movement stabilization algorithm, we must validate if the proposed approach finds the region of interest (ONH in our case) in real world datasets. Figure 4 shows the result of this approach for high (top two pairs) and low (bottom down pairs) resolution images. The images contain noise and different color scales, however, in all cases the regions of interest are found. As explained before, our pairs will be already aligned due to the equipment optics characteristics, but as we use a sequence of pairs to do the all-in-focus, patient and motor movement can create differences between the sequence of pair images. With this method we remove this differences and avoid noise in the all-in-focus image.

4.2 All-in-focus

Our next experiment compares the results of the all-in-focus technique applied to our context against user selected images. We used an average of 30 images for each image generation. The parameters used in these test were $\alpha = 0.001$, $S_{th} = 13$ and the size of the focus measure window was 36 in the notation of [23]. In Figure 6, the top two images are the pair with better focus from our data, chosen by the user and on bottom are the pair automatically generated by the all-in-focus technique. The images are very similar, but in the white border of the models, the bottom images have better focus.

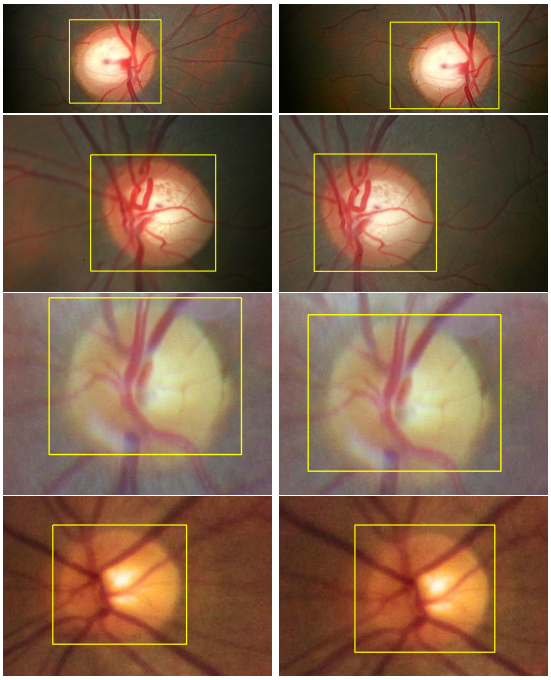

Figure 4: Image alignment tests in stereo pairs.

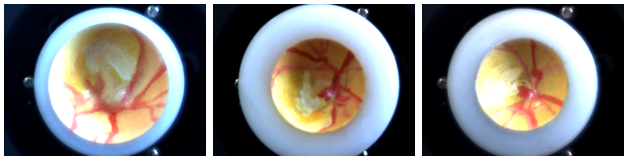

Figure 5: Experiment models.

4.3 Reconstruction

Our main experiment is to reconstruct the three models from stereo pairs and compare it to the real measurements of the models. Two reconstruction can be seen in Figure 7, first with user selected images with better focus and second using the all-in-focus automatic generated images. The form of the models are reproduced with some noise in the border of the surfaces due to occlusion. We proceed then to compare the results with the measures taken of the models. We compute error measures in the horizontal cuts and a global dimensionless metric defined as the volume of the ground truth volume divided by the volume of the reconstruction (close to 1 is better). The quality of the excavation of the all-in-focus reconstruction is preserved as it don't behaves differently from the user selected images. The volume metric is especially informative as it indicates that both reconstructions present good representation of the volumes, however, the all-in-focus reconstruction produced better results in this case. As can been seen in Tables 1 and Figure 8, the errors of the both reconstructions are similar. As at the first approach user interaction was needed, we accomplish similar or better results with the use of all-in-focus, our goal with this study.

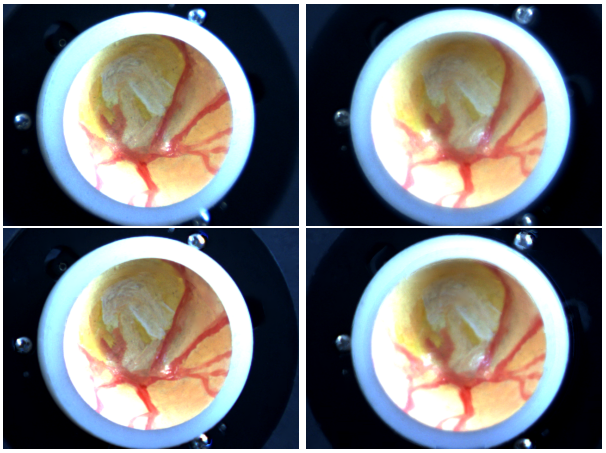

Figure 6: Two user selected images (top) and its correspondent all-in-focus automatic generated images (bottom).

5. DISCUSSION

The stacking technique is not new in other areas of image processing [19], but to our knowledge it has not been widespread in the medical area. There are some applications in biology and microscopy [9] but we have found very few examples for live *in vivo* images. In this work we have found that the depth computation for retinal nerve images is a very sensible technique, which can be easily affected by noise and therefore spoil promising results. Because of that one of the goals was to discover ways in which a series of images, even better if taken for slightly different focus positions, could be used to come up with a single best image.

6. CONCLUSIONS

The 3D reconstruction technique presented here may serve as a prototype for a growing market which is demanding faster, simpler, easier to operate and more cost effective retinal cameras. The main factor that makes these devices cheaper is that they do not require all the state-of-the-art and gold standard optics for angiography and fluoroscopy, and red-free that top retinal cameras usually incorporate. The trade-off is that, because of the simpler optics, these devices will capture good quality images only at the central retina, as opposed to devices that may capture 80% of the retina (`www.optos.com`). The technique presented allow 3D imaging without the use of special goggles, is easier to use and operate in certain conditions when limited physical space and time are required. The main application of this approach is in screening examinations and progression/treatment monitoring of glaucoma, and may also be applied to the diagnosis of early vision loss in diabetic retinopathy, age related macular degeneration and other retinal pathologies, therefore adding value to current state-of-the-art retinal cameras as a low-cost preliminary examination instrument. Our next goal is to apply these same algorithms on *in vivo* eyes in collaboration with a physician to evaluate the impact of our technique in medical diagnosis.

6.1 Acknowledgments

The authors are funded by CAPES, CNPq (#382234/2014-6 and #350232/2013-0) and FAPESP (#2014/09546-9).

Table 1: Reconstruction absolute errors, in mm, except for the dimensionless volume metric.

(a) user selected images

Model	Mean error	Maximum error	Standard deviation	Peaks error	Volume metric
1	2.8689	8.3554	2.9095	1.1359	0.9568
2	2.4205	6.9620	2.7142	0.0427	0.7984
3	1.5074	3.2711	1.0529	1.0285	0.7695

(b) all-in-focus automatic generated images

Model	Mean error	Maximum error	Standard deviation	Peaks error	Volume metric
1	2.4557	7.9311	2.6162	1.3615	0.9750
2	1.2791	4.3066	1.5282	0.3993	0.8508
3	1.4461	3.7170	1.0243	1.5002	0.8343

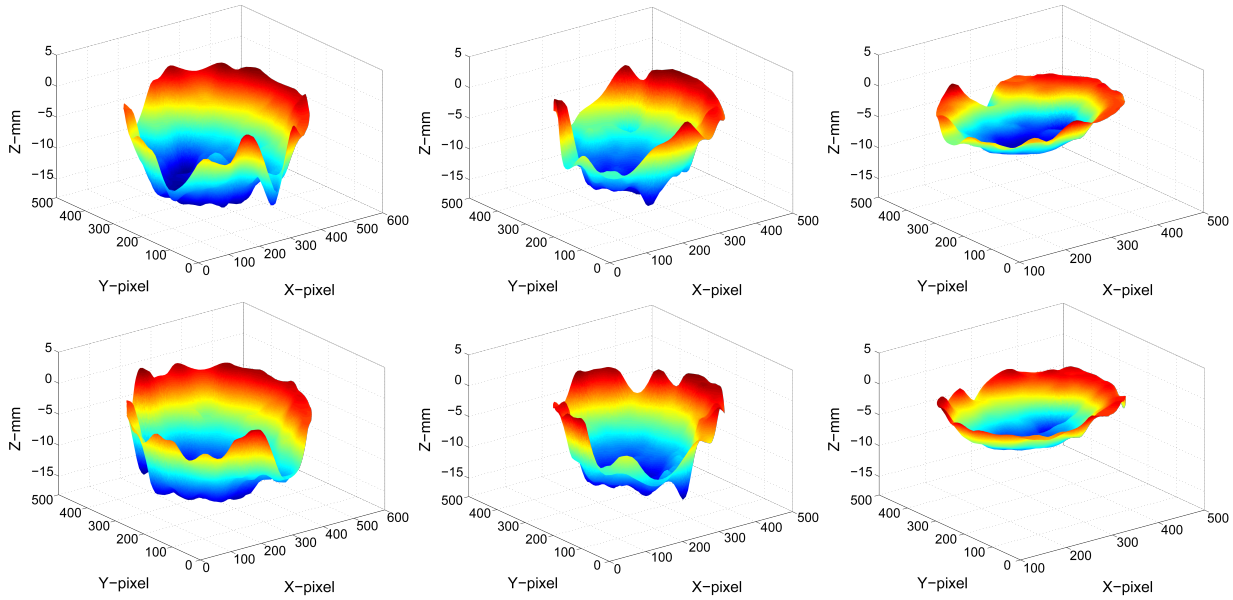

Figure 7: Top row is the reconstruction with the user selected images, on the bottom row is the all-in-focus automatic images reconstruction.

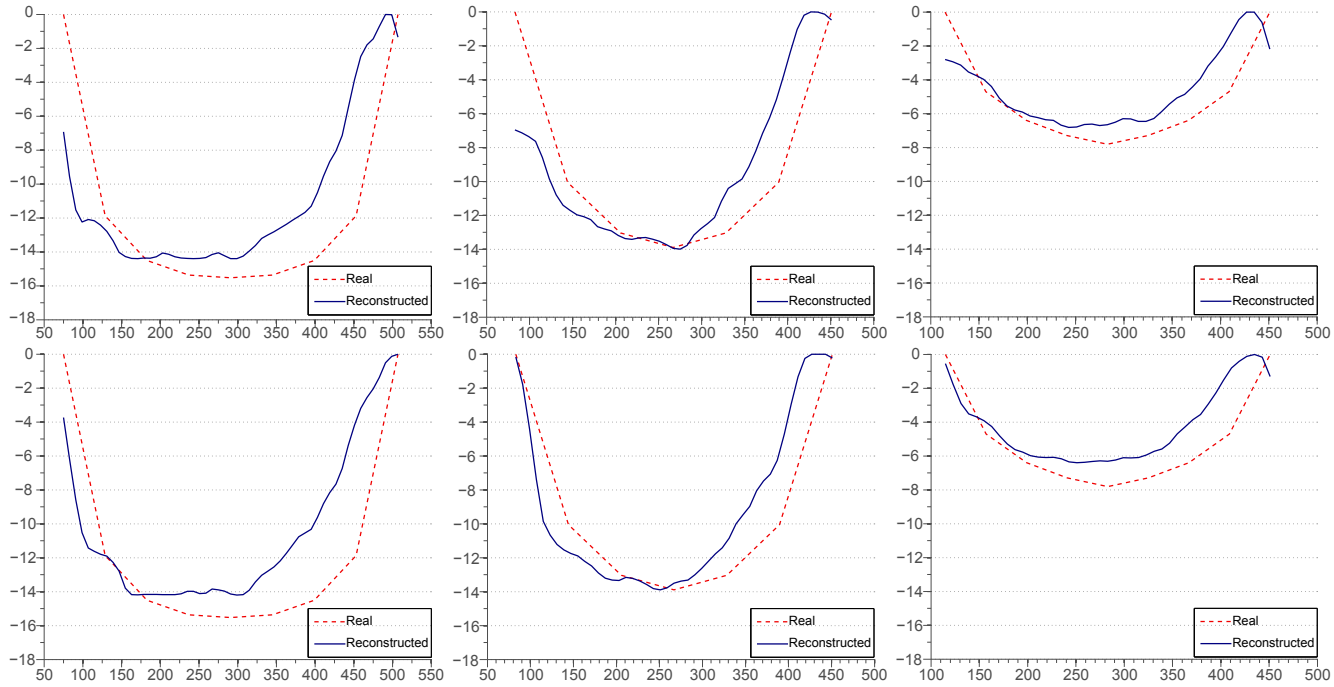

Figure 8: Top row is the reconstruction with the user selected images, on the bottom row is the all-in-focus automatic images reconstruction.

7. REFERENCES

[1] F. Aguet, D. Van De Ville, and M. Unser. Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy. *IEEE Trans. Image Process.*, 17(7):1144–1153, 2008.

[2] V. Algazi, J. Keltner, and C. Johnson. Computer analysis of the optic cup in glaucoma. *Invest. Ophthalmol. Vis. Sci.*, 26(12):1759–1770, 1985.

[3] L. Barleon, E. M. Hoffmann, M. Berres, N. Pfeiffer, and F. H. Grus. Comparison of dynamic contour tonometry and goldmann applanation tonometry in glaucoma patients and healthy subjects. *AJO*, 142(4):583–590, 2006.

[4] R. Bernardes, P. Serranho, and C. Lobo. Digital ocular fundus imaging: a review. *Ophthalmologica*, 226(4):161–181, 2011.

[5] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intell*, 25(8):993–1008, 2003.

[6] E. Corona, S. Mitra, M. Wilson, T. Krile, Y. H. Kwon, and P. Soliz. Digital stereo image analyzer for generating automated 3-d measures of optic disc deformation in glaucoma. *IEEE Trans. Med. Imag.*, 21(10):1244–1253, 2002.

[7] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[8] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser. Optical coherence tomography-principles and applications. *Rep. Prog. Phys.*, 66(2):239, 2003.

[9] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser. Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images. *Microsc. Res. Techniq.*, 65(1-2):33–42, 2004.

[10] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.*, 54(4):1167–1178, 2010.

[11] H. Han, J. Jeong, and E. Arai. Virtual out of focus with single image to enhance 3d perception. In *3DTV-CON*, pages 1–4, 2011.

[12] D. C. Hood, A. Slobodnick, A. S. Raza, C. G. De Moraes, C. C. Teng, and R. Ritch. Early glaucoma involves both deep local, and shallow widespread, retinal nerve fiber damage of the macular region. *IOVS*, 55(2):632–649, 2014.

[13] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[14] K. N. Kutulakos and S. W. Hasinoff. Focal stack photography: High-performance photography with a conventional camera. In *IAPR*, 2009.

[15] X. Lin, J. Suo, G. Wetzstein, Q. Dai, and R. Raskar. Coded focal stack photography. In *IEEE ICCP*, pages 1–9, 2013.

[16] S. Liu, S. Lam, R. N. Weinreb, C. Ye, C. Y. Cheung, G. Lai, D. S.-C. Lam, and C. K.-S. Leung. Comparison of standard automated perimetry, frequency-doubling technology perimetry, and short-wavelength automated perimetry for detection of glaucoma. *IOVS*, 52(10):7325–7331, 2011.

[17] D. A. Motta, A. Serillo, L. de Matos, F. M. Yasuoka, V. S. Bagnato, and L. A. Carvalho. 3d papillary image capturing by the stereo fundus camera system for clinical diagnosis on retina and optic nerve. In *SPIE BiOS*, pages 893614–893614, 2014.

[18] T. Nakagawa, T. Suzuki, Y. Hayashi, Y. Mizukusa, Y. Hatanaka, K. Ishida, T. Hara, H. Fujita, and T. Yamamoto. Quantitative depth analysis of optic nerve head using stereo retinal fundus image pair. *J. Biomed. Opt.*, 13(6):064026–064026, 2008.

[19] K. Ohba, J. C. P. Ortega, K. Tanie, G. Rin, R. Dangi, Y. Takei, T. Kaneko, and N. Kawahara. Real-time micro-environmental observation with virtual reality. In *ICPR*, volume 4, pages 487–490, 2000.

[20] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[21] D. Pascolini and S. P. Mariotti. Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.*, page 96:614–618, 2011.

[22] F. Perez Nava and J. Luke. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In *3DTV-CON, 2009*, pages 1–4, 2009.

[23] S. Pertuz, D. Puig, M. A. Garcia, and A. Fusiello. Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images. *IEEE Trans. Image Process.*, 22(3):1242–1251, 2013.

[24] H. A. Quigley and A. T. Broman. The number of people with glaucoma worldwide in 2010 and 2020. *Br. J. Ophthalmol.*, 90(3):262–267, 2006.

[25] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos. Registration of stereo and temporal images of the retina. *IEEE Trans. Med. Imag.*, 18(5):404–418, 1999.

[26] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *I. J. Comput. Vision*, 47(1-3):7–42, 2002.

[27] P. W. Smith and N. Nandhakumar. An improved power cepstrum based stereo correspondence method for textured scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(3):338–348, 1996.

[28] V. J. Srinivasan, M. Wojtkowski, A. J. Witkin, J. S. Duker, T. H. Ko, M. Carvalho, J. S. Schuman, A. Kowalczyk, and J. G. Fujimoto. High-definition and 3-dimensional imaging of macular pathologies with high-speed ultrahigh-resolution optical coherence tomography. *Ophthalmology*, 113(11):2054–2065, 2006.

[29] L. Tang, Y. H. Kwon, W. L. Alward, E. C. Greenlee, K. Lee, M. K. Garvin, and M. D. Abràmoff. 3d reconstruction of the optic nerve head using stereo fundus images for computer-aided diagnosis of glaucoma. In *SPIE Medical Imaging*, pages 76243D–76243D, 2010.

[30] D. Vaquero, N. Gelfand, M. Tico, K. Pulli, and M. Turk. Generalized autofocus. In *IEEE WACV*, pages 511–518, 2011.

[31] J. Xu and O. Chutatape. Auto-adjusted 3-d optic disk viewing from low-resolution stereo fundus image. *Comput. Biol. Med.*, 36(9):921–940, 2006.

[32] Z. Zhang and R. S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proc. of the IEEE*, 87(8):1315–1326, 1999.