3D Reconstruction of the Retinal Surface for Robot-Assisted Eye Surgery

M. Havlena[1], K.K. Maninis[1], D. Bouget[2], E. Vander Poorten[2], L. Van Gool[1]

[1]*Computer Vision Laboratory, D-ITET, ETH Zürich, 8092 Zürich, Switzerland*

[2]*Department of Mechanical Engineering, University of Leuven, 3001 Heverlee, Belgium*
michal.havlena@vision.ee.ethz.ch

INTRODUCTION

Robot-assisted eye surgery is the central topic of the EU funded project EurEyeCase. Major objectives of the project comprise of the development of methodologies to perform two surgical procedures that cannot be easily carried out by sole human surgeons, namely retinal vein cannulation and retinal membrane peeling. In the proposed assistive system, the surface of the retina is modeled prior to the operation using image data from the available stereo microscope.

Pre-operative 3D reconstruction aims to provide the surgeons with a detailed 3D mesh model of the surface of the retina to facilitate their pre-operative planning. Namely, the visualization of the detected vessel structure and/or the optic disc in 3D gives the surgeons a unique opportunity to examine the structure of the eye from novel viewpoints and to plan the path of the tool with a higher confidence.

MATERIALS AND METHODS

The proposed method relies on the availability of stereo image data, which is in our case provided by a stereo microscope. A Haag-Streit ophthalmologic surgical microscope augmented by a stereo vertiscope with two C-mounted Allied Vision Prosilica GC 1290 cameras is able to acquire synchronized stereo videos having a 1280x960 resolution at 25 fps, see Fig. 1.

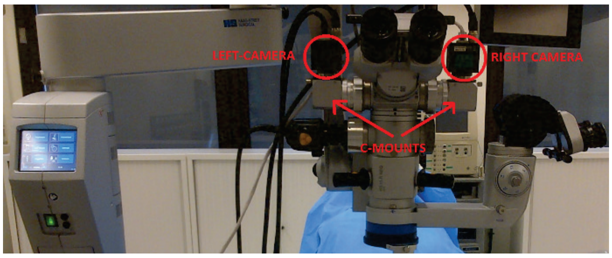

Figure 1. EurEyeCase data acquisition setup. A Haag-Streit ophthalmologic surgical microscope augmented by a stereo vertiscope (a 60° adapter to shorten the microscope body) with two C-mounted Allied Vision Prosilica GC 1290 cameras.

In the following, we describe in detail two practicable 3D visualizations of the retinal surface: (i) a fast depth-map generation from a single stereo-image pair and (ii) a complex textured 3D mesh modeling from multiple stereo pairs of images. We use a pre-calibrated stereo setup on a single zoom level with the focal length of both cameras fixed to 7680 pixels.

To generate a depth-map from a single stereo-image pair, we use the functionality of MATLAB Computer Vision System Toolbox to perform image rectification, disparity computation, and noise filtering of the resulting dense 3D pointcloud. The relative left↔right eye transformation in the form of a fundamental matrix [1] is known thanks to the stereo calibration.

Rectification is a commonly used technique which transforms the problem of depth-map estimation from a 2D to a 1D correspondence search, thus speeding up the computation significantly. Once the stereo-image pair is rectified, it is sufficient to search for correspondence only on the individual lines of the two images. To create the rectified pair, we employ two MATLAB functions `estimateUncalibratedRectification()` and `imwarp()`. Disparity, i.e., the reciprocal depth, is then computed by the semi-global matching algorithm (SGM) [2] using the function `disparity()`. Providing that the disparity map is quite complete, i.e., it does not contain many holes, it is possible to generate a visually pleasing dense 3D pointcloud by instantiating each of the pixels in 3D space at the depth inversely proportional to the estimated disparity. Function `pcdenoise()` is then employed to remove noise, i.e., inconsistent 3D points, from the resulting pointcloud using the method of [3].

In order to be able to fuse the information from multiple stereo-image pairs, the corresponding cameras have to be externally calibrated in the same coordinate system. To this end, we use a publicly available Structure-from-Motion (SfM) tool, VisualSFM [4]. The tool follows the standard SfM pipeline design where extracted natural image features, in this case SiftGPU [5], are matched between different image pairs and resulting multiview feature tracks are triangulated into 3D points. Multiple stereo-image pairs are then combined to estimate a consensual 3D surface mesh which matches well to the available observations. A state-of-the-art method is employed, based on Delaunay tetrahedralization of the 3D space and global surface estimation by a graph-cut [6] in the implementation of [7], available either as a binary or through a web service [8].

RESULTS

Due to the lack of human-eye targets, we limit ourselves to ex-vivo pig eyes for our experiments. Namely, we use the stereo video sequences of open-sky eyes, i.e., eyes of which the lenses have been removed. If a real in-vivo eye target were used, the optical effect of the actual eye lens would need to be accounted for [9].

The particular sequence we work with is a 110 seconds long stereo video viewing the eye at different zoom levels and slightly sliding to the sides. To reduce the number of processed data and remove the inherent redundancy, we restrict ourselves to a subsequence representing a single pre-calibrated zoom level and extract frames at 0.5 fps, leading to 10 stereo-image pairs.

First, we show how a depth-map and a 3D pointcloud can be generated for one of the stereo-image pairs. Note that one can easily visualize semantic information, e.g., the vessel segmentation results [10], in 3D. Instead of the original image, a "heatmap" is used to assign colors to the resulting 3D points, see Fig. 2.

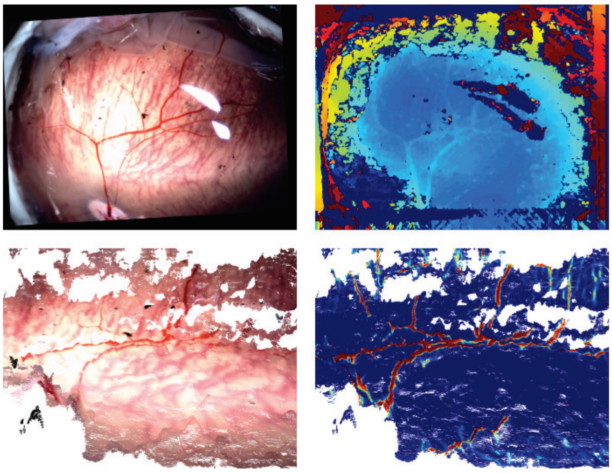

Figure 2. Rectification (top left), disparity computation (top right), and de-noised dense 3D pointcloud generation. The colors of the points correspond either to the pixel colors from the respective image (bottom left) or to the results of vessel segmentation [10] (bottom right). Blue denotes a very low while red a very high confidence of a certain area being a vein.

Second, we externally calibrate the 10 stereo-image pairs in few tens of seconds and submit a 3D surface reconstruction job to the web service [8]. Due to the specific acquisition conditions which do not bring a sufficient translation baseline for the frames in the sequence, we decreased the threshold for the minimum allowed triangulation angle `min_apical_angle` to 1 degree. The job exited in slightly over 10 minutes giving a textured 3D mesh consisting of more than 130 thousand triangles. Such a mesh gives the surgeons a nice overview of the relative configurations of different eye structures and helps them to plan the trajectories of the tools accordingly, see Fig. 3.

CONCLUSION AND DISCUSSION

We presented two techniques to construct the 3D retinal model from a stereo video sequence: (i) a fast depth-map generation from a single stereo-image pair and (ii) a complex textured 3D mesh modeling from multiple stereo pairs of images. Currently, no prior knowledge about the expected shape of the eye is input to the reconstruction, mainly due to the fact that open-sky eyes do not have the — close to perfect — spherical shape of the closed eyes. Incorporating such a shape prior, when relevant and satisfied, would make the reconstruction more robust and the 3D surfaces more accurate.

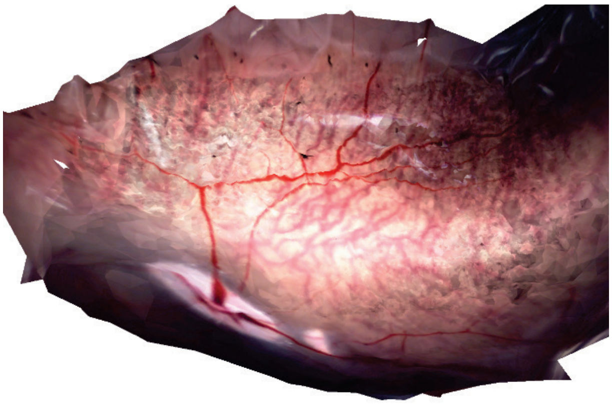

Figure 3. Reconstructed textured 3D surface mesh rendered from a novel viewpoint which cannot be attained by manipulating the microscope. Endowing the surgeons with the capability to generate such a view helps them to better understand the positions and relative configurations of the eye structures.

ACKNOWLEDGEMENT

Research funded by the EU's Horizon 2020 Programme under grant agreement No 645331 – EurEyeCase.

REFERENCES

[1] Hartley R., Zisserman A. "Multiple View Geometry in Computer Vision," 2nd edition, Cambridge University Press, 2003.

[2] Hirschmüller H. "Accurate and efficient stereo processing by semi-global matching and mutual information," in Proceedings of the 18th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2005; 807-814 vol. 2.

[3] Rusu R.B. et al. "Towards 3D point cloud based object maps for household environments," RAS; 2008; 56(11): 927-941.

[4] Wu C. "VisualSFM: A visual structure from motion system – http://ccwu.me/vsfm," 2013.

[5] Wu C. "SiftGPU: A GPU implementation of scale invariant feature transform (SIFT) – http://cs.unc.edu/~ccwu/siftgpu," 2011.

[6] Labatut P., Pons J.-P., Keriven R. "Efficient multi-view reconstruction of large-scale scenes using interest points, Delaunay triangulation and graph cuts," in Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV) 2007; 1-8.

[7] Jancosek M., Pajdla T. "Multi-view reconstruction preserving weakly supported surfaces," in Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2011; 3121-3128.

[8] Heller J. et al. "3D reconstruction from photographs by CMP SfM web service," in Proceedings of the 14th IAPR International Conference on Machine Vision Applications (MVA) 2015; 30-34.

[9] Bergeles C. et al. "Wide-Angle Intraocular Imaging and Localization," in Proceedings of the 12th International Conference on Medical Image and Computer-Assisted Intervention (MICCAI) 2009; 540-548.

[10] Maninis K.K., Pont-Tuset J., Arbeláez P., Van Gool L. "Deep Retinal Image Understanding," in Proceedings of the 19th International Conference on Medical Image and Computer-Assisted Intervention (MICCAI) 2016; 1-8.