Semantic Virtual Shadows (SVS) for Improved Perception in 4D OCT Guided Surgery

Michael Sommersperger[1]([✉]), Shervin Dehghani[1], Philipp Matten[2],
Kristina Mach[1], M. Ali Nasseri[3], Hessam Roodaki[4], Ulrich Eck[1],
and Nassir Navab[1]

[1] Technical University of Munich, Munich, Germany
`michael.sommersperger@tum.de`
[2] Medical University of Vienna, Vienna, Austria
[3] Klinikum Rechts der Isar, Augenklinik, Munich, Germany
[4] Carl Zeiss Meditec AG, Munich, Germany

Abstract. Swept-Source Optical Coherence Tomography (SS-OCT) integrated with surgical microscopes has enabled fast, high-resolution, and volumetric visualization of delicate tissue-instrument interactions. However, some visual features, which provide essential perceptual information in microscopic surgery, are not present in 4D OCT. Such a feature is the shadow of the surgical instruments cast onto the retina by the endo-illumination probe, which is among the most important cognitive cues for perceptual distance estimation. In this work, we propose *Semantic Virtual Shadows (SVS)*, a novel concept to artificially generate instrument-specific shadows in OCT volumes, enabling naturally non-existent but important perceptual cues that are present in microscopic surgery. Semantic scene information is leveraged by considering only voxels associated with shadow-casting and shadow-receiving objects, identified using a learning-based approach and efficient volume processing, respectively. Real-time performance is achieved by a precomputed semantic shadow volume texture that assigns a shadowing factor to each voxel associated with a shadow-receiving object. The novelty of the method includes not only instrument-specific shadowing on the surface anatomy but also exclusively on deep-seated subsurface structures, providing advantages for various vitreoretinal procedures. Our user study indicates the benefits of the method for 4D OCT-guided surgery in several cognitive and performance-specific aspects.

Keywords: Optical Coherence Tomography · Real-Time Visualization · Visual Perception

1 Introduction

Navigating surgical instruments in vitreoretinal procedures requires extreme manual dexterity. Surgeons need to manipulate fine structures, such as an
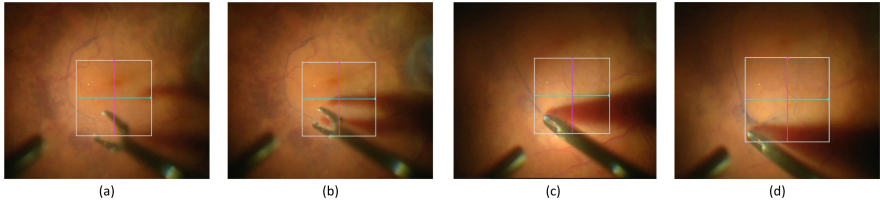
Fig. 1. A surgical forceps is carefully guided towards the retina (a-b), then grasps (c) and peels (d) a thin membrane from the surface. The instrument shadow cast onto the retina acts as a main visual cue for perceptual distance estimation.

Epiretinal Membrane (ERM), which can measure up to only $60\mu m$ [16], or carefully placing a microsurgical cannula within the on average $250\mu m$ [6] thick retina for intra or subretinal injections.

Accurate distance perception regarding the surgical instruments and the retina is of utmost importance for punctuated surgical action. Any unintentional contact with the retina could cause severe damage to important retinal cells. In conventional procedures, surgeons rely only on visual guidance from an operating microscope that only allows a top view. An endo-illumination probe is inserted into the eye to visualize the operating area. When surgical instruments are introduced, they cast a shadow onto the retina, relative to the position of the illumination probe. Such shadows provide one of the most essential cognitive cues to perceive the distance between the instrument and the retina and therefore are important for performing surgical action. Figure 1 shows frames of a video sequence during retinal membrane peeling. While the instrument approaches the retina, the distance to its shadow gradually decreases until the instrument tip coincides with the tip of its shadow at the retina.

Besides performing a procedure through this conventional stereoscopic view, recently, other imaging modalities have also become available. Technical advances have led to the integration of Swept-Source Optical Coherence Tomography (SS-OCT) into surgical microscopes [2,3,7], enabling near video-rate, depth-resolved imaging. Such 4D OCT systems have the potential to generate advanced visualizations of surgical interactions and improved treatments. While previous studies have demonstrated the feasibility of 4D OCT-guided surgery [1,9], some important perceptual cues that are available in the microscopic images are not present in OCT, such as the instrument shadow generated by the illumination probe. The absence of such perceptual cues increases the burden for surgeons to adapt to 4D OCT. While instrument shadows still exist in OCT, they are generated by the OCT laser and occur when the laser beam is blocked by the instrument such that structures below the instrument are fully obscured in the imaging modality. This OCT shadow is fixed in perspective and always occurs directly beneath the instrument. Hence, it does not provide the same intuitive cognitive cues, especially when viewing the volume from a top view, which is the usual view during ophthalmic surgery. To enable perceptually similar features as generated by the illumination probe, advanced visualization techniques need to be explored. Such visualizations could

provide cues that improve the general usability of 4D-OCT without deviating from the standard workflow and steepen the learning curve for surgeons.

In this paper, we propose *Semantic Virtual Shadows (SVS)*, a concept that integrates such visual cues for perceptual distance estimation in 4D OCT. By identifying shadow-casting and shadow-receiving objects in the OCT volume, we augment the shadow of surgical instruments on the retina. In particular, instrument-specific shadows can be generated in one case on the retinal surface, and in other cases, exclusively on deep-seated layers below the surface. Precomputing a semantic shadow volume texture prior to direct volume rendering (DVR) enables real-time performance and more flexibility in the rendering approach. We demonstrate the versatility of SVS by proposing two visualization approaches for different scenarios in retinal surgery, including retinal membrane peeling and subretinal injection.

2   Related Work

So far, only a few works have addressed the integration of perceptual cues into OCT volume rendering algorithms for interactive surgical visualizations in 4D OCT. Among these, approaches to color voxels based on distance information have been explored as a means of conveying spatial perceptual understanding. In an early study [1], a color transfer function was applied to each voxel, based on the distance to the center of mass of the volume. In a later work, [14] proposed a visualization that leverages a perceptually linear color map, which is anchored by a retinal layer segmentation, thus encoding distance information to an anatomical reference depth in a color map. They furthermore employ shadow rays [10] to enhance surface structures. However, their method is limited to local shadow rays to achieve real-time performance. There are only a few works that deal with the generation of a shadow for distance estimation in 4D OCT. Only in [5] do the authors propose to define a virtual light direction orthogonal to the OCT A-scans and employ shadow rays to create a shadow projection on an external plane positioned next to the volume. Such visualizations, however, are perceptually similar to 2D lateral volume projections, do not convey correct distance cues in presence of curved structures, and require removing the gaze from the surgical area to perceive instrument distance.

As opposed to previous works, we integrate an instrument-specific shadow augmented directly in the rendered OCT volume. Hence, our method aims to provide surgeons with familiar cognitive cues for perceptual distance estimation, as described in Sect. 1. Additionally, identifying shadow-casting and shadow-receiving objects allows the generation of instrument shadows on a specific retinal layer below the surface, which is not possible with previous approaches due to the self-shadowing of the retina.

3   Methodology

We define a *Semantic Virtual Shadow (SVS)* as a shadow that is only generated by identified shadow-casting objects and is explicitly cast onto identified
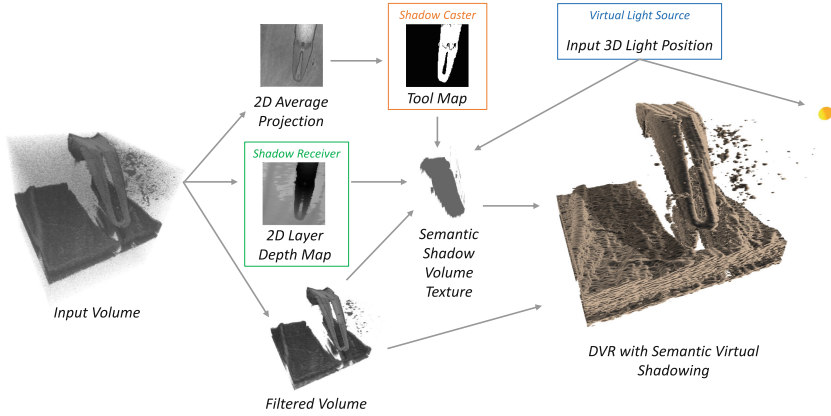
Fig. 2. Method overview: given an OCT volume and a virtual light source position, first shadow-casting and shadow-receiving objects are identified. These allow the generation of a *semantic shadow volume texture*, which is finally consumed by the DVR.

shadow-receiving objects. In our application, the primary use case of SVS is to generate visual cues for spatial distance perception. Taking into consideration shadow-casting and shadow-receiving objects as well as a virtual light source, a semantic shadow volume V_s can be constructed as a 3D texture that assigns a shadowing factor to each voxel associated with a shadow-receiving object in the OCT volume. This precomputed V_s is directly consumed by the direct volume rendering (DVR) to generate the instrument shadow augmentation. An overview of the proposed method is shown in Fig. 2.

While in theory any segmentation method could be employed to identify, we combine efficient volume processing with a learning based approach to approximate the segmentations and achieve the required processing rates for 4D OCT. For our interventional scenarios, we define the instrument as the shadow-casting object. Inspired by [15], to segment the instrument we first generate a 2D average projection image along the OCT A-scan direction (Z-axis) using

$$P_{\mathrm{avg}}(x,y) = \frac{1}{N} \sum_z I(x,y,z) \tag{1}$$

from which A-scans containing parts of the instrument signal can be identified. We train a U-net style [8] architecture with a ResNet34 [4] backbone to generate a 2D binary instrument segmentation map M_{sc}. Since the OCT signal is blocked at the instrument surface, the A-scans corresponding to identified pixels in M_{sc} contain only instrument information. Therefore the instrument can be identified by thresholding the voxel intensities of A-scans obtained by M_{sc}. We further define the retinal layer, on which the SVS is generated as the shadow-receiving object. Due to their terrain-like surface properties, the retinal layer surface can be defined by a 2D depth map M_{sr}, indicating the relative surface position for each A-scan in the volume. Further, a distance parameter d_r is introduced,

defining the extent of the shadow-receiving object. To generate V_s, a shadowing factor is calculated for each voxel p associated with a shadow-receiving object. Given the position of the virtual light source p_l and $L(p) = \frac{p_l - p}{|p_l - p|}$ defining the light direction, $V_s(p)$ is calculated with:

$$V_s(p) = \begin{cases} \prod_i (1 - (I(p + i \cdot L(p))^\alpha) \cdot M_{sc}(p + i \cdot L(p)) & \text{if } |M_{sr}(p_x, p_y) - p_z| < d_r \\ 1.0 & \text{otherwise.} \end{cases} \quad (2)$$

sampling volume intensity values $I(p)$ along the light direction through the entire volume. Note that, in our equations, the z axis corresponds to the axial A-scan direction. We precompute V_s prior to volume raymarching using a compute shader program, enabling high update rates and flexibility in the design of the DVR algorithm that directly consumes V_s. In the following, we propose two specific visualization approaches that integrate SVS according to equation (2) with different shadow-receiving objects, demonstrating the versatility and advantages of the method for different vitreoretinal procedures.
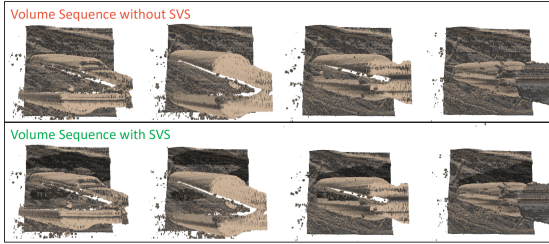

Fig. 3. Comparison of temporal OCT volumes rendered with Phong shading (top) and integrated SVS generating shadows on the retina surface (bottom).
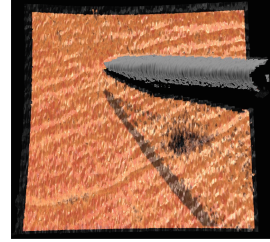
Fig. 4. SVS on the RPE layer. The retina above the RPE does not generate shadow.

3.1 Shadow Augmentation on Surface Structures

During surgical phases, in which the instrument is located above the retinal surface, unintentional contact needs to be avoided. By defining the retinal surface as the shadow-receiving object, the SVS generates cognitive cues for surface distance perception. We use the following 2D projection for M_{sr}:

$$P_{\text{surf}}(x, y) = \min_z \ I(x, y, z) > t_s \quad (3)$$

where $t_s = 0.17$ is an empirically chosen threshold. We render the OCT volume using classic Phong shading to visualize surface structure details while integrating V_s to augment the instrument-specific shadow $C(p) = C_{Phong}(p) \cdot V_s(p)$. The top row of Fig. 3 shows frames of a volume sequence with Phong shading, while in the bottom the same volumes are rendered with both Phong shading and SVS.

3.2 Shadow Augmentation on Subsurface Structures

In intra- and subretinal injection procedures, once reached the retinal surface, surgeons need to carefully guide the cannula to the injection target without damaging the underlying Retinal Pigment Epithelium (RPE) which is a comparably deep-seated layer in OCT. We propose to augment an instrument shadow explicitly on the RPE and visualize the retinal surface semi-transparent. Compared to the previous section, we define the RPE as the shadow-receiving object. Since in OCT imaging, the RPE typically is a hyper-reflective layer [17], M_{sr} can be efficiently approximated by:

$$P_{\text{subsurf}}(x, y) = \arg\max_z \ I(x, y, z) \tag{4}$$

During volume raymarching, we apply Phong shading models to visualize the instrument and render the retinal surface semi-transparent while preserving surface highlights, as presented in [12]. As previously, Eq. (3) is used to obtain the surface depth. For the RPE and underlying structures, we integrate V_s and apply an RGB color map $C_{RGB}(p)$. During volume raymarching, this leads to the following convention, integrated via a piece-wise linear opacity function:

$$C(p) = \begin{cases} C_{PhongTool}(p) & \text{if } M_{sc}(p) \\ C_{PhongILM}(p) & \text{if } |P_{surf}(p_x, p_y) - p_z| < d_{surf} \\ C_{RGB}(p) \cdot V_s(p) & \text{if } P_{subsurf}(p_x, p_y) - p_z < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Here, $C_{PhongTool}(p)$ and $C_{PhongILM}(p)$ are the Phong shading models for the instrument and the retinal surface, respectively. We refer to Fig. 4 and in particular to our supplementary video demonstrating visualization examples.

4 Experiments and Results

4.1 Implementation and Comparative Time Profiling

Our instrument segmentation network was trained on 2D projection images generated from OCT volumes using Eq. 1. In total 330 OCT volumes of resolution $391 \times 391 \times 720$ (spiral scanning) acquired from a model eye and 160 volumes of resolution $128 \times 512 \times 1024$ (linear scanning) acquired from ex-vivo porcine eyes were used to generate a data set of 3928 images, including data augmentation strategies. The axial projection images were manually labeled by two biomedical engineers. We used Pytorch 1.13 for model training and TensorRT 8.4 for optimization. We implemented our method using C++ and OpenGL 4.6 (Windows 10 with Intel Core i7-8700K @3.7 GHz, Nvidia RTX 3090Ti), employing a compute shader to generate the 2D projection images and V_s, and a fragment shader for DVR.

Table 1 shows the inference times of SVS compared to baseline shadow rays (SR), which were previously integrated for OCT visualization [14]. For ablation,

we also compare SVS to only using a shadow volume buffer without semantic information (SB). The visualization parameters were adjusted in all methods to achieve a similar visual outcome. The volumes were rendered at a resolution of 3840×2160. To further demonstrate the potential of SVS to support surgical tasks, our method was integrated into the 4D SS-OCT system presented in [2] with a volume acquisition rate of $10Hz$. Surgical actions with a forceps in a phantom eye model are demonstrated in our supplementary material.

Table 1. Average inference time in milliseconds comparing *SVS*, a 3D shadow volume buffer without semantic information (SB) and shadow rays (SR) as baseline.

Method	Filter	Enface	Depth Map	Seg	Shadow	Render	Total
SVS	7.7 ± 0.7	0.7 ± 0.13	0.7 ± 0.1	3.8 ± 0.1	2.3 ± 0.2	13.5 ± 0.2	$\mathbf{28.8 \pm 0.8}$
SB	7.7 ± 0.7	-	-	-	51.9 ± 6.8	14.6 ± 0.3	74.3 ± 6.8
SR	7.7 ± 0.7	-	-	-	-	331.3 ± 4.2	339.1 ± 4.6

4.2   User Study

To evaluate the effect of SVS for distance perception compared to baseline DVR in 4D OCT guided surgery, we conducted a user study with 12 biomedical experts (10 male, 2 female) familiar with OCT. We conducted the study in a virtual environment simulating 4D OCT using the method proposed in [11], which was displayed using stereo rendering in a VR (HTC VIVE Pro) headset. A 3D input device (3D Systems Geomagic Touch) was used to control a calibrated virtual surgical tool, where a motion scaling of $4:1$ was applied to not affect the results by the limitations of manual dexterity. Prior to the study, participants were asked to perform vision tests to ensure normal visual acuity, including stereo vision, contrast vision, and normal visual field. The study data was anonymized and performed in accordance with the declaration of Helsinki. Users were asked to familiarize themselves with the interaction in the virtual environment before the study, reducing the impact of a learning curve on the study results. For each trial, we randomly positioned a small target point close to the retina, simulating anatomical structures that could be targeted in a surgical scenario. Participants were asked to move the instrument tip to the target and press a button to confirm the positioning. Each uses performed 10 trials, each in the following three variants: (i) baseline DVR with Phong shading as in [13], (ii) (SVS_{Tool}) treating the instrument as a shadow-casting object and the retinal surface as a shadow-receiving object, and (iii) $SVS_{ToolTarget}$ with both the instrument and the target as shadow-casting objects, as the target represents an anatomical structure that could theoretically be segmented in OCT. We provide examples of the study variants in our supplementary material. During the study, the order of the trial variants and the position of the virtual light source was randomized, while the view onto the retina was fixed at $10°C$ from the axial

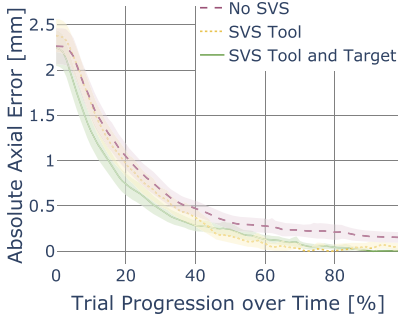**Fig. 5.** Depth-specific error over the progression of the trials.
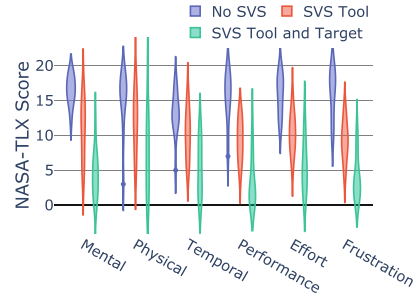
Fig. 6. Task load rating comparing baseline, SVS_{Tool} and $SVS_{ToolTarget}$.

OCT direction. Over the progression of each trial, we extracted the distance error to the target and to the retina. With baseline DVR, a final mean targeting error of 0.29 mm (± 0.18) was achieved, while with SVS_{Tool} and $SVS_{ToolTarget}$, mean errors of 0.19 mm (± 0.15) and 0.12 mm (± 0.07) where achieved respectively. After identifying unequal variances in the distributions, a Kruskal-Wallis test obtained significant differences between the methods ($p < 0.001$). Figure 5 shows the distance error to the target in axial OCT direction measured over the trial progression, along with a 95% confidence interval of the error values. With SVS, users approached with faster convergence to the target and less error variance toward the end of the trial. In addition, Fig. 6 shows the results of the NASA-TLX survey. Statistically significant differences ($p < 0.001$) in all categories could be obtained based on the ANOVA test, while equal variances were found in all categories.

5   Discussion and Conclusion

The results of our user study indicate that the proposed method is able to generate effective perceptual cues for distance estimation in 4D OCT. This is reflected in improved targeting performance, as well as a high acceptance of the generated perceptual cues. Figure 5 suggests more robust targeting in both SVS_{Tool} and $SVS_{ToolTarget}$ variants, implying higher confidence in the approach. Preliminary qualitative feedback from four clinicians confirmed the intuitiveness of SVS and its strong integration potential without changing the existing clinical workflow. An essential component of SVS is the identification of shadow-casting and shadow-receiving objects in the volume. Compared to previous methods, SVS is able to generate perceptual cues exclusively on subsurface structures. Precomputing a semantic shadow volume texture was shown to achieve processing rates suitable for 4D OCT systems with volume acquisition rates of $24.2 Hz$ [7]. In the future, prior knowledge of the instrument's geometrical model could be integrated to enable more accurate shadow representations. We also plan

to investigate optimal virtual light source positions and envision employing the illumination probe to control the virtual light for more intuitive interactions.

In conclusion, the proposed SVS augments visual cues that are naturally not present in OCT, but essential in microscopic vitreoretinal surgery. The flexibility of our approach enables object-specific shadow generation not only on surface structures but also exclusively on subsurface structures, supporting various surgical procedures. We provided a general definition of object-specific shadow generation and demonstrated our method on a 4D SS-OCT system for live display. In our user study, SVS was shown to provide intuitive and effective visual cues for targeted instrument maneuvers.

References

1. Bleicher, I.D., Jackson-Atogi, M., Viehland, C., Gabr, H., Izatt, J.A., Toth, C.A.: Depth-based, motion-stabilized colorization of microscope-integrated optical coherence tomography volumes for microscope-independent microsurgery. Transl. Vis. Sci. Technol. **7**(6), 1–1 (2018)
2. Britten, A., et al.: Surgical microscope integrated MHz SS-OCT with live volumetric visualization. Biomed. Optics Express **14**(2), 846–865 (2023)
3. Carrasco-Zevallos, O.M., Viehland, C., Keller, B., McNabb, R.P., Kuo, A.N., Izatt, J.A.: Constant linear velocity spiral scanning for near video rate 4D OCT ophthalmic and surgical imaging with isotropic transverse sampling. Biomed. Optics Express **9**(10), 5052 (2018)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
5. Jeong, H., Kim, H.J., Hyeon, M.G., Kim, P., Choi, Y., Kim, B.M.: Shadow extension for ray casting enhances volumetric visualization in real-time 4D-OCT. Optics Commun. **460**, 125237 (2020)
6. Jo, Y.J., Heo, D.W., Shin, Y.I., Kim, J.Y.: Diurnal variation of retina thickness measured with time domain and spectral domain optical coherence tomography in healthy subjects. Invest. Ophthalmol. Vis. Sci. **52**(9), 6497–6500 (2011)
7. Kolb, J.P., Draxinger, W., Klee, J., Pfeiffer, T., Eibl, M., Klein, T., Wieser, W., Huber, R.: Live video rate volumetric OCT imaging of the retina with multi-MHz A-scan rates. PLoS ONE **14**(3), e0213144 (2019)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Roodaki, H., di San Filippo, C.A., Zapp, D., Navab, N., Eslami, A.: A surgical guidance system for big-bubble deep anterior lamellar keratoplasty. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 378–385. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46720-7_44

10. Ropinski, T., Kasten, J., Hinrichs, K.: Efficient shadows for GPU-based volume raycasting (2008)
11. Sommersperger, M., et al.: Surgical scene generation and adversarial networks for physics-based iOCT synthesis. Biomed. Optics Express **13**(4), 2414–2430 (2022)
12. Trout, R.M., et al.: Methods for real-time feature-guided image fusion of intrasurgical volumetric optical coherence tomography with digital microscopy. Biomed. Optics Express **14**(7), 3308–3326 (2023)
13. Viehland, C., et al.: Enhanced volumetric visualization for real time 4d intraoperative ophthalmic swept-source oct. Biomed. Optics Express **7**(5), 1815–1829 (2016)
14. Weiss, J., Eck, U., Nasseri, M.A., Maier, M., Eslami, A., Navab, N.: Layer-aware iOCT volume rendering for retinal surgery. In: Kozlíková, B., Linsen, L., Vázquez, P.P., Lawonn, K., Raidou, R.G. (eds.) Eurographics Workshop on Visual Computing for Biology and Medicine. The Eurographics Association (2019)
15. Weiss, J., Sommersperger, M., Nasseri, A., Eslami, A., Eck, U., Navab, N.: Processing-aware real-time rendering for optimized tissue visualization in intraoperative 4D OCT. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 267–276. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_26
16. Wilkins, J.R., et al.: Characterization of epiretinal membranes using optical coherence tomography. Ophthalmology **103**(12), 2142–2151 (1996)
17. Zhang, T., Kho, A.M., Yiu, G., Srinivasan, V.J.: Visible light optical coherence tomography (OCT) quantifies subcellular contributions to outer retinal band 4. Transl. Vis. Sci. Technol. **10**(3), 30–30 (2021)