Real-time intraoperative 4D full-range FD-OCT based on the dual graphics processing units architecture for microsurgery guidance

Kang Zhang* and Jin U. Kang

Department of Electrical and Computer Engineering, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 USA
*kzhang8@jhu.edu

Abstract: Real-time 4D full-range complex-conjugate-free Fourier-domain optical coherence tomography (FD-OCT) is implemented using a dual graphics processing units (dual-GPUs) architecture. One GPU is dedicated to the FD-OCT data processing while the second one is used for the volume rendering and display. GPU accelerated non-uniform fast Fourier transform (NUFFT) is also implemented to suppress the side lobes of the point spread function to improve the image quality. Using a 128,000 A-scan/second OCT spectrometer, we obtained 5 volumes/second real-time full-range 3D OCT imaging. A complete micro-manipulation of a phantom using a microsurgical tool is monitored by multiple volume renderings of the same 3D date set with different view angles. Compared to the conventional surgical microscope, this technology would provide the surgeons a more comprehensive spatial view of the microsurgical site and could serve as an effective intraoperative guidance tool.

OCIS codes: (100.2000) Digital image processing; (100.6890) Three-dimensional image processing; (110.4500) Optical coherence tomography; (170.3890) Medical optics instrumentation.

References and links

1. K. Zhang, W. Wang, J. Han, and J. U. Kang, "A surface topology and motion compensation system for microsurgery guidance and intervention based on common-path optical coherence tomography," IEEE Trans. Biomed. Eng. **56**(9), 2318–2321 (2009).
2. Y. K. Tao, J. P. Ehlers, C. A. Toth, and J. A. Izatt, "Intraoperative spectral domain optical coherence tomography for vitreoretinal surgery," Opt. Lett. **35**(20), 3315–3317 (2010).
3. A. Stephen, Boppart, Mark E. Brezinski and James G. Fujimoto, "Surgical guidance and intervention," in *Handbook of Optical Coherence Tomography*, B. E. Bouma and G. J Tearney, ed. (Marcel Dekker, New York, NY, 2001).
4. W.-Y. Oh, B. J. Vakoc, M. Shishkov, G. J. Tearney, and B. E. Bouma, ">400 kHz repetition rate wavelength-swept laser and application to high-speed optical frequency domain imaging," Opt. Lett. **35**(17), 2919–2921 (2010).
5. B. Potsaid, B. Baumann, D. Huang, S. Barry, A. E. Cable, J. S. Schuman, J. S. Duker, and J. G. Fujimoto, "Ultrahigh speed 1050nm swept source/Fourier domain OCT retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second," Opt. Express **18**(19), 20029–20048 (2010).
6. W. Wieser, B. R. Biedermann, T. Klein, C. M. Eigenwillig, and R. Huber, "Multi-megahertz OCT: High quality 3D imaging at 20 million A-scans and 4.5 GVoxels per second," Opt. Express **18**(14), 14685–14704 (2010).
7. T. Bonin, G. Franke, M. Hagen-Eggert, P. Koch, and G. Hüttmann, "*In vivo* Fourier-domain full-field OCT of the human retina with 1.5 million A-lines/s," Opt. Lett. **35**(20), 3432–3434 (2010).
8. K. Zhang and J. U. Kang, "Real-time 4D signal processing and visualization using graphics processing unit on a regular nonlinear-k Fourier-domain OCT system," Opt. Express **18**(11), 11772–11784 (2010).
9. M. Sylwestrzak, M. Szkulmowski, D. Szlag, and P. Targowski, "Real-time imaging for spectral optical coherence tomography with massively parallel data processing," Photonics Lett. Poland **2**(3), 137–139 (2010).
10. J. Probst, D. Hillmann, E. Lankenau, C. Winter, S. Oelckers, P. Koch, and G. Hüttmann, "Optical coherence tomography with online visualization of more than seven rendered volumes per second," J. Biomed. Opt. **15**(2), 026014 (2010).

11. K. Zhang and J. U. Kang, "Graphics processing unit accelerated non-uniform fast Fourier transform for ultrahigh-speed, real-time Fourier-domain OCT," Opt. Express **18**(22), 23472–23487 (2010).
12. Y. Watanabe, S. Maeno, K. Aoshima, H. Hasegawa, and H. Koseki, "Real-time processing for full-range Fourier-domain optical-coherence tomography with zero-filling interpolation using multiple graphic processing units," Appl. Opt. **49**(25), 4756–4762 (2010).
13. B. Baumann, M. Pircher, E. Götzinger, and C. K. Hitzenberger, "Full range complex spectral domain optical coherence tomography without additional phase shifters," Opt. Express **15**(20), 13375–13387 (2007), http://www.opticsinfobase.org/abstract.cfm?URI=oe-15-20-13375.
14. NVIDIA, "NVIDIA CUDA C Programming Guide Version 3.2" (2010).

1. Introduction

Microsurgeries require both physical and optical access to limited space in order to perform task on delicate tissue. The ability to view critical parts of the surgical region and work within micron proximity to the fragile tissue surface requires excellent visibility and precise instrument manipulation. The surgeon needs to function within the limits of human sensory and motion capability to visualize targets, steadily guide microsurgical tools and execute all surgical targets. These directed surgical maneuvers must occur intraoperatively with minimization of surgical risk and expeditious resolution of complications. Conventionally, visualization during the operation is realized by surgical microscopes, which limits the surgeon's field of view (FOV) to the *en face* scope [1], with limited depth perception of micro-structures and tissue planes.

As a noninvasive imaging modality, optical coherence tomography (OCT) is capable of cross-sectional micrometer-resolution images and a complete 3D data set could be obtained by 2D scanning of the targeted region. Compared to other modalities used in image-guided surgical intervention such as MRI, CT, and ultrasound, OCT is highly suitable for applications in microsurgical guidance [1–3]. For clinical intraoperative purposes, a FD-OCT system should be capable of ultrahigh speed raw data acquisition as well as matching-speed data processing and visualization. In recent years, the A-scan acquisition rate of FD-OCT systems has been generally reached multi-hundred-of-thousand line/second level [4,5] and approaching to multi-million line/second level [6,7]. The recent developments of graphics processing unit (GPU) accelerated FD-OCT processing and visualization have enabled real-time 4D (3D + time) imaging at the speed up to 10 volume/second [8–10]. However, these systems all work in the standard mode, and therefore suffer from spatially reversed complex-conjugate ghost images. During intraoperative imaging, for example, when long-shaft surgical tools are used, such ghost images could severely misguide the surgeons. As a solution, the GPU-accelerated full-range FD-OCT has been utilized and real-time B-scan images was demonstrated with effective complex-conjugate suppression and doubled imaging range [11,12].

In this work, we implemented the real-time 4D full-range complex-conjugate-free FD-OCT based on the dual-GPUs architecture, where one GPU is dedicated to the FD-OCT data processing while the second one is used for the volume rendering and display. GPU based non-uniform fast Fourier transform (NUFFT) [12] is also implemented to suppress the side lobes of the point spread function and to improve the image quality. With a 128,000 A-scan/second OCT engine, we obtained 5 volumes/second 3D imaging and display. We have demonstrated the real-time visualization capability of the system by performing a micro-manipulation process using a vitro-retinal surgical tool and a phantom model. Multiple volume rendering of the same 3D data set were preformed and displayed with different view angles. This technology can provide the surgeons with a comprehensive intraoperative imaging of the microsurgical region which could improve the accuracy and safety of the microsurgical procedures.

2. System configuration and data processing

The system configuration is shown in Fig. 1. In the FD-OCT system section, a 12-bit dual-line CMOS line-scan camera (Sprint spL2048-140k, Basler AG, Germany) is used as the detector

of the OCT spectrometer. A superluminescence diode (SLED) (λ_0 = 825nm, $\Delta\lambda$ = 70nm, Superlum, Ireland) is used as the light source, giving a theoretical axial resolution of 5.5μm in air. The transversal resolution was approximately 40μm assuming Gaussian beam profile. The CMOS camera is set to operate at the 1024-pixel mode by selecting the area-of-interest (AOI). The minimum line period is camera-limited to 7.8μs, corresponding to a maximum line rate of 128k A-scan/s, and the exposure time is 6.5 μs. The beam scanning was implemented by a pair of high speed galvanometer mirrors controlled by a function generator and a data acquisition (DAQ) card. The raw data acquisition is performed using a high speed frame grabber with camera link interface. To realize the full-range complex OCT mode, a phase modulation is applied to each B-scan's 2D interferogram frame by slightly displacing the probe beam off the first galvanometer's pivoting point (only the first galvanometer is illustrated in Fig. 1) [11–13].
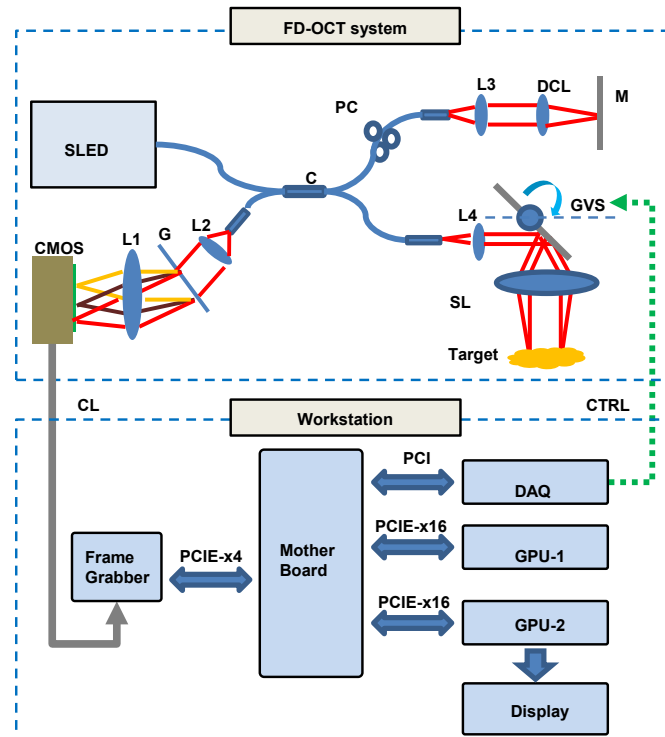

Fig. 1. System configuration: CMOS, CMOS line scan camera; G, grating; L1, L2, L3, L4 achromatic collimators; C, 50:50 broadband fiber coupler; CL, camera link cable; CTRL, galvanometer control signal; GVS, galvanometer pairs (only the first galvanometer is illustrated for simplicity); SL, scanning lens; DCL, dispersion compensation lens; M, reference mirror; PC, polarization controller.

A quad-core Dell T7500 workstation was used to host the frame grabber (PCIE-x4 interface), DAQ card (PCI interface), GPU-1 and GPU-2 (both PCIE-x16 interface), all on the same mother board. GPU-1 (NVIDIA GeForce GTX 580) with 512 stream processors, 1.59GHz processor clock and 1.5 GBytes graphics memory is dedicated for raw data processing of B-scan frames. GPU-2 (NVIDIA GeForce GTS 450) with 192 stream processors, 1.76GHz processor clock and 1.0 GBytes graphics memory is dedicated for the volume rendering and display of the complete C-scan data processed by GPU-1. The GPU is programmed through NVIDIA's Compute Unified Device Architecture (CUDA) technology [14]. The software is developed under the Microsoft Visual C + + environment with National Instrument's IMAQ Win32 APIs.

The signal processing flow chart of the dual-GPUs architecture is illustrated in Fig. 2, where three major threads are used for the FD-OCT system raw data acquisition (Thread 1), the GPU accelerated FD-OCT data processing (Thread 2), and the GPU based volume rendering (Thread 3). The three threads synchronize in the pipeline mode, where Thread 1 triggers Thread 2 for every B-scan and Thread 2 triggers Thread 3 for every complete C-scan, as indicated by the dashed arrows. The solid arrows describe the main data stream and the hollow arrows indicate the internal data flow of the GPU. Since the CUDA technology currently does not support direct data transfer between GPU memories, a C-Scan buffer is placed in the host memory for the data relay.
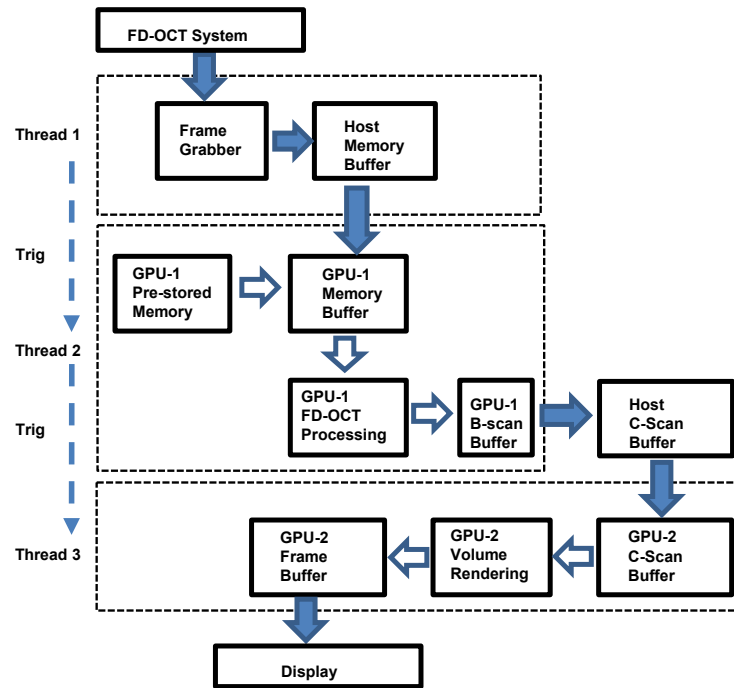

Fig. 2. Signal processing flow chart of the dual-GPUs architecture. Dashed arrows, thread triggering; Solid arrows, main data stream; Hollow arrows, internal data flow of the GPU. Here the graphics memory refers to global memory.

Compared to previously reported systems, this dual-GPUs architecture separates the computing task of the signal processing and the visualization into different GPUs, which has the following advantages:

(1) Assigning different computing tasks to different GPUs makes the entire system more stable and consistent. For the real-time 4D imaging mode, the volume rendering is only conducted when a complete C-scan is ready, while B-scan frame processing is running continuously. Therefore, if the signal processing and the visualization are performed on the same GPU, competition for GPU resource will happen when the volume rendering starts while the B-scan processing is still going on, which could result in instability for both tasks.

(2) It will be more convenient to enhance the system performance from the software engineering perspective. For example, the A-scan processing could be further accelerated and the point spread function (PSF) could be refined by improving algorithm with GPU-1, while more complex 3D image processing task such as segmentation or target tracking can be added to GPU-2.

In our experiment, the B-scan size is set to 256 A-scans with 1024 pixel each. Using the GPU based NUFFT algorithm, GPU-1 achieved a peak A-scan processing rate of 252,000 lines/s and an effective rate of 186,000 lines/s when the host-device data transferring bandwidth of PCIE-x16 interface was considered, which is higher than the camera's acquisition line rate. The NUFFT method was effective in suppressing the side lobes of the PSF and in improving the image quality, especially when surgical tools with metallic surface are used. The C-scan size is set to 100 B-scans, resulting in $256 \times 100 \times 1024$ voxels (effectively $250 \times 98 \times 1024$ voxels after removing of edge pixels due to fly-back time of galvanometers), and 5 volumes/second. It takes GPU-2 about 8ms to render one 2D image with 512×512 pixel from this 3D data set using the ray-casting algorithm [8].
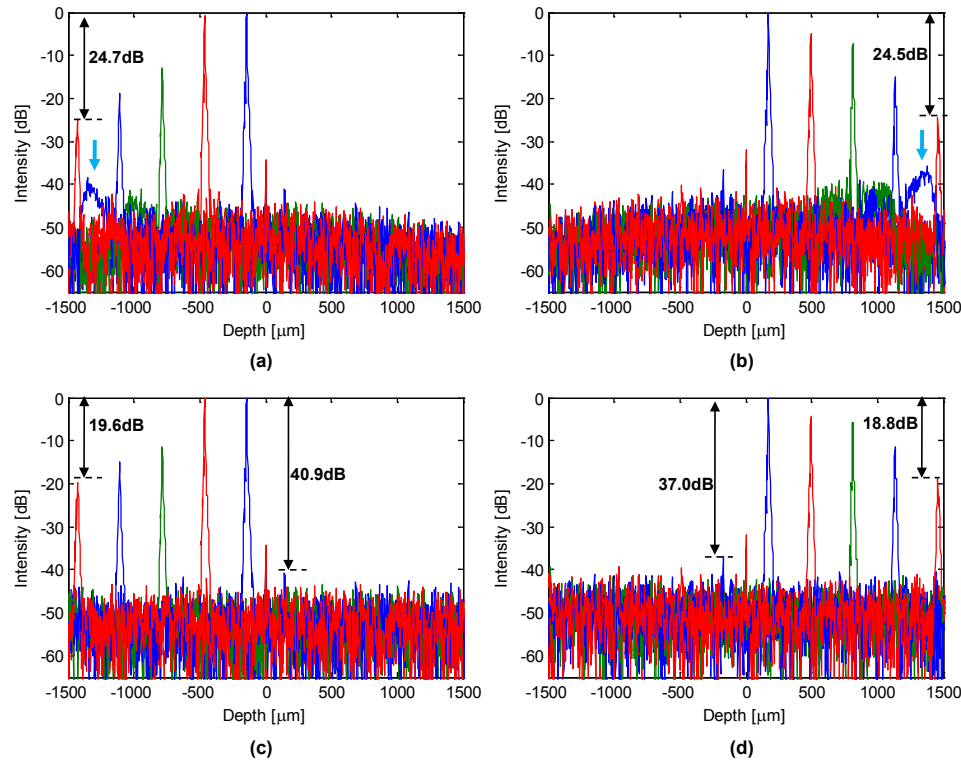
3. Results and discussion

Fig. 3. Optical performance of the system: (a) and (b), PSFs processed by linear interpolation with FFT, blue arrows indicate the side lobes of PSFs near positive and negative edges due to interpolation error. (c) and (d), PSFs processed by NUFFT.

First, we tested the optical performance of the system using a mirror as the target. At both sides of the zero-delay, PSFs at different positions are processed as A-scans using linear interpolation with FFT (Figs. 3(a) and 3(b)) and NUFFT (Figs. 3(c) and 3(d)), respectively. As one can see, using NUFFT processing, the system obtained a sensitivity fall-off of 19.6 dB from position near zero-delay to the negative edge, and 18.8 dB to the positive edge, which is lower than using linear interpolation with FFT (24.7 dB to the negative edge and 24.5 dB to the positive edge). Moreover, compared to the linear interpolation method, NUFFT obtained a constant background noise level over the whole A-scan range. The blue arrows in Figs. 3(a) and 3(b) indicates side lobes in the PSFs near both positive and negative edges as a result of interpolation error. By applying a proper super-Gaussian filter to the modified Hilbert transform [12], the conjugate suppression ratios of 37.0 dB and 40.9 dB are obtained

respectively at the positive and negative sides near zero-delay. Therefore by applying NUFFT in GPU-1, we can obtain high quality, low noise image sets for later volume rendering in GPU-2.

Then, the *in vivo* human finger imaging is conducted to test the imaging capability of biological tissue. The scanning range is 3.5mm (X) × 3.5mm (Y) lateral and 3mm (Z) for the axial full-range. The finger nail fold region is imaged as Fig. 4 (screen captured as Media 1 at 5 frame/second), where 4 frames are rendered from the same 3D data set with different view angles. The green arrows/dots on each 2D frame correspond to the same edges/vertexes of the rendering volume frame, giving comprehensive information of the image volume. As noted in Fig. 4(d), the major dermatologic structures such as epidermis (E), dermis (D), nail plate (NP), nail root (NR) and nail bed (NB) are clearly distinguishable.
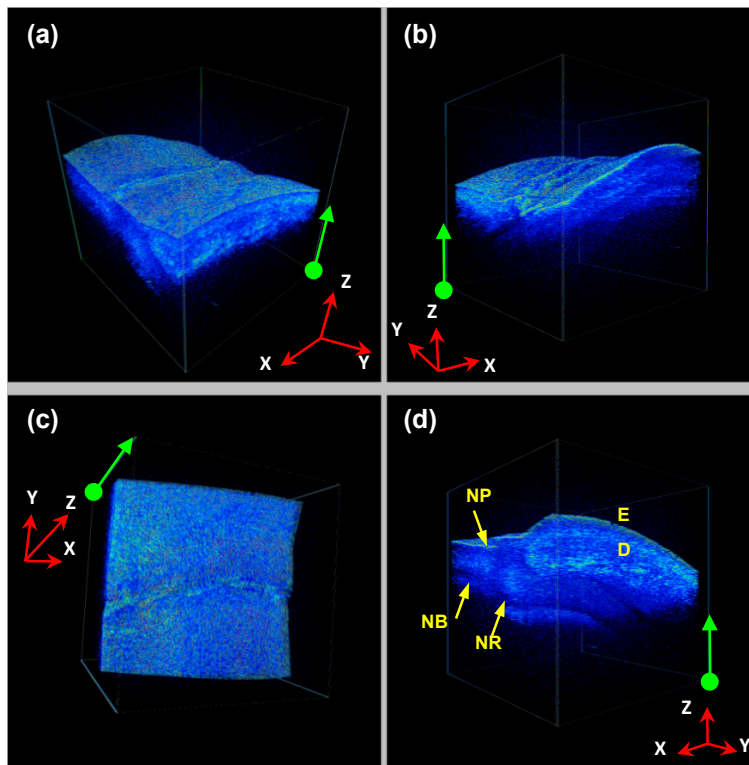

Fig. 4. (Media 1) *In vivo* human finger nail fold imaging: (a)~(d) are rendered from the same 3D data set with different view angles. The green arrows/dots on each 2D frame correspond to the same edges/ vertexes of the rendering volume frame. Volume size: 256(Y) × 100(X) × 1024(Z) voxels/ 3.5mm (Y) × 3.5mm (X) × 3mm (Z).

Finally, we performed a real-time 4D full-range FD-OCT guided micro-manipulation using a phantom model and a vitreoretinal surgical forceps, with the same scanning protocol as Fig. 5. As shown in Fig. 5, a sub-millimeter particle is attached on a multi-layered surface made of polymer layers. The mini-surgical forceps was used to pick up the particle from the surface without touching the surface. As shown in Media 2, multiple volume rendering of the same 3D date set were displayed with different view angles to allow accurate monitoring of the micro-procedure, and the tool-to-target spatial relation is clearly demonstrated in real-time. Compared to the conventional surgical microscope, this technology can provide the surgeons with a comprehensive spatial view of the microsurgical region and the depth perception. Therefore, this technology could become an effective intraoperative surgical guidance tool that could improve the accuracy and safety of the microsurgical procedures.
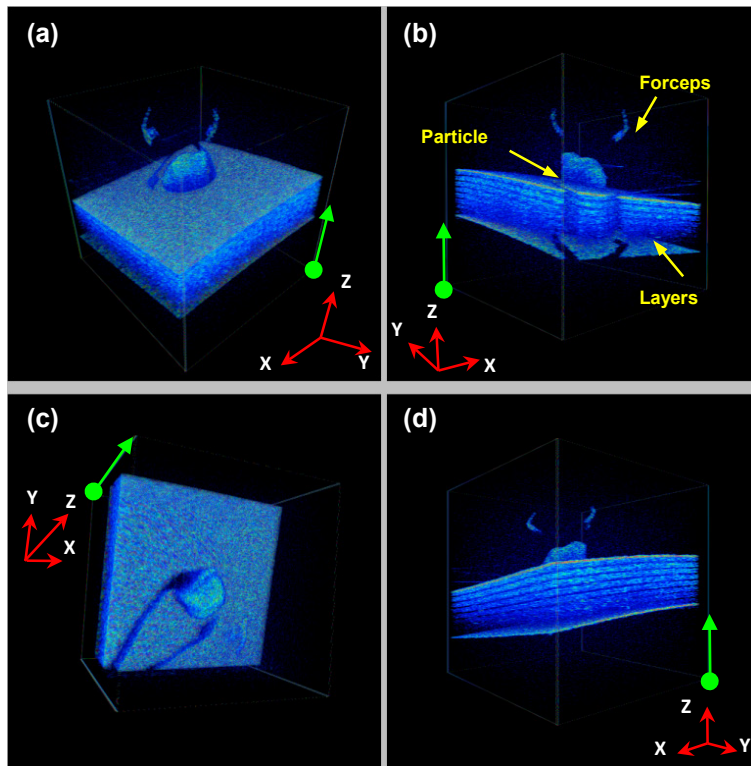
Fig. 5. (Media 2) Real-time 4D full-range FD-OCT guided micro-manipulation using a phantom model and a vitreoretinal surgical forceps. The green arrows/dots on each 2D frame correspond to the same edges/ vertexes of the rendering volume frame. Volume size: 256(Y) × 100(X) × 1024(Z) voxels/3.5mm (Y) × 3.5mm (X) × 3mm (Z).

4. Conclusion

In this work, a real-time 4D full-range FD-OCT system is implemented based on the dual-GPUs architecture. The computing task of signal processing and visualization into different GPUs and real-time 4D imaging and display of 5 volume/second has been obtained. A real-time 4D full-range FD-OCT guided micro-manipulation is performed using a phantom model and a vitreoretinal surgical forceps. This technology can provide the surgeons a comprehensive spatial view of the microsurgical site and can be used to guide microsurgical tools during microsurgical procedures effectively.

Acknowledgments