

# Real-time GPU-accelerated processing and volumetric display for wide-field laser-scanning optical-resolution photoacoustic microscopy

Heesung Kang,<sup>1,5</sup> Sang-Won Lee,<sup>1,2,3,5,6</sup> Eun-Soo Lee,<sup>1</sup> Se-Hwa Kim,<sup>1,2,4</sup>  
and Tae Geol Lee<sup>1,3,7</sup>

<sup>1</sup>Center for Nano-Bio Measurement, Korea Research Institute of Standards and Science, Daejeon 305-340, South Korea

<sup>2</sup>Center for Nanosafety Metrology, Korea Research Institute of Standards and Science, Daejeon 305-340, South Korea

<sup>3</sup>Department of Nano Science, University of Science and Technology, Daejeon 305-350, South Korea

<sup>4</sup>Department of Nano and Bio Surface Science, University of Science and Technology, Daejeon 305-350, South Korea

<sup>5</sup>These authors contributed equally to this work

<sup>6</sup>swlee76@kriss.re.kr

<sup>7</sup>tglee@kriss.re.kr

**Abstract:** Fast signal processing and real-time displays are essential for practical imaging modality in various fields of applications. However, the imaging speed in optical-resolution photoacoustic microscopy (OR-PAM), in particular, depends on factors such as the pulse repetition rate of the laser, scanning method, field of view (FOV), and signal processing time. In the past, efforts to increase acquisition speed either focused on developing new scanning methods or using lasers with higher pulse repetition rates. However, high-speed signal processing is also important for real-time volumetric display in OR-PAM. In this study, we carried out parallel signal processing using a graphics processing unit (GPU) to enable fast signal processing and wide-field real-time displays in laser-scanning OR-PAM. The average total GPU processing time for a B-mode PAM image was approximately 1.35 ms at a display speed of 480 fps when the data samples were acquired with 736 (axial)  $\times$  500 (lateral) points/B-mode-frame at a pulse repetition rate of 300 kHz. In addition, we successfully displayed maximum amplitude projection images of a mouse's ear as volumetric images with an FOV of 3 mm  $\times$  3 mm (500  $\times$  500 pixels) at 1.02 s, corresponding to 0.98 fps.

©2015 Optical Society of America

**OCIS codes:** (110.5120) Photoacoustic imaging; (170.3880) Medical and biological imaging; (170.5120) Photoacoustic imaging; (170.5810) Scanning microscopy; (180.5810) Scanning microscopy.

## References and links

1. X. Wang, Y. Pang, G. Ku, X. Xie, G. Stoica, and L. V. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain," *Nat. Biotechnol.* **21**(7), 803–806 (2003).
2. H. F. Zhang, K. Maslov, G. Stoica, and L. V. Wang, "Functional photoacoustic microscopy for high-resolution and noninvasive in vivo imaging," *Nat. Biotechnol.* **24**(7), 848–851 (2006).
3. M. L. Li, J. C. Wang, J. A. Schwartz, K. L. Gill-Sharp, G. Stoica, and L. V. Wang, "In-vivo photoacoustic microscopy of nanoshell extravasation from solid tumor vasculature," *J. Biomed. Opt.* **14**(1), 010507 (2009).
4. K. Maslov, H. F. Zhang, S. Hu, and L. V. Wang, "Optical-resolution photoacoustic microscopy for in vivo imaging of single capillaries," *Opt. Lett.* **33**(9), 929–931 (2008).
5. Z. Xie, S. Jiao, H. F. Zhang, and C. A. Puliafito, "Laser-scanning optical-resolution photoacoustic microscopy," *Opt. Lett.* **34**(12), 1771–1773 (2009).
6. B. Rao, L. Li, K. Maslov, and L. Wang, "Hybrid-scanning optical-resolution photoacoustic microscopy for in vivo vasculature imaging," *Opt. Lett.* **35**(10), 1521–1523 (2010).

7. L. Wang, K. Maslov, J. Yao, B. Rao, and L. V. Wang, "Fast voice-coil scanning optical-resolution photoacoustic microscopy," *Opt. Lett.* **36**(2), 139–141 (2011).
8. B. Rao, K. Maslov, A. Danielli, R. Chen, K. K. Shung, Q. Zhou, and L. V. Wang, "Real-time four-dimensional optical-resolution photoacoustic microscopy with Au nanoparticle-assisted subdiffraction-limit resolution," *Opt. Lett.* **36**(7), 1137–1139 (2011).
9. W. Shi, P. Hajireza, P. Shao, A. Forbrich, and R. J. Zemp, "In vivo near-realtime volumetric optical-resolution photoacoustic microscopy using a high-repetition-rate nanosecond fiber-laser," *Opt. Express* **19**(18), 17143–17150 (2011).
10. W. Shi, P. Shao, P. Hajireza, A. Forbrich, and R. J. Zemp, "In vivo dynamic process imaging using real-time optical-resolution photoacoustic microscopy," *J. Biomed. Opt.* **18**(2), 026001 (2013).
11. S. P. Mattison, R. L. Shelton, R. T. Maxson, and B. E. Applegate, "Continuous real-time photoacoustic demodulation via field programmable gate array for dynamic imaging of zebrafish cardiac cycle," *Biomed. Opt. Express* **4**(8), 1451–1463 (2013).
12. Y. Watanabe and T. Itagaki, "Real-time display on Fourier domain optical coherence tomography system using a graphics processing unit," *J. Biomed. Opt.* **14**(6), 060506 (2009).
13. N. H. Cho, U. Jung, S. Kim, W. Jung, J. Oh, H. W. Kang, and J. Kim, "High speed SD-OCT system using GPU accelerated mode for in vivo human eye imaging," *J. Opt. Soc. Korea* **17**(1), 68–72 (2013).
14. T. Shimobaba, Y. Sato, J. Miura, M. Takenouchi, and T. Ito, "Real-time digital holographic microscopy using the graphic processing unit," *Opt. Express* **16**(16), 11776–11781 (2008).
15. A. Narayanaswamy, S. Dwarakapuram, C. S. Bjornsson, B. M. Cutler, W. Shain, and B. Roysam, "Robust adaptive 3-D segmentation of vessel laminae from fluorescence confocal microscope images and parallel GPU implementation," *IEEE Trans. Med. Imaging* **29**(3), 583–597 (2010).
16. J. Yuan, G. Xu, Y. Yu, Y. Zhou, P. L. Carson, X. Wang, and X. Liu, "Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization," *J. Biomed. Opt.* **18**(8), 086001 (2013).
17. X. L. Deán-Ben, A. Ozbek, and D. Razansky, "Volumetric real-time tracking of peripheral human vasculature with GPU-accelerated three-dimensional photoacoustic tomography," *IEEE Trans. Med. Imaging* **32**(11), 2050–2055 (2013).
18. K. Peng, L. He, Z. Zhu, J. Tang, and J. Xiao, "Three-dimensional photoacoustic tomography based on graphics-processing-unit-accelerated finite element method," *Appl. Opt.* **52**(34), 8270–8279 (2013).
19. V. P. Zharov, E. I. Galanzha, E. V. Shashkov, N. G. Khlebtsov, and V. V. Tuchin, "In vivo photoacoustic flow cytometry for monitoring of circulating single cancer cells and contrast agents," *Opt. Lett.* **31**(24), 3623–3625 (2006).
20. K. Tsoi and W. Luk, "Axel: A heterogeneous cluster with FPGAs and GPUs," in *Proceedings of the 18<sup>th</sup> annual ACM/SIGDA international symposium*, pp. 115–124 (2010).
21. D. H. Jones, A. Powell, C.-S. Bouganis, and P. Y. K. Cheung, "GPU versus FPGA for high productivity computing," in *Proceedings of the 2010 International Conference on Field Programmable Logic and Applications*, pp. 119–124 (2010).
22. J. Yao, C.-H. Huang, L. Wang, J.-M. Yang, L. Gao, K. I. Maslov, J. Zou, and L. V. Wang, "Wide-field fast-scanning photoacoustic microscopy based on a water-immersible MEMS scanning mirror," *J. Biomed. Opt.* **17**(8), 080505 (2012).
23. J. Y. Kim, C. Lee, K. Park, G. Lim, and C. Kim, "Fast optical-resolution photoacoustic microscopy using a 2-axis water-proofing MEMS scanner," *Sci. Rep.* **5**, 7932 (2015).
24. J. Yao, L. Wang, J.-M. Yang, K. I. Maslov, T. T. W. Wong, L. Li, C.-H. Huang, J. Zou, and L. V. Wang, "High-speed label-free functional photoacoustic microscopy of mouse brain in action," *Nat. Methods* **12**(5), 407–410 (2015).

---

## 1. Introduction

Photoacoustic imaging is widely used as a noninvasive imaging technology that can be combined with optical absorption contrast and ultrasound spatial resolution for structural, functional, and molecular imaging [1–3]. In particular, optical-resolution photoacoustic microscopy (OR-PAM), which was first introduced by Maslov et al. [4], can provide capillary-level spatial resolution as a result of its tightly focused micron-scale laser spot size. The imaging speed in OR-PAM depends on factors such as the pulse repetition rate of the laser, scanning methods, field of view (FOV), and signal processing time. The imaging speed in the first generation OR-PAM was relatively slow because of the mechanical scanning involved and the low pulse repetition rate (10–20 Hz) of the laser utilized. The higher acquisition speed can minimize motion artifacts, reduce anesthetic duration for animal experiments, and facilitate quantitative analysis of 3D and 4D image data. Recently, since the commercialization of nanosecond pulsed lasers with a few kilohertz to a few hundred kilohertz repetition rates, several groups have actively studied methods of accelerating the

imaging acquisition speed and obtaining maximum amplitude projection (MAP) images with wide FOVs as volumetric images.

For example, Xie et al. [5] and Rao et al. [6] proposed a laser-scanning method that uses a 2D galvanometer scanner or hybrid scanning method combined with a 1D galvanometer scanner and a 1D motorized stage at a 5-kHz repetition rate. Wang et al. [7] presented a fast mechanical scanner based on a voice-coil stage with a laser repetition rate of 4 kHz. In addition, real-time volumetric laser-scanning OR-PAM images have been exhibited by Rao et al. [8] and Shi et al. [9, 10] at pulse repetition rates of 100 kHz, 320 kHz, and 600 kHz. To acquire the real-time volumetric PAM data, either small amounts of volumetric data pixels were collected for small FOVs (a few hundred micrometers) [8, 10] or block-data (or 3D data) with a large number of pixels for large FOVs (a few millimeters) and saved them without data transfer until the acquisition was complete [9, 10]. Data acquisition speed was increased by using new scanning methods and lasers with higher pulse repetition rates, but most of the volumetric OR-PAM images were still displayed via post-signal processing because of the long computation time involved.

Mattison et al. [11] introduced real-time photoacoustic demodulation and PA images using a field programmable gate array (FPGA). They collected and processed a  $1000 \times 1000 \times 1000$  pixel PAM data set in 10.575 s at a pulse repetition rate of 100 kHz. In addition, they stated that the results of simulations conducted indicated that real-time processing was capable up to a line rate of 1 MHz. Finally, the heart of a zebrafish embryo was dynamically and volumetrically imaged with an FOV of 1 mm,  $200 \times 200 \times 200$  pixels, and a period of 1.76 s at a pulse repetition rate of 25 kHz.

In recent times, graphics processing units (GPUs) have been used in the place of central processing units (CPUs) to accelerate numerical computations in optical imaging techniques such as optical coherence tomography, digital holography microscopy, and confocal microscopy [12–15]. This is because a GPU has many stream processors, which can enable highly parallel processing. Consequently, GPU techniques have been applied to integrated photoacoustic tomography (PAT), ultrasound dual-modality systems, and 3D PAT image reconstruction algorithms [16–18]. Yuan et al. [16] and Beán-Ben et al. [17] used a configuration comprising an arrayed ultrasound transducer for detection and a fiber bundle for illumination at a low pulse repetition rate of 10 Hz. They also applied a GPU technique incorporating the back projection algorithm for real-time display [16, 18]. Further, Peng et al. [18] proposed a finite element (FEM) based 3D reconstruction algorithm that exploits GPUs.

In this study, we realized real-time display of 2D B-mode PAM images and MAP PAM images on a laser-scanning OR-PAM at a pulse repetition rate of 300 kHz. Photoacoustic signals were collected frame-by-frame (2D data set) instead of block-by-block (3D data set), and then processed using CPU multithreading technology and GPU parallel processing technology. Respective display rates of 480 fps for B-mode PAM images and 0.98 fps for MAP PAM images at a resolution of  $500 \times 500$  pixels and an FOV of  $3 \text{ mm} \times 3 \text{ mm}$  were achieved.

## 2. Material and methods

### 2.1. Laser-scanning OR-PAM

A schematic of a laser-scanning OR-PAM is shown in Fig. 1(a). Ytterbium-doped fiber laser (YLP-G10, IPG Photonics Corp.), which has variable pulse repetition rate from 20 kHz to 600 kHz with a 1-ns pulse width at 532 nm, was used as a light source. In this experiment, the repetition rate was set to 300 kHz using a function generator. Light from the laser was incident on a single-mode optical fiber via a  $5 \times$  objective lens with 0.13 NA in a fiber launch platform. The light that passed through the optical fiber was scanned using a 2D galvanometer scanning mirror and focused by means of a scan lens (LSM03-VIS, Thorlabs Inc.). The scan lens had an effective focal length of 39 mm, a working distance of 25.1 mm,

and a maximum FOV of  $10.3 \times 10.3$  mm. The focusing beam passed through a glass window with a thickness of approximately 1 mm and was incident on the sample, as shown in Fig. 1(a). The 2D galvanometer scanning mirror was controlled via a data acquisition board (NI USB-6363, National Instruments Corp.) with two saw-tooth functions for the  $x$ -axis and  $y$ -axis. The  $x$ -axis was driven at 480 Hz with a duty cycle of 80%, and the  $y$ -axis was set at 0.98 Hz with a duty cycle of 100%. When the scan angle of all axes was set to  $\pm 1.5$  degree, the measured FOV was approximately  $3 \text{ mm} \times 3 \text{ mm}$ . Geometrically, the FOV should be approximately  $2 \text{ mm} \times 2 \text{ mm}$ . This difference occurred for several reasons: (1) the focal position shifted slightly because the focusing beam passed through a glass window. (2) A distance error occurred between the 2D galvanometer and the objective lens. (3) A scan angle error occurred because the scan angle was controlled by the applied voltage. The acoustic waves that occurred at the focal plane were detected using an unfocused ultrasound transducer (V316-N-SU, Olympus-NDT) with a center frequency of 20 MHz and a  $-6$  dB bandwidth of 71%. The acoustic signals detected were amplified by two amplifiers (ZFL-500LN + , Mini-Circuits) with a gain of 24 dB (total 48 dB) and filtered using a low-pass filter (VLF-52 + , Mini-Circuits). These signals were finally converted using a high-speed digitizer (ATS9350, AlazarTech) at a sampling rate of 250 MSamples/s. The digitizer was able to convert with 12-bit resolution and transfer data to a personal computer (PC) at a speed of 1.6 GB/s. To obtain PAM images, samples were positioned on the transparent glass of the sample stand, as shown in Fig. 1(a).

Figure 1(b) shows the synchronized signals for the line-trigger (green line) of 300 kHz, frame-trigger (blue line) of 480 Hz, and fast-axis ( $x$ -axis, red line) galvanometer. In previous studies, only line-triggers generated by a function generator or a photodiode were used [4–10]. In this study, we added frame-trigger signals to reduce the jitter generated by different main clocks between the function generator, analog-output board, and the digitizer. The fast-axis galvanometer was moved (magenta line) with a delay time of approximately 0.43 ms compared with the function signal from the A/D board, as shown in Fig. 1(b). Consequently, the TTL signal for the frame-trigger was delayed compared with the saw-tooth function for the fast-axis galvanometer. The data acquisition started at the positive edge of the frame-trigger signal. Thus, 736 points per line were sampled by the line-trigger and 500 lines per B-mode-frame were collected.

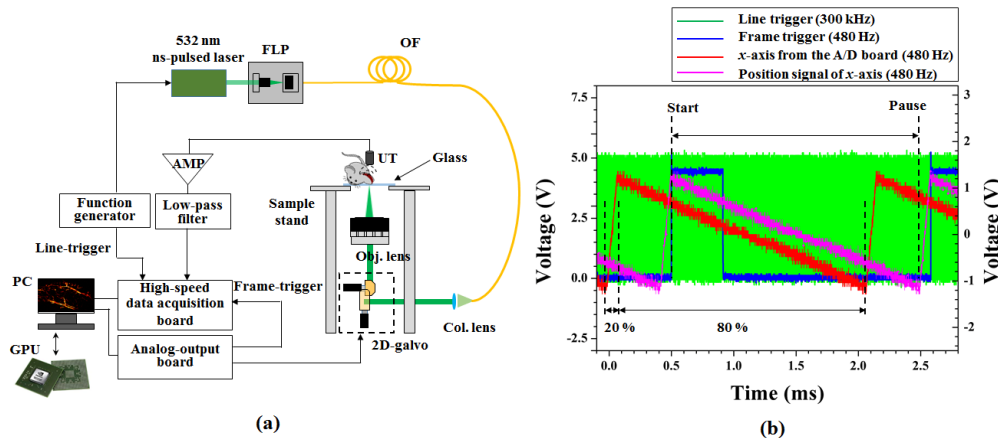


Fig. 1. (a) Schematic of laser-scanning OR-PAM. FLP: fiber launch platform, OF: optical fiber, Col. Lens: collimation lens, Obj. lens: objective lens, UT: ultrasound transducer, and AMP: amplifier. (b) Synchronized signals for line-trigger (green), frame-trigger (blue), fast-axis galvanometer signal (red) from the A/D board, and position signal of the fast-axis galvanometer (magenta).

## 2.2. Parallel signal processing implementation

The sequence flow for the real-time display PAM imaging is shown in Fig. 2. First, thread 1 of the CPU generates an analog-output signal to control the 2D galvanometer, trigger frames, and acquire 736 (axial)  $\times$  500 (lateral) samples per frame acoustic signals. Next, thread 2 of the CPU calls the signal processing sequences of the GPU with five kernels to perform the following functions:

- (1) Type conversion from 16-bit unsigned integer to 64-bit double (Kernel 1)
- (2) Hilbert transform (Kernels 2 and 3, as shown in the dotted box in Fig. 2)
- (3) Computation of the amplitude from real values and imaginary values of the Hilbert-transformed data (Kernel 4)
- (4) Type conversion from 64-bit double to 8-bit unsigned character for a 256 colormap (Kernel 4)
- (5) Find maximum amplitude (Kernel 5)

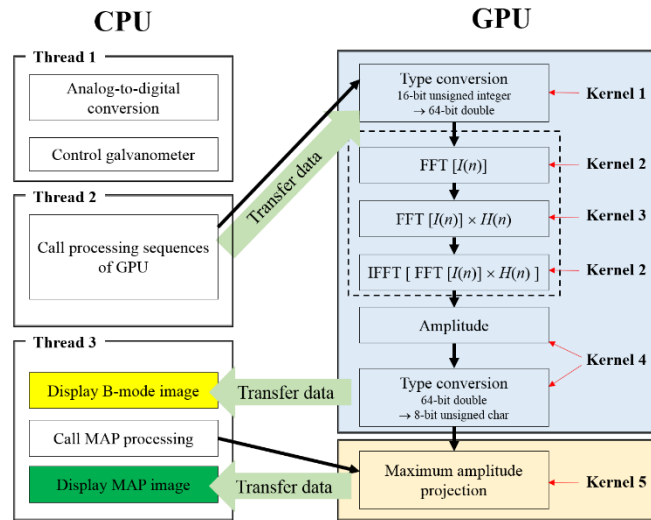


Fig. 2. Signal processing sequence flowchart in the CPU and GPU.

For the Hilbert transform, fast Fourier transform (FFT) and inverse-FFT were carried out using the CUDA FFT library (cuFFT). The Fourier transformed signals were multiplied by  $H(n)$  as follows:

$$H(n) = \begin{cases} 1 & n = 0 \text{ and } n = N/2 \\ 2 & 0 < n < N/2 \\ 0 & N/2 < n < N-1 \end{cases} \quad (1)$$

where  $N$  is the number of sampled even numbers per A-line. The multiplied signals were then converted back to the time domain via inverse-FFT. In thread 3 of the CPU, the B-mode PAM image with 8-bit unsigned character was transferred to the memory of the PC and displayed on the monitor. If the MAP PAM images needed to be displayed on the monitor, then thread 3 of the CPU would call for the processing of MAP on the GPU, transfer the MAP data to PC memory, and display the MAP PAM images on the monitor.

The GPU-based signal processing was implemented as shown in Fig. 3. To obtain B-mode PAM images and MAP PAM images, three types of parallel implementations are needed. The

first is parallel computations of pixel-to-pixel for type conversion (Kernels 1 and 4) and multiplication of  $H(n)$  (Kernel 3), as shown in Fig. 3(a). The second is line-to-line parallel computation for FFT or IFFT (Kernel 2, Fig. 3(b)). The third is line-to-pixel parallel computations to find the maximum amplitude (Kernel 5), as shown in Fig. 3(c).

A personal computer with a quad-core CPU (Intel Core i7-4790, Intel) and a graphics card (ASUS GTX780 Ti, ASUSTeK Computer Inc.) based GeForce series GPU (NVIDIA Corp.), comprising 2880 stream processors, a 7000 MHz memory clock, and 3 GB of RAM were used to process the acquired data. To accelerate numerical calculations and display real-time B-mode and MAP PAM images, we developed custom software using Microsoft Visual C++ and NVIDIA's compute unified device architecture (CUDA) 6.5 technology.

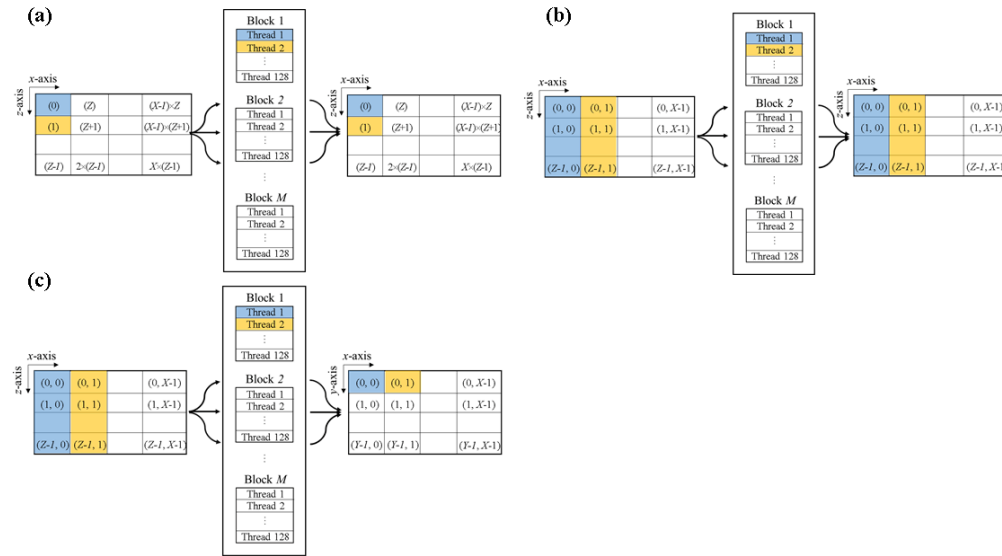


Fig. 3. Illustration of the parallel implementation in the GPU: (a) For Kernels 1, 3, and 4, (b) for Kernel 2 (FFT or IFFT), and (c) for Kernel 5.

### 3. Results and discussion

In GPU processing, the computation time is affected by the number of threads per block. We measured the average time using 500 frames of B-mode PAM images, which had  $736$  (axial)  $\times$   $500$  (lateral) sampling points. In general, NVIDIA recommends that the number of threads per block should be set to multiples of 64. Therefore, we measured the average calculation time according to the changing number of threads per block, 64, 128, 256, and 512, in the GPU. As shown in Fig. 4(a), when the number of threads per block was set to 128, the computation time for the B-mode PAM image was the shortest. Table 1 summarizes the processing time of each sequence when the number of threads per block was 128. The total processing time added to each processing time was 1.35 ms. This computing time was shorter than the frame interval time (2.08 ms). Consequently, real-time display of the processed B-mode PAM images was achieved. Figure 4(b) shows the GPU-accelerated image display speed that we subsequently compared to the image display speed achieved based on a regular CPU. We confirmed that the image display speed based on the GPU was 60 times and 30 times faster than those using single thread and two threads in the CPU, respectively.

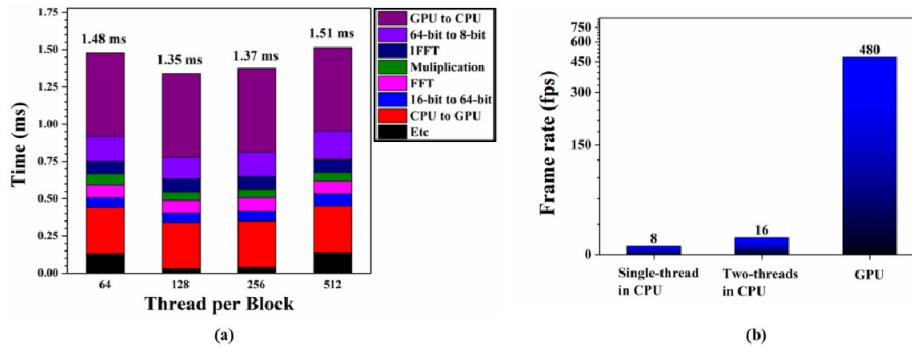


Fig. 4. (a) Computation time versus number of threads per block in the GPU, (b) comparison of the real-time display performance for B-mode PAM images on three different computational platforms: Intel i7 CPU using a single thread, Intel i7 CPU using two threads, and GeForce GTX780 GPU.

To evaluate the performance of our laser-scanning OR-PAM system, a USAF 1951 resolution target (Edmunds Optics, Barrington, NJ, USA), coated negatively with chrome, was imaged as shown in Fig. 5(a). In contrast to the mechanical scanning OR-PAM, laser-scanning OR-PAM has a limited FOV because of the acceptance angle and element size of the ultrasound transducer. In our system, an area  $3 \text{ mm} \times 3 \text{ mm}$  was achieved as the maximum FOV with  $500 \times 500$  pixels. The physical pixel size was calculated as  $6 \mu\text{m}$ . Therefore, theoretically, the minimum lateral resolution was twice the physical pixel size, i.e.,  $12 \mu\text{m}$ . In fact, in Fig. 5(a), the lines at group 5 and element 6, which correspond to a lateral resolution of  $17.5 \mu\text{m}$ , could be distinguished. Further, when the FOV was reduced to  $1 \text{ mm} \times 1 \text{ mm}$  with  $500 \times 500$  pixels, we could clearly recognize the lines at group 7 and element 1, corresponding to a lateral resolution of  $7.8 \mu\text{m}$ , as shown in Fig. 5(b). This value was lower than the theoretical mean spot size of the focused light, which was  $9.9 \mu\text{m}$  at  $633 \text{ nm}$ . For this experiment, when the resolution target was exposed to a laser pulse energy of  $80 \text{ nJ}$ , acoustic signals from the chrome layer were saturated. Under this condition, we measured a signal-to-noise ratio (SNR) of approximately  $23 \text{ dB}$ .

**Table 1. Average GPU processing time using 500 frames of B-mode PAM images when the number of threads per block was 128.**

Processing	Time (ms)
Data transfer (CPU $\rightarrow$ GPU)	0.31
Type conversion (16-bit $\rightarrow$ 64-bit)	0.06
cuFFT	0.09
Multiplication with $H(n)$	0.06
Inverse cuFFT	0.09
Type conversion (64-bit $\rightarrow$ 8-bit)	0.15
Data transfer (GPU $\rightarrow$ CPU)	0.56
etc.	0.03
<b>Total</b>	<b>1.35</b>

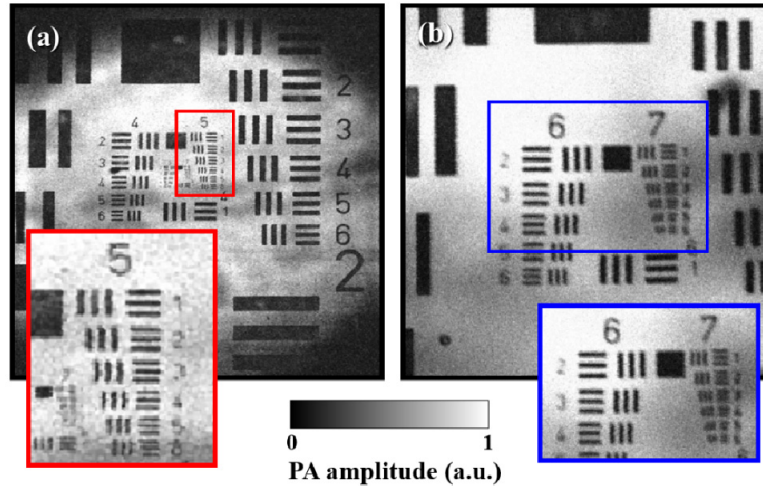


Fig. 5. MAP OR-PAM images of the USAF1951 resolution target: (a) FOV of  $3\text{ mm} \times 3\text{ mm}$  and  $500 \times 500$  pixels, (b) FOV of  $1\text{ mm} \times 1\text{ mm}$  and  $500 \times 500$  pixels.

Finally, we obtained *in-vivo* B-mode PAM images and MAP PAM images of micro-vasculatures in a BALB/c-nude mouse's ear. Figure 6(a) shows the cross-sectional B-mode PAM image at the distal end of the mouse's left ear. [Visualization 1](#) was constructed with 500 B-mode PAM images, which are one volumetric PAM data. These B-mode PAM images were displayed at 480 fps on the monitor, but we converted [Visualization 1](#) to 50 fps. In Fig. 6(a), a mirrored image (red arrows) is presented owing to the reflected acoustic wave from the glass of the sample stand. Figure 6(b) shows the MAP PAM image of blood vessels in the ear. Capillaries (white arrows) can clearly be seen in Fig. 6(b). A total time of 1.02 s was needed to acquire and display one volumetric PAM data with  $500 \times 500$  pixels and FOV of  $3\text{ mm} \times 3\text{ mm}$ . In addition, we were able to obtain a total of 23 frames of MAP PAM images within a period of 23.47 s, which we used to make a movie ([Visualization 2](#)) and converted [Visualization 2](#) to 5 fps.

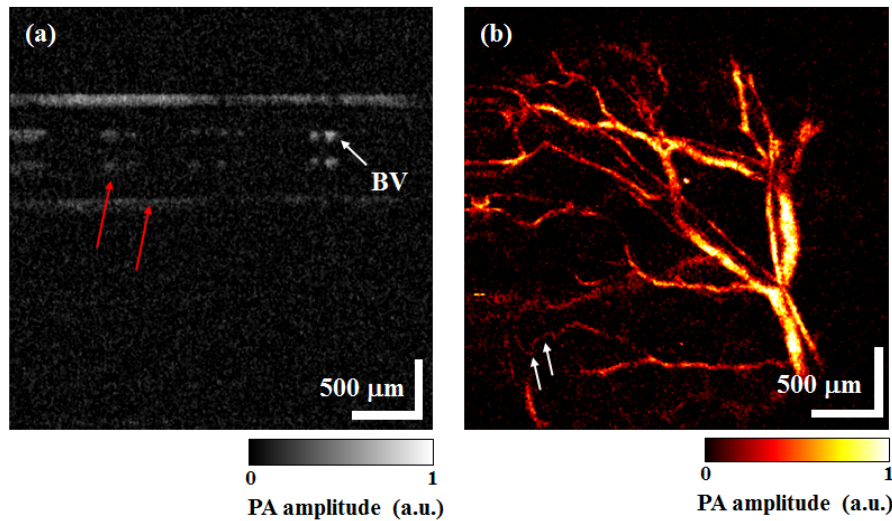


Fig. 6. OR-PAM images of a BALB/c-nude mouse's ear. (a) B-mode PAM image ([Visualization 1](#)), (b) MAP PAM image ([Visualization 2](#)). BV: Blood vessel.



In the *in-vivo* study, the laser pulse energy on the mouse's skin was approximately 100 nJ. Estimating that the depth of the focused spot was  $\sim 180$   $\mu\text{m}$  below the skin surface, the surface spot size was calculated to be approximately 9.3  $\mu\text{m}$ . Therefore, the calculated surface laser fluence was 36.8  $\text{mJ}/\text{cm}^2$ . Although the ANSI laser safety limit is 20  $\text{mJ}/\text{cm}^2$ , the value obtained is below the skin damage threshold [5, 19]. In practice, we observed no thermal damage on the skin surface, as shown in [Visualization 2](#). If the optic and ultrasound beams are aligned confocally and the focused ultrasound transducer is used, the laser pulse energy can be reduced. We believe that this would result in a laser fluence that is below the ANSI safety limit.

If the volumetric PAM image data are saved as a block, only five volumetric PAM images, corresponding to 4.9 s, can be acquired because data size per volume is 351 MB (2 bytes  $\times$  736  $\times$  500  $\times$  500) and the digitizer only has 2 GB of onboard memory. The digitizer can also continuously save data to hard disk in streaming mode. In streaming mode, post-processing should be carried out to display B-mode and MAP PAM images. Therefore, it would be impossible to achieve real-time display in such a scenario. If the pixel size of a MAP PAM image is reduced to 500 (x-axis)  $\times$  250 (y-axis) pixels with an FOV of 3 mm  $\times$  1.5 mm, the display speed can be doubled (1.96 fps).

In the previous study by Mattison et al. [11], the results of simulations indicated that their PAM system using an FPGA was capable of a maximum line rate of 1 MHz. Nevertheless, only volumetric images with an FOV of 1 mm  $\times$  1 mm and 200  $\times$  200  $\times$  200 pixels were obtained at a period of 1.76 s using a laser with a pulse repetition rate of 25 kHz, owing to a limitation of the laser. However, even if the laser was at a pulse repetition rate of 1 MHz and overcame the limitations in obtaining PAM images, the PAM system would still have the disadvantage of a limited imaging depth of  $\sim 1.5$  mm, because the speed of acoustic waves in tissues is approximately 1500 m/s. In our OR-PAM system based on GPU-accelerated signal processing at a pulse repetition rate of 300 kHz, continuous real-time MAP display with an FOV of 3 mm  $\times$  3 mm (9  $\times$  larger) and 736  $\times$  500  $\times$  500 pixels (23  $\times$  larger) is available at a period of 1.02 s (1.7  $\times$  faster). Comparison of the time consumed for parallel processing by the GPU method to that consumed by the FPGA method is rather difficult because parallel processing time is dependent on factors such as data size, processing algorithm, and type of application. Although controversy has arisen regarding the performance of GPU compared to that of FPGA [20, 21], it can be stated unequivocally that the GPU technique has merits such as relative ease of software development without low-level programming language due in large part to its many available programming libraries.

In related studies, Beán-Ben et al. [17] utilized the GPU technique in a real-time PAT system. They used an arrayed ultrasound transducer instead of a single ultrasound transducer for detection because a pulse repetition rate of the laser, which was used, was very slow (10 Hz). Using the arrayed ultrasound transducer should need the back projection algorithm as image reconstruction algorithm. Because the back projection algorithm involves complicated computations, long computation time is required unless parallel processing is accomplished. However, even though the data size was 128  $\times$  128  $\times$  64 pixels (smaller by a factor of 175) and the GPU technique was used, they were only able to realize a computation time 52 times faster (19.5 ms) than that realized in this study.

Continuous real-time MAP display has several advantages as microscopic imaging modality such as monitoring of system performance, immediate confirmation of animal condition, and biological transition. Figure 7 ([Visualization 3](#)) shows the movements of the sample (mouse ear), which can be generated during *in-vivo* studies in both macroscopic and microscopic imaging. These movements occurred image distortion (white arrow, Fig. 7(c)) owing to movement of the sample position or changes of shapes and intensities (white arrows, Fig. 7(d)) owing to movement of focal position in the sample. Although dynamic variance results can be observed using post-processing [8–10, 19] and real-time processing [11] in OR-PAM with a laser at a high repetition rate, movements such as these can be monitored only in

real-time processing and display. Catching these movements is important in *in-vivo* studies and reduces the time wasted as a result of failed experiments. Visualization 3 was also converted from 0.98 fps to 5 fps with 23 frames of MAP images.

The galvanometer scanning mirror used in our OR-PAM system has a trade-off between scan angle and scan frequency. When a large scan angle (large FOV) is needed, the fast-axis scan frequency (B-mode frequency) has to be lower than it is at a small scan angle. On the other hand, if the scan angle of the fast-axis galvanometer is set to a small angle, the scan frequency can be increased. A water-immersible MEMS scanning mirror for wide-field fast-scanning OR-PAM has recently been reported [22–24]. Combining the MEMS scanning mirror with a laser that has a high pulse repetition rate bestows MEMS-OR-PAM with powerful advantages such as fast acquisition with wide scanning range and high SNR owing to the confocal alignment of the optical and acoustic beams; further, it facilitates the application of a handheld probe or endoscopy. However, for volumetric MAP image display, MEMS-OR-PAM still depends on post-signal processing. In further study, we expect to display real-time volumetric MAP PAM images with a larger FOV of  $6\text{ mm} \times 3\text{ mm}$  and  $1000 \times 500$  pixels at 600 kHz pulse repetition rate by applying the GPU-accelerated signal processing technique to the MEMS-OR-PAM system.

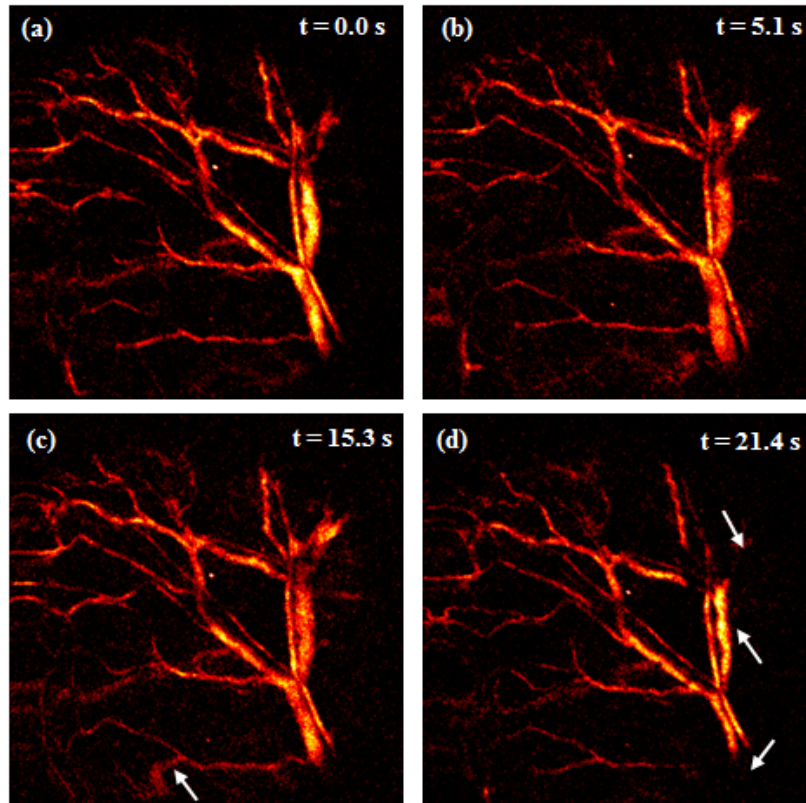


Fig. 7. Image distortion owing to movement of the sample position and changes of shapes and intensities owing to movement of focal position in the sample when there were breathing or momentary movement (Visualization 3).

#### 4. Conclusion

In conclusion, we demonstrated real-time volumetric display in laser-scanning OR-PAM using GPU-accelerated signal processing at a 300-kHz pulse repetition rate. We generated five kernels in the GPU to process Hilbert transform and maximum amplitude projection, and

then optimized the number of threads per block in the GPU. When the number of threads per block was set to 128, the computation time for B-mode PAM image was the shortest. The averaged total time for GPU processing of a B-mode PAM image was approximately 1.35 ms and the display speed was 480 fps when the data samples were acquired with 736 (axial)  $\times$  500 (lateral) points/B-mode-frame. In addition, we were able to display MAP PAM images of a mouse's ear as volumetric images at 0.98 fps (500  $\times$  500 pixels, 3 mm  $\times$  3 mm).

### **Acknowledgments**

This work was supported by the "Development of Nanobio Convergence Measurement Technology Program (GP2015-0034)" from the Korea Research Institute of Standards and Science. It was also supported by grants from the "Pioneer Research Center Program (2012-0009541)" and the "Nano Material Technology Development Program (2014M3A7B6020163)" through the National Research Foundation (NRF) of Korea.