


Five-dimensional ultrasound system for soft tissue visualization

Nishikant P. Deshmukh¹  · Jesus J. Caban³ · Russell H. Taylor¹ · Gregory D. Hager¹ · Emad M. Boctor^{1,2}

Received: 15 April 2015 / Accepted: 29 July 2015 / Published online: 15 August 2015
© CARS 2015

Abstract

Purpose A five-dimensional ultrasound (US) system is proposed as a real-time pipeline involving fusion of 3D B-mode data with the 3D ultrasound elastography (USE) data as well as visualization of these fused data and a real-time update capability over time for each consecutive scan. 3D B-mode data assist in visualizing the anatomy of the target organ, and 3D elastography data adds strain information.

Methods We investigate the feasibility of such a system and show that an end-to-end real-time system, from acquisition to visualization, can be developed. We present a system that consists of (a) a real-time 3D elastography algorithm based on a normalized cross-correlation (NCC) computation on a GPU; (b) real-time 3D B-mode acquisition and network transfer; (c) scan conversion of 3D elastography and B-mode volumes (if acquired by 4D wobbler probe); and (d) visualization software that fuses, visualizes, and updates 3D B-mode and 3D elastography data in real time.

Results We achieved a speed improvement of 4.45-fold for the threaded version of the NCC-based 3D USE versus the non-threaded version. The maximum speed was 79 volumes/s for 3D scan conversion. In a phantom, we validated the dimensions of a 2.2-cm-diameter sphere scan-converted to B-mode volume. Also, we validated the 5D US system visualization transfer function and detected 1- and 2-cm spherical objects (phantom lesion). Finally, we applied the system to a

phantom consisting of three lesions to delineate the lesions from the surrounding background regions of the phantom.

Conclusion A 5D US system is achievable with real-time performance. We can distinguish between hard and soft areas in a phantom using the transfer functions.

Keywords Five-dimensional ultrasound system · GPU · Elastography · Ultrasound · Visualization · 3D elastography · 3D scan conversion · OpenGL · Computer graphics · Medical imaging

Introduction

In 2013, 1,660,290 cancer incidences and 580,350 cancer deaths were reported in the USA [1]. Ultrasound elastography (USE) is an imaging tool often used in the diagnosis and treatment of cancer [2]. USE involves comparing pre- and post-compression ultrasound data to map tissue stiffness [2]. A wide array of applications have emerged using USE such as image-guided thermal ablation monitoring [3], neoadjuvant therapy monitoring [4], and intraoperative robotic surgery guidance [5,6].

US technology is safe, low cost, mobile, and real time, and it emits zero ionizing radiation [3]. However, 2D US has problems in tracking because of the absence of elevation information in the image [3]. The introduction of 3D transducers [3] that can give information about all three orthogonal planes alleviates this problem. USE combined with B-mode scan acts as a complementary technology to reduce the need for unnecessary biopsies [7]. 2D/3D B-mode image lacks the ability to differentiate between the background tissue and the *iso-echoic* tumors because the speckle information and impedance are the same in tumors and surrounding tissues [8]. The 2D/3D USE systems calculate the displacement of

✉ Nishikant P. Deshmukh
nishikant@cs.jhu.edu

¹ Department of Computer Science, The Johns Hopkins University, Baltimore, MD, USA

² Department of Radiology, The Johns Hopkins Medical Institutions, Baltimore, MD, USA

³ National Intrepid Center of Excellence, Walter Reed National Military Medical Center, Bethesda, MD, USA

speckles in the radio frequency (RF) data and generate a strain map; this strain map helps to identify the boundary between the tumor and the surrounding tissues providing more accurate information [2]. 3D USE has been used for diagnosis, monitoring and treatment of breast cancer [9, 10], testicular adrenal rest tumors [11], cervical lymph nodes [12], and ablation monitoring [3, 13, 14].

3D B-mode/USE imaging does not reveal temporal changes in the underlying organ [7], but 4D US with scanning 3D B-mode images over time at the same or different locations does offer real-time, continuous feedback [7]. 4D US has been used in advanced in vivo studies of fetal face expression [15], determination of heart aorta elasticity [16], pelvic floor muscle monitoring [17, 18], fetal heart monitoring [19], and motion tracking of the liver [20], among other applications. 4D volume contrast imaging (VCI) uses contrast enhancing measures to add either elasticity data [21] or color Doppler data [22]. USE has been used in conjunction with B-mode images to increase specificity and maintain the sensitivity of tumor detection [7]. Automatic segmentation of the prostate has been achieved by the fusion and then extraction of contour boundary information from both B-mode and vibro-elastography images [23].

Ablation therapy involves detecting the tumor in a pre-operative CT/MRI scan and then manually registering the location of the tumor intraoperatively using the US B-mode image guidance [3]. Ablation therapy involves inserting an ablation needle under this B-mode image guidance, and ablating the target tumor using RF ablation or high-intensity focused ultrasound (HIFU) [3]. However, there is a risk of incorrect tumor detection and needle placement. Furthermore, the treatment may not ablate the entire tumor area or may extend to affect surrounding healthy tissues [3]. To minimize this risk, 2D/3D USE-guided ablation therapy can offer precise locations of the tumor and the ablated region [3]. However, it is difficult for surgeons to monitor two different feedbacks while performing the surgery. There is a need for overlaid mechanisms to precisely locate needle position. This requirement also exists with a US-guided biopsy to accurately monitor needle placement to extract malignant cells. The feedback needs to be fast enough to allow the surgeon to monitor this process in real time. During thermal ablation monitoring, the ablation process self-induces noise in the collection of ultrasound RF data; therefore, the time window to collect US data is quite small. The system should be fast and efficient enough to collect these data and visualize them over time to increase the accuracy of thermal ablation. Controlling the transfer function would allow distinction between USE and B-mode data because the contour of the organ will encapsulate the lesion/tumor inside it. Finally, fast scan conversion is needed to convert both the USE and B-mode volumes for the 3D wobbler probe that acquires RF data along a spherical sector.

A 5D US system involves the fusion of 3D B-mode images and 3D USE data visualized over time. The live feedback of both 3D B-mode and 3D USE data will improve the early detection and treatment of cancer. 3D B-mode provides information about *hyper-* and *hypo-echoic* tumors [8]. Apart from enhancing the boundary of the *hyper-* and *hypo-echoic* tumors, 3D USE can give information about the *iso-echoic* tumors [8]. The additional strain information from 3D USE offers more diagnostic and monitoring information in terms of shape, size, and position of the lesion [3]. In the 5D US, a multi-dimensional transfer function allows efficient segregation of the B-mode and USE data in a single texture volume. This segregation will allow for advanced segmentation where we can isolate a tumor, cyst, and organ contour information depending on the transfer function as well as a future multi-modality registration [8]. The B-mode modality can potentially be replaced with an advanced US time series tissue typing methods [24, 25]. However, in this paper, we focus on the more ubiquitously used B-mode modality for wider acceptance and clinical trials. Several reports [26, 27] indicate that both specificity and sensitivity are very high considering the relative size of the lesion in strain to that of the B-mode. Malignant tumors tend to infiltrate into surrounding healthy tissue that necessitates real-time visualization with an adapted opacity function to show the extent of strain and B-mode of an all-in-one, 3D-rendered scene.

Contributions

This paper presents, to the best of our knowledge, the first implementation of a real-time 5D US system based on the fusion of 3D B-mode and 3D USE data updated over time. The existing 3D US systems display B-mode and elastography volumes in separate windows, which make them difficult to view, synchronize, and tag 3D B-mode data with strain information. A combined 3D B-mode and elastography visualized over time in 5D US solves this problem. The absence of real-time computation hardware limited the development of essential components of 5D US system such as a sophisticated real-time elastography software, real-time 3D scan conversion, and a real-time visualization system. The contributions reported in this paper include GPU-based real-time 3D elastography using a multi-stream technique, GPU-based real-time 3D volume scan conversion, and a real-time volume renderer that updates every time the 3D USE, and 3D B-mode volume is received over the network using the OpenIGTLinkMusiic library [28, 29]. The definition of a real-time computing system is a system in which all of the components combine to finish the task in a given time [30]. In our case, we define the upper bound as any system that allows a surgeon to operate the system during the real-time thermal ablation monitoring that takes several minutes to perform [3]. A response time of 5–10s would be sufficient for

the 5D US system given that the two RF data volume capture times are in the order of 4–6 s (estimate based on 2D acquisition speed [28]). Through this paper, we attempt to answer the following questions: (1) Is it possible to achieve a real-time 5D US system? and (2) What should we expect the computation time to be?

Traditional general-purpose graphic processing units (GPGPUs) have a single instruction multiple data (SIMD) architecture. The SIMD allowed only one compute-unified data architecture (CUDA) kernel to execute on the GPGPU at a given time. With the advent of Fermi architecture [31] and the above processors, modern GPGPUs now support multiple instructions and multiple data stream capability of facilitating the execution of multiple kernels simultaneously. We use the multi-stream capability of these GPGPUs to give us a faster 3D USE by extending our work in 2D elastography as described in [32]. Scan conversion is essential to convert the data acquired in the polar coordinates by a 3D wobbler probe into the Cartesian coordinates [33]. This paper presents a simple GPGPU-based scan conversion to convert 3D B-mode and 3D USE data in real time. The 3D USE data have been visualized using a volume rendering technique [34–36], but it lacks a real-time data receiver to refresh the volume data. This paper extends the volume renderer in [35] to receive the data in real time.

The 3D B-mode and USE encounter noise artifacts like speckles and de-correlation-induced artifacts. A multi-dimensional transfer function [35] is used to reduce this effect to map B-mode and USE values as well as different color and opacity values. This transfer function gives 5D US the ability to visualize the surface of the tumor and inner–outer surface of the cyst with a particular color and opacity from the USE data. Concurrently, the contour of the organ and biopsy/ablation needle can be visualized using different color and opacity values derived from the B-mode data. Thus, it is necessary to fuse together the 3D B-mode and 3D USE values to complement each other. In addition to a feasibility study about achieving these updates in real time, this paper also presents a validation study on the size of an elastography phantom lesion. While a 5D system based on Doppler ultrasound for monitoring the heart to visualize blood flow already exists, it requires specialized Doppler hardware [37].

General-purpose graphics processing unit (GPGPU)

A GPGPU contains many computation cores that parallel run the similar code. Many of the components such as filters, cross-correlation, and scan conversion can independently work on subsets of data. A GPGPU allows parallelization of such components and offsets the workload from the primary CPU, which allows additional processing for the foreground CPU. Many workstations, including the ones

embedded in the US systems, have onboard GPGPU and also allow the addition of extra GPGPUs via PCI Express slots [32]. The current Ultrasonix (Richmond, BC, Canada) ultrasound machine used in our experiments can acquire RF data at ~ 100 frames/s (fps) [29]. In advanced techniques like parallel beam forming where envelope detection occurs via onboard hardware, the acquisition rate has increased to nearly 860 fps [38], necessitating a real-time GPGPU-based architecture. The design choice of the architecture gives us the flexibility to run various components on the same machine or different machines depending on resource availability. Apart from that, a GPGPU frees the main CPU of the US machine to do other essential tasks; this decreases slowdown of the US system. A GPGPU will frequently be referred to as simply a GPU in the remainder of this paper.

Normalized cross-correlation (NCC)-based elastography on GPU

NCC helps track speckle movement when the palpation motion displaces the tissue. We assume that the direction of palpation motion is parallel to the axial direction of the RF images obtained from the ultrasound acquisition system. We calculate displacement along the axial direction by selecting a template window in a pre-compression RF image and the source window in a post-compression RF image. The template window is searched in the source window to estimate displacement using cosine curve fitting [3]. The outliers are corrected by median and averaging filters. Strain estimation uses linear regression [2]. The following equation defines the NCC score:

$$\begin{aligned} \gamma(u, v) &= \frac{\sum_{x,y} [f(x, y) - \overline{f_{u,v}}] [t(x-u, y-v) - \overline{t}]}{\left\{ \sum_{x,y} [f(x, y) - \overline{f_{u,v}}]^2 \sum_{x,y} [t(x-u, y-v) - \overline{t}]^2 \right\}^{0.5}} \end{aligned} \quad (1)$$

where $f(x, y)$ is the search window that is searched in the template window $t(x, y)$ for the displacement u and v in the x (axial) and y (lateral) directions, respectively. The variables $\overline{f_{u,v}}$ and \overline{t} are the mean of the search and template window, respectively.

Each NCC window comparison can be computed efficiently on the GPU because there is inherent data independence in the processing. This data independence explicitly reduces the need to synchronize the elastography image computation. Similarly, the median filter, average filter, and strain estimation are independent for every pixel calculation and can be efficiently parallelized on the GPU [39].

Methods

This section details the data flow diagram of the real-time 5D US with its subcomponents: 3D USE, 3D B-mode acquisition, 3D scan conversion, and visualization modules. First, we explain the overall system diagram followed by the implementation of 3D USE, 3D scan conversion, and the visualizer. The most common type of 3D probe is a wobbler probe, which has a 2D US array mounted on a motor and is attached to a rotating motor shaft. The 4D probe creates a scan depicting a spherical sector along the elevational direction. The 3D B-mode and USE data need to be scan-converted to reflect the correct shape of the underlying objects.

Five-dimensional ultrasound system

Figure 1 and Algorithm 1 show the overall system diagram and steps needed to create the 5D ultrasound system. The system is highly modular, and each component can run on the same machine or different machines. The OpenIGTLinkMusiic library [28,29] helps us to achieve this modularity. The RF server resides on a US machine and helps to collect real-time 3D RF data and 3D B-mode data. These

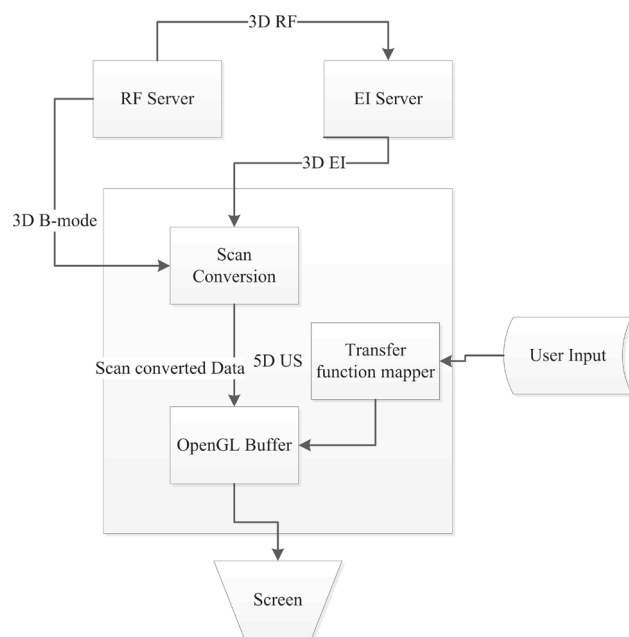


Fig. 1 Data flow diagram of 5D US: the RF server collects the 3D data using a wobbler probe that performs a sector scan using a 2D probe in a particular field of view. The 3D RF data are passed to the elastography image (EI) server, which calculates the 3D USE data and passes them to the scan conversion module in the 5D US system. The RF server also sends the 3D B-mode data directly to the scan conversion module of the 5D US system. The 3D B-mode and the 3D USE scan-converted data are then passed to the visualization system. The users select the transfer function values to highlight different areas of the volume with *different colors*

3D RF data are sent to USE/EI (elastography image) server, and the 3D B-mode data are dispatched to a 5D US visualizer. The hardware synchronizes the 3D RF data and the 3D B-mode data. Thus, it is not necessary to register them separately. After an image pair is received by the 3D EI server, it computes 3D USE to give 3D EI data for 5D US visualization.

If we visualize 3D B-mode and 3D EI as stored in the memory, then the data will appear as a rectangle. However, in our case, a 4D wobbler probe collects the data. This probe requires a 3D scan conversion module to convert the rectangular coordinates to polar coordinates depicting a spherical sector. Thus, a GPU-based scan conversion module is internally embedded inside the 5D visualizer. The scan conversion module sends scan-converted 3D B-mode and 3D EI volume to the OpenGL shading language buffer. We provide the user with an interactive transfer function mapper to draw 2D transfer function. The transfer function assists the ray tracer to assign a color value to the 3D EI and 3D B-mode voxels. The ray tracer module fuses each voxel from the two 3D datasets and displays them on the screen.

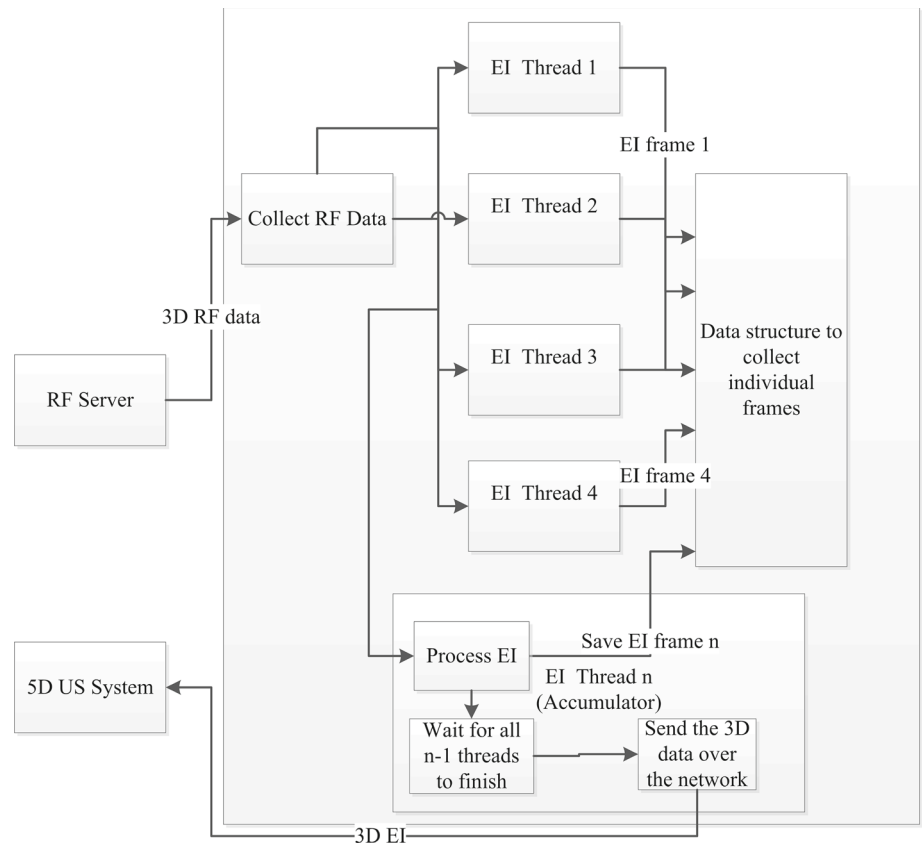
Algorithm 1: Five-dimensional ultrasound system

- 1: The RF server collects the US machine-generated 3D RF data and 3D B-mode data and transfers them to the elastography (EI) server and 5D US system, respectively
- 2: The EI server generates 3D USE/EI data using the GPU by calculating the estimated displacement per slice in the two volumes of the RF data
- 3: The 3D EI data are sent to the 5D US system
- 4: 5D US system receives 3D B-mode and EI data volumes that are scan-converted by a GPU to give a fast volumetric scan conversion
- 5: The scan-converted 3D volumes of 3D B-mode and 3D USE/EI are then passed to the OpenGL shader on the GPU that registers the two volumes
- 6: The ray tracer highlights the color values for pixels in the buffer depending on the transfer function selection by the user
- 7: New data overwrite the buffer, and the pixels are updated as per the transfer function values

Multi-threaded 3D ultrasound elastography

Figure 2 and Algorithm 2 detail how we accelerate 3D USE. An earlier version of the GPU-based 3D EI server ran on a single operating system thread [3]. A thread controls the entire pipeline of the GPU and consists of tasks such as input/output (I/O), memory allocation, kernel function invocation, wait for kernel function invocations, and memory de-allocation [32]. Due to the SIMD architecture of the earlier version of the GPU, this thread could not be executed in parallel because they would wait for the thread to complete their task [32]. In our version, we bind this pipeline together with the help of the CUDA stream functionality similar to a previous report

Fig. 2 3D USE: a block diagram of the 3D USE system that collects the data. The process is distributed among the elastography image (EI) thread to calculate each slice independently. These slices are then collected in the accumulator thread that waits for the remaining threads to finish their task and then sends these USE data as one volume to the 5D US system



[32]. The CUDA streams bind the entire GPU pipeline for elastography consisting of several kernels such as displacement estimation by NCC, moving the average filter, median filter, and strain [32]. These CUDA streams are then assigned to separate threads. The CUDA maintains data independence between streams that implies that the threads do not interfere with each other and provide a robust implementation. The number of threads that create n matches is the number of slices in a scan. The n threads then execute in parallel. If the thread ID starts with 1, we assign a thread with an ID equal to n as an accumulator. The accumulator thread waits for the other $(n - 1)$ threads in a batch to complete execution of elastography for their respective assigned pair in the RF frame (slice) in consecutive RF volumes. After execution is complete, the n threads store their elastography image into a commonly shared buffer with an index as the thread ID. The accumulator thread with ID n after finishing the wait for $(n - 1)$ thread accesses the shared buffer and creates an OpenIGTLink message for volume consisting of n elastography image slices. This thread then sends these data over a TCP/IP network using an OpenIGTLinkMusiiic thread [28, 29].

In the case of 4D wobbler probes, the 2D transducer moves along a spherical sector with a fixed step angle. If the number of slices collected is n , then the field of view of the 3D scan

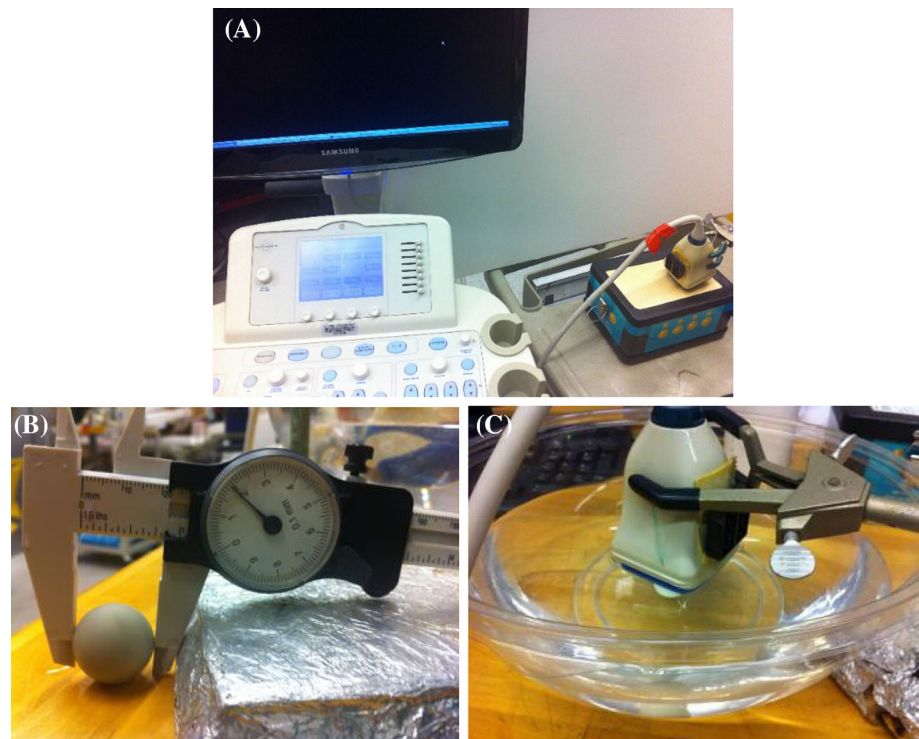
is equal to n times the step angle. However, the 3D image data are stored in the memory as a 2D image per slice in consecutive memory locations. A source transmits these 3D image data over the network. We performed scan conversion from rectangular coordinates to polar coordinates to convert these image data to its correct spherical sector shape from the rectangular form when accumulated together. For efficient and real-time implementation, we map these 3D image data on the GPU and perform scan conversion. This visualizer embeds the scan converter within its structure because of its small execution time. A detailed explanation of the 3D scan conversion is in “Appendix.”

Algorithm 2: Multi-threaded 3D ultrasound elastography

- 1: Receive 3D RF data from the RF server with n slices
- 2: *Spawn* n threads to calculate 2D elastography on the GPU. The components are connected by a CUDA stream to protect the data among the threads
- 3: The GPU threads execute in parallel calculating strain images and scan conversion according to the depth
- 4: The n th thread, at the point of its completion, waits for the $n - 1$ thread to complete its task and waits on the join command
- 5: All $n - 1$ threads deposit their data in the shared buffer, which is combined with the n th thread. This volume of size n slices is sent over the network

Fig. 3 Experimental setup: this figure shows the experimental setup for our experiments.

a Experimental phantom setup where a 4D probe is held by a passive arm on top of CIRS Elasticity 049A QA phantom.
b, c Experimental setup to validate scan conversion of 2.2-cm sphere under water



Five-dimensional ultrasound visualizer transfer function

The y axis corresponds to the grayscale intensities for B-mode images, and the x axis corresponds to the grayscale intensities for the strain (elastography) images. The user can draw various shapes such as an ellipse or rectangle of different sizes. The program then creates a Gaussian circle with a radius as the maximum dimension of the ellipse or rectangle that, in turn, defines the transfer function. These Gaussian circles have centers equal to the center of the corresponding ellipse or rectangle drawn by the user. This transfer function maps different intensity values of B-mode and strain images. The user can also change the opacity of each ellipse or rectangle and create the corresponding transfer function. During the ray tracing, each pixel value is indexed into this transfer function, and the corresponding color and opacity value is assigned to that pixel during the accumulation process of the ray tracer. This ray tracer then displays the combined pixel intensity along a given line in the perspective view. This process is repeated for each line path until the entire volume is complete. The user can zoom in and out. The user can change the position, color, and opacity of each ellipse or rectangle, and the program updates the corresponding 3D volume with new values. This highly dynamic visualization system may help in searching for different features such as strain, needle, and tissue boundaries as well as any foreign objects in the volume depending on the dataset.

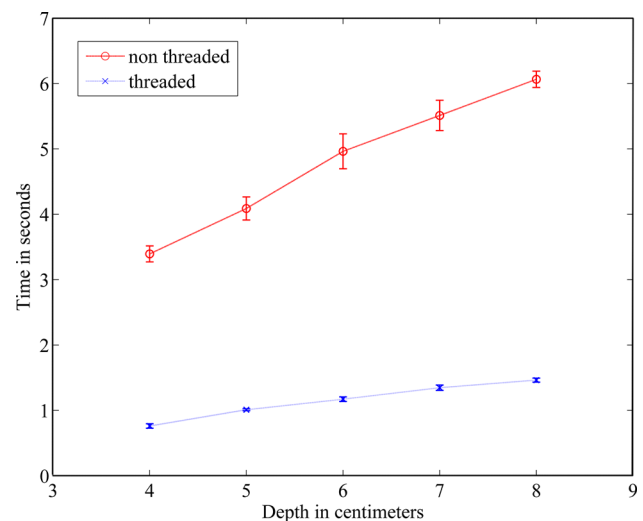


Fig. 4 Performance of non-threaded versus threaded 3D EI: this figure shows the comparison of time taken to compute a volume of elastography for threaded versus non-threaded version to determine whether our algorithm led to any improvement in performance. As indicated in this graph, a threaded version led to a 4.45 \times improvement in speed. This proves that our algorithm has managed to keep the EI volume generation to just below 1 s for 4 cm imaging depth (approximately 0.75 s/volume) and 1.46 s/volume for 8 cm imaging depth

Experiments

This section details the experimental setup, planning, and expected outcomes regarding the experiments and results.

Table 1 Throughput of elastography algorithm: this table lists the throughput of the input RF data volumes for real-time elastography

Imaging depth (cm)	Volume size (kB)	Throughput (kB/s)	
		Non-threaded	Threaded
4	30,720	18,103	80,717
5	38,880	19,017	76,973
6	46,560	18,763	79,461
7	54,240	19,679	80,563
8	61,920	20,419	84,643

The values are multiplied by 2 because generating an elastography image volume needs a pair of RF data. As can be seen, the 8 cm imaging depth for threaded 3D EI gives maximum throughput

We evaluated 3D USE timing information and 5D US visualization on a system with Intel Xeon CPU 2.13 GHz, 12 GB RAM, Windows 7–64 b, NVidia C2070 GPU. We measured the 3D scan conversion timing information on a computer with Intel i7 3.2 GHz, Windows XP 64 b, and an NVidia C1060 GPU. The probe used to collect the 3D RF data is Ultrasonix 4DL14-5/38 attached to the Ultrasonix-CEP machine.

We used a CIRS Elasticity 049A QA phantom with background elasticity of 33 kPa with lesions of varying elasticity. We validated the 5D US system with a 1- and 2-cm lesions. The elasticity of the lesions is 58 and 39 kPa. While scanning the phantom surface, we generated the timing diagram of 3D USE as well as a scan conversion on the 3D B-mode data. The phantom setup is shown in Fig. 3a. The setup used to validate the scan conversion module is shown in Fig. 3b, c.

An offline phantom study has been reported previously [35]. By offline, we mean that the data are collected using a staged robot with a 2D transducer. The stage robot performs a palpation and sends a signal to the US machine after each pre- and post-compression motion. The US machine then collects the corresponding RF data. The B-mode and elastography images in this experiment are calculated offline in MATLAB because the purpose of this experiment is to show the difference between B-mode, elastography, and fused volumes as reported in Fig. 10. For both the 2D and 4D transducer, the number of pixels in the lateral direction of the RF data remains constant at 128 for a 4-cm 2D transducer. The number of pixels in the axial direction changes as a function of depth—this is 1024 pixels for 4 cm, 1296 pixels for 5 cm, 1552 pixels for 6 cm, 1808 pixels for 7 cm, and 2064 for 8 cm of imaging depth. The voxels per pixel are 1. The size of each pixel in the RF data is 16 b.

Results

We next compare the two 3D USE implementations to understand the advantages of *threaded* versus *non-threaded* 3D USE application. We set the timer just before 3D RF data are

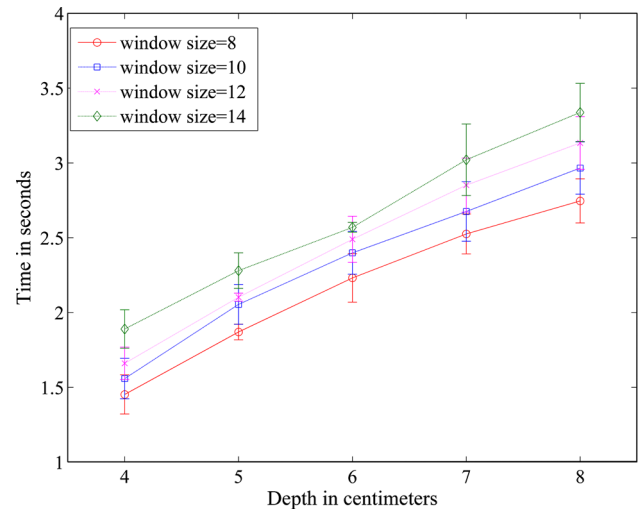


Fig. 5 Performance of non-threaded 3D EI for different window sizes of NCC: this figure shows comparison between different window sizes of 3D NCC. The forward image search is restricted to maximum 2 mm, and window overlap is kept constant at 98 %. It has been observed that the runtime in general decreases with decreasing window size. The average standard deviation (0.13 s/volume) is lowest for window size 8 and 12 and highest (0.16 s/volume) for window size 10. This indicates that the speed for window size 8 is stable and faster than other window sizes. A difference in standard deviation of 0.03 s/volume is notable because the fastest runtime is 1.45 s/volume

passed to 3D USE processing engine, and the program stops the timer just after the USE frames are generated. All of the threads finish the execution. In the case of single-threaded 3D USE application, only one thread calculates the USE for all the slices. Thus, we set the timer after the calculation of the whole volume is completed. Similarly, the speed of the 3D scan conversion is calculated after the thread that calculates the scan conversion has finished.

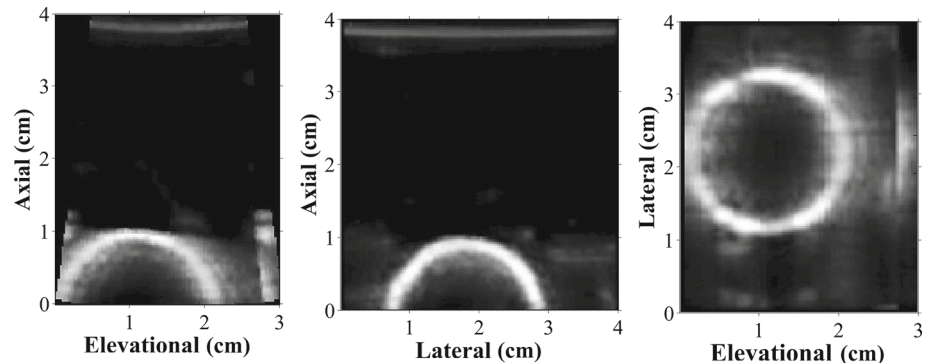
As shown in Fig. 4, the *non-threaded* version of the GPU-based 3D elastography (simply referred as *non-threaded version*) could achieve a maximum averaged runtime of 3.39 s/volume at 4 cm of imaging depth. The volume contains 30 slices. The corresponding average *maximum runtime* for threaded GPU-based 3D elastography (referred to simply as *threaded version*) is 0.12 s/volume at 4 cm depth. The aver-

Table 2 Speed of 3D scan conversion for a B-mode volume: the table lists the speed of real-time 3D scan conversion in volumes per second (vps)

	31 fpv (volumes/s)	60 fpv (volumes/s)	90 fpv (volumes/s)	120 fpv (volumes/s)
4 cm	52.45 (± 2.29)	25.82 (± 0.40)	19.30 (± 0.26)	13.81 (± 0.09)
5 cm	60.00 (± 2.33)	33.33 (± 0.39)	22.91 (± 0.55)	16.18 (± 0.27)
6 cm	66.14 (± 2.33)	38.42 (± 1.55)	25.99 (± 0.16)	17.75 (± 0.31)
7 cm	72.44 (± 3.97)	40.65 (± 0.35)	27.77 (± 0.61)	16.41 (± 0.26)
8 cm	79.40 (± 2.62)	46.18 (± 0.87)	30.89 (± 0.31)	22.50 (± 0.11)

The columns are arranged per number of frames in each volume or frames per volume (fpv). The highest speed is observed for 8 cm imaging depth at 79.40 fpv. The slowest 13.81 vps is observed for 120 fpv at 4 cm imaging depth

Fig. 6 Validation of size after scan conversion: validation of size of an object after scan conversion is performed by imaging a 2.2-cm-diameter sphere inside a water tank. The scan-converted output is shown in the images with an approximate diameter of 2.2 cm in all the three views



age execution time increased gradually with an increase in depth because the number of samples increases as a function of depth. The minimum average runtime recorded for the *non-threaded version* is 6.06 s/volume when collected at 8 cm of imaging depth. We observed that the standard deviation for the timing value was minimal for the threaded version. This indicates a stable execution time where the runtime of generation of volumes remains constant. The maximum average standard deviation in the case of the *threaded version* is 0.041 s/volume at 7 cm of imaging depth. The minimum average standard deviation in the case of the *non-threaded version* is 0.122 s/volume at 4 cm of imaging depth.

This result suggests that the *threaded version* performs better in execution time and stability than the *non-threaded version* of the GPU-based 3D elastography. For 5 cm imaging depth, the threaded version recorded the lowest standard deviation of 0.015 s and the *non-threaded version* recorded the second lowest standard deviation of 0.17 s. The small standard deviation corresponds to an imaging depth of 5 cm—this indicates that 5 cm of imaging depth is an ideal imaging depth for both cases. It results in a stable stream of volumes. Please note that the results are average timing values calculated for just above 200 volumes. The window size for this result of 3D NCC is 12 with 2 mm maximum forward search along the axial direction and 98 % window overlap. As listed in Table 1, the maximum throughput of 84643.02 kB/s is obtained for imaging depth of 8 cm for a

threaded version of elastography. This is slightly better than the average throughput of 80471.80 kB/s for a non-threaded version of the 3D elastography.

In Fig. 5, we further investigated the impact of different window sizes (8, 10, 12, and 14) on the runtime of the 3D NCC volume with constant 2 mm maximum forward search in the axial direction and 98 % overlap. (Please note that the quality of the resulting elastography images changes according to the window size as reported in [32].) Window size is equal to the number of samples in the axial direction used for comparison of source and target images. The runtime of the 3D NCC elastography is best for window size 8; the *minimum time* is 1.45 s/volume. In terms of average time for all imaging depths, the average seconds/volume runtime for window size 8 is 2.16 (± 0.13), window size 10 is 2.33 (± 0.16), window size 12 is 2.45 (± 0.13), and window size 14 is 2.62 (± 0.14). Here, we can see that window size 10 is slightly better than window size 12 and 14, whereas the time is taken for window size 8 is clearly much lower.

The 3D scan converter is needed to reconstruct the geometry for 4D wobbler probes where the 2D array is moving around a fixed axis of rotation along a spherical sector with a limited field of view. As shown in Table 2, the maximum speed observed is 79.40 volumes/s for 8 cm depth for 31 frames per volume. The lowest speed is 13.81 volumes/s for 120 frames per volume at 4 cm depth. This result corresponds to the scan conversion of B-mode volumes. The speed increases with increasing depth in B-mode volume

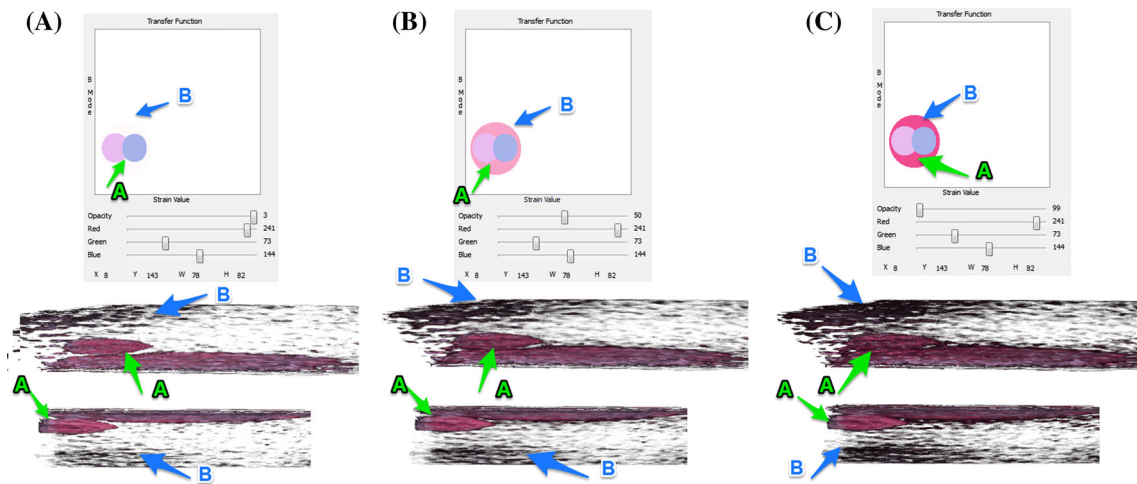


Fig. 7 Impact of changing opacity values: *ellipse A* indicates the region of high strain value where the lesion is found; *ellipse B* indicates the region around *ellipse A*. The opacity value for **a** is set at 3; in **b**, it is set at 50; in **c**, it is set at 100. The *arrow in the output section* indicates the

corresponding regions highlighted by *each ellipse*. As shown in **a**, the low opacity value reduces unneeded noise and B-mode values while the lesion indicated by *arrow A* remains visible

Fig. 8 B-mode and strain volume fused together: region *A* on the transfer function map emphasizes the hyper-echoic region as a spherical region in the output. The rest of the B-mode values are in the surrounding region of the lesion

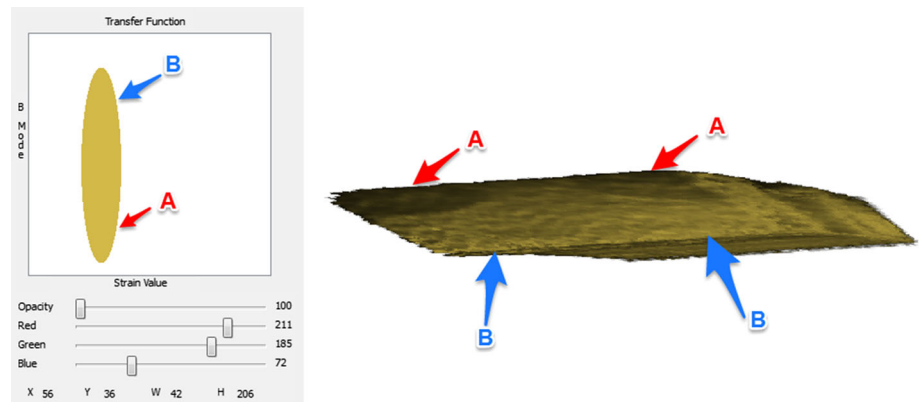
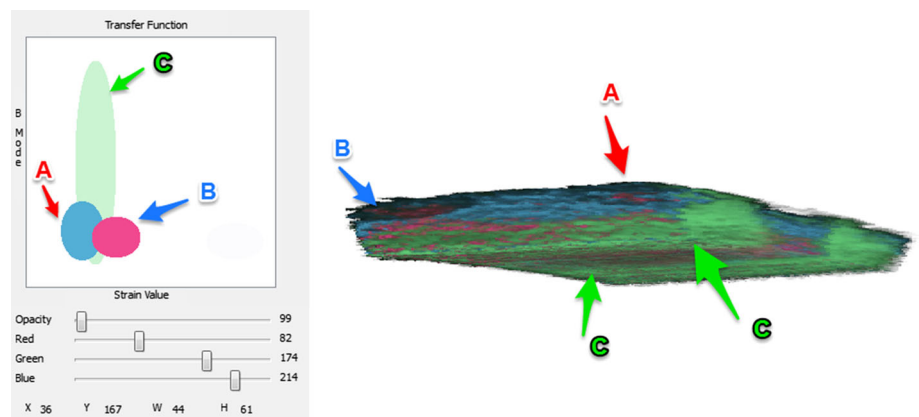


Fig. 9 Differentiating hard and soft lesion: the soft lesion (elasticity 39 kPa) is highlighted with *pink color (ellipse B)* and hard lesion (elasticity 58 kPa) is highlighted with *blue color (ellipse A)*. We can differentiate the soft lesion (label *B*) from the hard lesion (label *A*) with a subtle difference. There is an overlap of *colors* where the soft lesion is partially *green* due to the small difference in elasticity of the lesions



because to adjust the aspect ratio on the display screen the US machine reduces the number of pixels in the lateral direction with increasing depth. Thus, the effective size of the volume decreases with an increase in depth. This leads to an increase in speed of scan conversion as the imaging depth increases.

We validated whether the scan conversion correctly converts the dimensions by imaging a solid sphere of 2.2 cm diameter underwater with B-mode. As shown in Fig. 6, the dimension of the sphere in all three views is found to be approximately 2.2 cm to the scale.

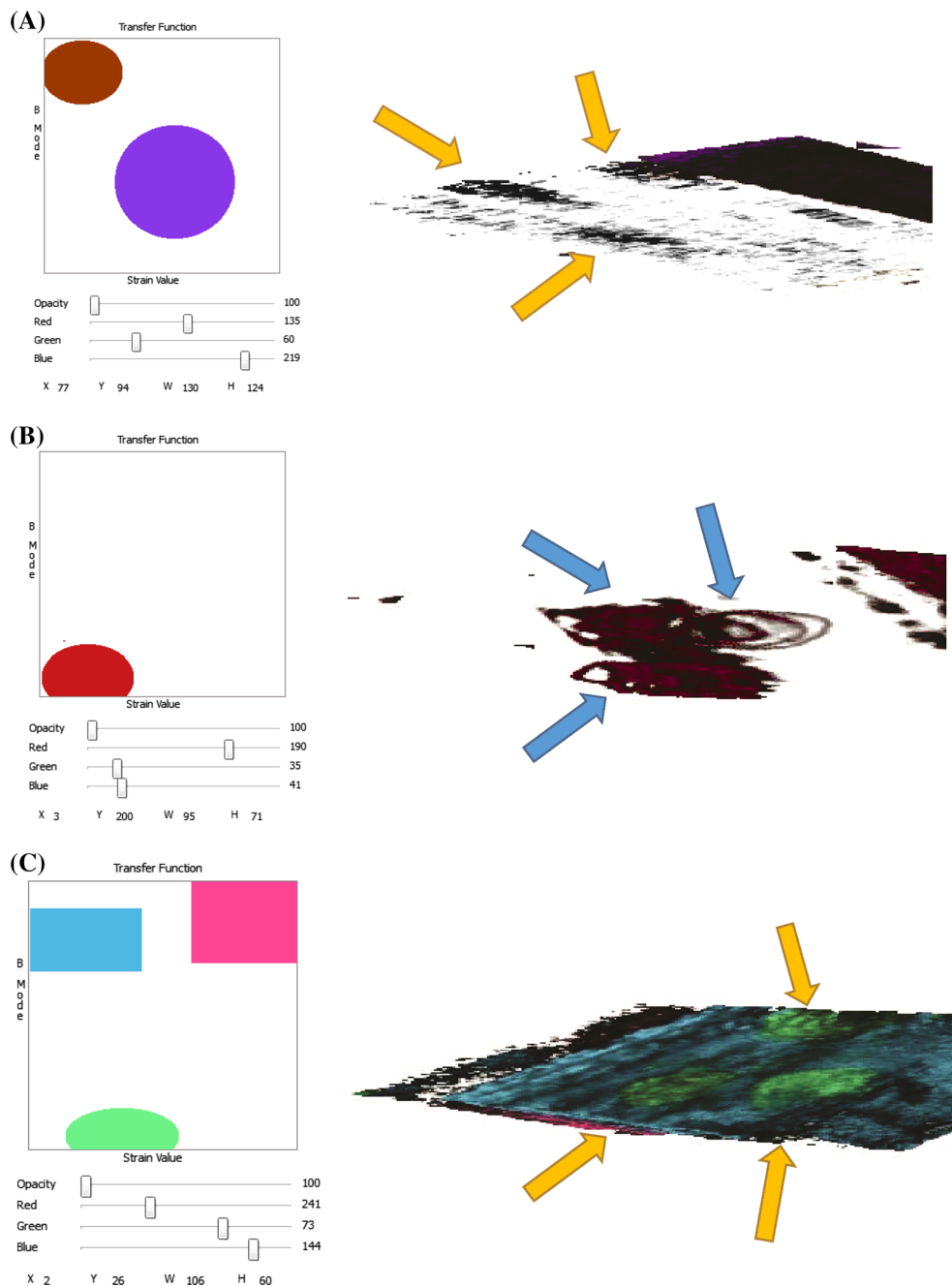


Fig. 10 Differentiating B-mode, elastography, and fused volume: this figure shows volume rendering with different input data types. **a** shows only B-mode volume to display contour of the lesion, **b** shows only elas-

tography volume to show the ablated region, and **c** shows both B-mode and elastography fused volume. In **c**, the lesion region is displayed as *green* region surrounded by background in *blue* and *pink*

Figure 7 shows the effect of choosing different opacity values for ellipse B. As shown in Fig. 7a, the opacity value is set to 3, but it is 50 in Fig. 7b and 100 in Fig. 7c. This shows that region B (indicated by blue arrow) corresponding to ellipse B changes the opacity in the output. The two ellipses indicated by label A have constant opacity, and the lesions are constant in the output (arrow A in the output). The

opacity value of 50 for ellipse B in Fig. 7b reduces the effect of unneeded noise and B-mode data by increasing the transparency. The unneeded noise and B-mode corresponding to ellipse B are further diluted by reducing the opacity to 3. This is indicated by label B in the output. This result suggests that we can keep the strain or elastography values visible and at the same time draw additional circles to illuminate features,

objects, and tissues that might be useful in the corresponding B-mode images. Figure 8 shows a fusion of 3D B-mode and strain values where one ellipse highlights maximum possible dynamic range in 3D B-mode and 3D strain values. The primary strain locations are shown by arrow/label A; B-mode values are indicated by label B.

In Fig. 9, we investigated whether we can differentiate between hard and soft lesions. In this case, the soft tissue/lesion has an elasticity value of 39 kPa, and the harder tissue/lesion has an elasticity value of 58 kPa. As seen in the transfer function, ellipse A corresponds to the hard lesion, and ellipse B corresponds to the soft lesion. From the output, we can see that label B clearly shows a softer lesion. There is an overlap of colors between label A and label B due to the small difference in elasticity between lesions. More rigorous evaluation of different strain strength will be done in future studies.

Figure 10 shows the results of a phantom experiment where three lesions are surrounded by background material. As shown in Fig. 10a B-mode volume, the lesion is slightly visible. In Fig. 10b, an elastography volume is displayed, and the lesion is clearly visible. In Fig. 10c, the fused B-mode and elastography volume are displayed. The selected transfer functions display the lesion region in green and the background regions in blue and pink. This verifies our claim that the volume fusion can improve the feedback of the underlying parameters in B-mode and elastography volumes.

Discussion

We demonstrated the feasibility of a 5D US system by implementing, evaluating, and validating each component of the system. Our highly modular system led to a 5D US system. This end-to-end system facilitated data acquisition from a US machine that was distributed to various elements. The timing graphs in Figs. 4 and 5 demonstrate that our new 3D elastography algorithm is fast and stable. 3D elastography is computationally expensive, and reducing the execution time of a volume comparison was the first step. The threaded elastography that we implemented outperformed the non-threaded elastography version by a factor of 4.45×. The 3D scan converter on a GPU gave a maximum of 79 volumes/s. This high speed and small standard deviation ensured that we could embed the scan conversion module inside the 5D visualizer. The multi-stream 3D elastography is stable and executes on the same computer where the visualizer was running. Because the GPU performs most workload on the GPU, the US system can also execute all parts if equipped with a GPU. Thus, the implementation of a 5D US system in the operating room is a practical solution.

3D USE can be improved by externally tracking the ultrasound probe with an optical tracking system. The tracked

3D USE can be an extension of online tracked USE [32]. The multi-stream 3D USE introduced here can easily accommodate tracking information, and each multi-stream thread can then find in-plane slices among the slices across multiple scans allocated to it. The transfer function can be improvised in multiple ways to study and determine the multi-dimensional transfer functions best suited for the underlying tissue conditions.

Conclusion

We demonstrated the first known implementation of a five-dimensional ultrasound system that involves an end-to-end system for acquisition of 3D RF and B-mode data. We transferred them over the network, calculated the GPU-based 3D elastography, and GPU-based scan conversion of volumes and visualization using multi-dimensional transfer functions. The GPU-based multi-threaded 3D elastography gave us 4.45-fold better performance versus single-threaded GPU-based 3D elastography. We achieved a maximum speed of 79 B-mode volumes/s for 3D scan conversion. We validated the size and shape of the 3D B-mode scan conversion output for a 2.2-cm spherical ball—in 3D elastography, we were able to visualize the 1- and 2-cm-diameter phantom lesion. We then distinguished between the lesion and the surrounding tissue in the phantom using transfer functions.

Acknowledgments NIH SBIR Grant (PI: Everett Burdette) CA134169, NIH Grant CA112852, and The Johns Hopkins University internal funds supported this work. The authors thank Dr. Pezhman Foroughi, Dr. Daniel Carnegie, Dr. Philipp Stolka, Dr. Terry S. Yoo, Dr. Ioana N. Fleming, Balázs P. Vágvölgyi, Dr. Hyun Jae Kang, Joyce Choi, Seth D. Billings, Tina Vashistha, Joseph Wang, and David Mann for their valuable inputs.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: 3D scan conversion on GPU

Scan conversion helps to convert the data collected on the surface of 4D probe that depicts spherical sector, which is an array of images, to the appropriate location in space. This section details the equations used for scan conversion on GPU. We calculate φ , the angle of field of view by

$$\varphi = N \times \delta \quad (2)$$

where N is the number of slices per volume and δ is the step angle of the motor. We then determine the dimension of the scan-converted volume by

$$y' = \frac{x' \times (\tau + y - \tau \times \cos(\frac{\varphi}{2}))}{(2 \times (\tau + y) \times \tan(\frac{\varphi}{2}))} \quad (3)$$

where (x', y') is the new (elevation, axial) size in pixels. In our case, x' is user defined, τ is the radius of the curvature in pixels and y is the number of pixels in the axial direction in the original volume. We then calculate the step size in the axial direction by

$$r_i = \tau + i + 1 \quad (4)$$

where r_i is the step size in the axial direction (pixels) from the origin of the probe for all $i = 0-y$. We then calculate the angle θ_j from the center slice by

$$\theta_j = -\frac{\varphi}{2} + j \times \delta \quad (5)$$

where θ_j is negative or positive on the opposite sides of the center slice, $j = 0-x$, where x is the number of pixels in the elevational direction. Typically, $x = N$ when the slice thickness is 1 pixel. We now calculate the forward index i'' into the final volume along the axial direction by

$$i'' = \frac{x' \times (\text{round}(r_i \times \cos(\theta_j)) - \tau \times \cos(\frac{\varphi}{2}))}{\text{round}(2 + (\tau + y) \times \tan(\frac{\varphi}{2}))}. \quad (6)$$

We now calculate the forward index j'' in the final volume along the elevation direction by

$$j'' = \frac{y' \times ((\tau + y) \times \tan(\frac{\varphi}{2}) + \text{round}(r_i \times \sin(\theta_j)))}{(\tau + y - \tau \times \cos(\frac{\varphi}{2}))}. \quad (7)$$

The following equation gives final scan conversion after substituting values of y' , i'' , and j'' from Eqs. (3), (6), and (7), respectively:

$$\begin{aligned} \text{output} [k \times x' \times y' + i'' \times x' + j''] \\ = \text{input} [k \times x \times y + i \times x + j] \end{aligned} \quad (8)$$

where output is the output volume buffer which will contain the scan-converted volume, input is the input volume buffer which contains the US machine acquired volume, $k = 0-z$, z is the number of pixels in the lateral direction.

GPU kernels execute Eqs. (3)–(8) that are highly parallel since index mappings are independent of each other. In the case of conflict, where multiple pixels from the different locations in the original buffer get mapped to the same pixel in the output buffer, we simply consider the maximum pixel value. There are holes created into the output slices, which are recovered using interpolation by a simple GPU-based averaging filter.

For this scan conversion to work and to keep the equations simple, we need to rotate the volume around an axis at the center of the volume and axis-align to elevation–axial–lateral directions. The rotation depicts swapping of elevation–axial and lateral–axial slices. The following equations give us an easy way to swap the values in the 3D buffers:

$$\begin{aligned} \text{postswap} [j \times y \times x + k \times x + i] \\ = \text{preswap} [i \times y \times z + k \times z + (z - j - 1)] \quad (9) \\ \text{preswap} [j \times y \times x + k \times x + (x - i - 1)] \\ = \text{postswap} [i \times y \times z + k \times z + j] \quad (10) \end{aligned}$$

where pre- and post-swap are the buffers before and after swapping of the pixels. A GPU kernel executes Eq. 9 just before the scan conversion and Eq. 10 after the scan conversion.

References

1. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. *CA Cancer J Clin* 63(1):11–30
2. Ophir J, Céspedes I, Ponnekanti H, Yazdi Y, Li X (1991) Elastography: a quantitative method for imaging the elasticity of biological tissues. *Ultrason Imaging* 13(2):111–134
3. Boctor EM, Deshmukh N, Ayad MS, Clarke C, Dickie K, Choti MA, Burdette EC (2009) Three-dimensional heat-induced echo-strain imaging for monitoring high-intensity acoustic ablation. In: McAleavey SA, D'hooge J (eds) *Medical imaging 2009: ultrasonic imaging and signal processing*, Lake Buena Vista, FL, p 72650R. doi:10.1117/12.811287
4. Falou O, Sadeghi-Naini A, Prematilake S, Sofroni E, Papanicolaou N, Iradji S, Jahedmotlagh Z, Lemon-Wong S, Pignol J-P, Rakovitch E, Zubovits J, Spayne J, Dent R, Trudeau M, Boileau JF, Wright FC, Yaffe MJ, Czarnota GJ (2013) Evaluation of neoadjuvant chemotherapy response in women with locally advanced breast cancer using ultrasound elastography. *Transl Oncol* 6(1):17–24
5. Billings S, Deshmukh N, Kang HJ, Taylor R, Boctor EM (2012) System for robot-assisted real-time laparoscopic ultrasound elastography. In: Holmes III DR, Wong HW (eds) *Medical imaging 2012: image-guided procedures, robotic interventions, and modeling*, San Diego, California, USA, p 83161W. doi:10.1117/12.911086
6. Sen HT, Deshmukh N, Goldman R, Kazanzides P, Taylor RH, Boctor E, Simaan N (2012) Enabling technologies for natural orifice transluminal endoscopic surgery (NOTES) using robotically guided elasticity imaging. In: Holmes III DR, Wong KH (eds) *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*, San Diego, California, USA, p 83161Y. doi:10.1117/12.912383
7. Weismann C, Mayr C, Egger H, Auer A (2011) Breast sonography—2D, 3D, 4D ultrasound or elastography? *Breast Care (Basel)* 6(2):98–103
8. Boctor EM, Choti M, Fictinger G, Taylor R, Prince JL (2011) *Robotic 5-dimensional ultrasound*. US Patent 7901357 B2
9. Torres F, Fanti Z, Arambula Cosío F (2013) 3D freehand ultrasound for medical assistance in diagnosis and treatment of breast cancer: preliminary results. In: IX international seminar on medical information processing and analysis, p 89220K

10. Sayed A, Layne G, Abraham J, Mukdadi O (2013) Nonlinear characterization of breast cancer using multi-compression 3D ultrasound elastography in vivo. *Ultrasonics* 53(5):979–991
11. Jedrzejewski G, Ben-Skowronek I, Wozniak MM, Brodzisz A, Budzyska E, Wieczorek AP (2013) Testicular adrenal rest tumors in boys with congenital adrenal hyperplasia: 3D US and elastography—do we get more information for diagnosis and monitoring? *J Pediatr Urol* 9(6):1032–1037
12. Ying M, Zheng Y-P, Kot BC-W, Cheung JC-W, Cheng SC-H, Kwong DL-W (2013) Three-dimensional elastography for cervical lymph node volume measurements: a study to investigate feasibility, accuracy and reliability. *Ultrasound Med Biol* 39(3):396–406
13. Rivaz H, Fleming I, Assumpcao L, Fichtinger G, Hamper U, Choti M, Hager G, Bector E (2008) Ablation monitoring with elastography: 2D in vivo and 3D ex vivo studies. In: Metaxas D, Axel L, Fichtinger G, Székely G (eds) *Medical image computing and computer-assisted intervention (MICCAI 2008)*, vol 5242. Springer, Berlin, pp 458–466
14. Foroughi P, Burgner J, Choti MA, Webster III RJ, Hager GD, Bector EM (2012) Towards intra-operative monitoring of ablation using tracked 3D ultrasound elastography and internal palpation. In: Bosch JG, Doyley MM (eds) *Medical imaging 2012: ultrasonic imaging, tomography, and therapy*, San Diego, California, USA, p 83200T. doi:10.1117/12.913988
15. Kanenishi K, Hanaoka U, Noguchi J, Marumo G, Hata T (2013) 4D ultrasound evaluation of fetal facial expressions during the latter stages of the second trimester. *Int J Gynecol Obstet* 121(3):257–260
16. Wittek A, Karatolios K, Bihari P, Schmitz-Rixen T, Moosdorf R, Vogt S, Blase C (2013) In vivo determination of elastic properties of the human aorta based on 4D ultrasound data. *J Mech Behav Biomed Mater* 27:167–183
17. Hilde G, Staer-Jensen J, Siafarikas F, Gjestland K, Ellström Engh M, Bø K (2013) How well can pelvic floor muscles with major defects contract? A cross-sectional comparative study 6 weeks after delivery using transperineal 3D/4D ultrasound and manometer. *BJOG* 120(11):1423–1429
18. Braekken IH, Majida M, Engh ME, Bø K (2009) Test–retest reliability of pelvic floor muscle contraction measured by 4D ultrasound. *Neurourol Urodyn* 28(1):68–73
19. Yagel S, Cohen SM, Shapiro I, Valsky DV (2007) 3D and 4D ultrasound in fetal cardiac scanning: a new look at the fetal heart. *Ultrasound Obstet Gynecol* 29(1):81–95
20. Vijayan S, Klein S, Hofstad EF, Lindseth F, Ystgaard B, Lango T (2013) Validation of a non-rigid registration method for motion compensation in 4D ultrasound of the liver. In: *IEEE 10th international symposium on biomedical imaging*, pp 792–795
21. Gritzmann N, Weismann CF, Datz L (2007) Diagnostic algorithm: how to make use of new 2D, 3D and 4D ultrasound technologies in breast imaging. *Eur J Radiol* 64(2):250–257
22. Achiron R, Gindes L, Zalel Y, Lipitz S, Weisz B (2008) Three- and four-dimensional ultrasound: new methods for evaluating fetal thoracic anomalies. *Ultrasound Obstet Gynecol* 32(1):36–43
23. Mahdavi SS, Moradi M, Morris WJ, Goldenberg SL, Salcudean SE (2012) Fusion of ultrasound B-mode and vibro-elastography images for automatic 3D segmentation of the prostate. *IEEE Trans Med Imaging* 31(11):2073–2082
24. Moradi M, Abolmaesumi P, Mousavi P (2010) Tissue typing using ultrasound RF time series: experiments with animal tissue samples. *Med Phys* 37(8):4401
25. Uniyal N, Eskandari H, Abolmaesumi P, Sojoudi S, Gordon P, Warren L, Rohling RN, Salcudean SE, Moradi M (2015) Ultrasound RF time series for classification of breast lesions. *IEEE Trans Med Imaging* 34(2):652–661
26. Itoh A, Ueno E, Tohno E, Kamma H, Takahashi H, Shiina T, Yamakawa M, Matsumura T (2006) Breast disease: clinical application of US elastography for diagnosis. *Radiology* 239(2):341–350
27. Burnside ES, Hall TJ, Sommer AM, Hesley GK, Sisney GA, Svensson WE, Fine JP, Jiang J, Hangiandreou NJ (2007) Differentiating benign from malignant solid breast masses with US strain imaging. *Radiology* 245(2):401–410
28. Kang H-J, Deshmukh NP, Stolka P, Burdette EC, Bector EM (2012) Ultrasound imaging software framework for real-time monitoring of acoustic ablation therapy. In: Bosch JG, Doyley MM (eds) *Medical imaging 2012: ultrasonic imaging, tomography, and therapy*, San Diego, California, USA, p 83201E. doi:10.1117/12.912362
29. Stolka P, Kang H, Bector E (2010) The MUSiiC Toolkit: modular real-time toolkit for advanced ultrasound research. <http://hdl.handle.net/10380/3172>
30. Shin KG, Ramanathan P (1994) Real-time computing: a new discipline of computer science and engineering. *Proc IEEE* 82(1):6–24
31. Heinecke A, Klemm M, Bungartz H-J (2012) From GPGPU to many-core: Nvidia Fermi and intel many integrated core architecture. *Comput Sci Eng* 14(2):78–83
32. Deshmukh NP, Kang HJ, Billings SD, Taylor RH, Hager GD, et al (2014) Elastography using multi-stream GPU: an application to online tracked ultrasound elastography, in-vivo and the da Vinci surgical system. *PLoS One* 9(12):e115881. doi:10.1371/journal.pone.0115881
33. Pospisil E, Rohling R, Azar R, Salcudean S (2010) 4-D × 3-D ultrasound: real-time scan conversion, filtering, and display of displacement vectors with a motorized curvilinear transducer. *IEEE Trans Ultrason Ferroelectr Freq Control* 57(10):2271–2283
34. Bector EM, Matinfar M, Ahmad O, Rivaz H, Choti M, Taylor RH (2009) Elasticity-based three dimensional ultrasound real-time volume rendering. In: Miga MI, Wong KH (eds) *Medical imaging 2009: visualization, image-guided procedures, and modeling*, Lake Buena Vista, FL, , p 72612V. doi:10.1117/12.815166
35. Mann D, Caban JJ, Stolka PJ, Bector EM, Yoo TS (2011) Multi-dimensional transfer functions for effective visualization of streaming ultrasound and elasticity images. In: Wong KH, Holmes III DR (eds) *Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, Lake Buena Vista, FL, p 796439. doi:10.1117/12.878935
36. Weis JA, Yankeelov TE, Munoz SA, Sastry RA, Barnes SL, Arlinghaus LR, Li X, Miga MI (2013) A consistent pre-clinical/clinical elastography approach for assessing tumor mechanical properties in therapeutic systems. In: Weaver JB, Molthen RC (eds) *Medical imaging 2013: biomedical applications in molecular, structural, and functional imaging*, Lake Buena Vista (Orlando Area), FL, USA, p 86721F. doi:10.1117/12.2007425
37. Heid V, Evers H, Henn C, Glombitza G, Meinzer H-P (2000) 5D interactive real time Doppler ultrasound visualization of the heart. *Med Imaging* 2000:494–500
38. Takahashi H, Hasegawa H, Kanai H (2013) Improvement of automated identification of the heart wall in echocardiography by suppressing clutter component. *Jpn J Appl Phys* 52(7S):07HF17
39. Deshmukh N, Rivaz H, Bector E (2009) GPU-based elasticity imaging algorithms. In: *Proceedings of the international conference in medical imaging computing and computer assisted intervention*