Direct three-dimensional ultrasound-to-video registration using photoacoustic markers

Alexis Cheng
Jin U. Kang
Russell H. Taylor
Emad M. Boctor

Direct three-dimensional ultrasound-to-video registration using photoacoustic markers

Alexis Cheng,[a] **Jin U. Kang,**[b] **Russell H. Taylor,**[a] **and Emad M. Boctor**[a,c]
[a]Johns Hopkins University, Department of Computer Science, 3400 North Charles Street, Baltimore, Maryland 21218
[b]Johns Hopkins University, Department of Electrical and Computer Engineering, 3400 North Charles Street, Baltimore, Maryland 21218
[c]Johns Hopkins University, Department of Radiology, 3400 North Charles Street, Baltimore, Maryland 21218

Abstract. Modern surgical procedures often have a fusion of video and other imaging modalities to provide the surgeon with information support. This requires interventional guidance equipment and surgical navigation systems to register different tools and devices together, such as stereoscopic endoscopes and ultrasound (US) transducers. In this work, the focus is specifically on the registration between these two devices. Electromagnetic and optical trackers are typically used to acquire this registration, but they have various drawbacks typically leading to target registration errors (TRE) of approximately 3 mm. We introduce photoacoustic markers for direct three-dimensional (3-D) US-to-video registration. The feasibility of this method was demonstrated on synthetic and *ex vivo* porcine liver, kidney, and fat phantoms with an air-coupled laser and a motorized 3-D US probe. The resulting TRE for each experiment ranged from 380 to 850 μm with standard deviations ranging from 150 to 450 μm. We also discuss a roadmap to bring this system into the surgical setting and possible challenges along the way. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JBO.18.6.066013]

Keywords: video registration; planning and image guidance of interventions; oncology applications; photoacoustic imaging; optical imaging; clinical ultrasound.

Paper 130074R received Feb. 8, 2013; revised manuscript received Apr. 30, 2013; accepted for publication May 15, 2013; published online Jun. 17, 2013.

1 Introduction

Interventional guidance systems are becoming increasingly common in modern surgical procedures including open, laparoscopic, and robotic surgeries.[1] During such procedures, surgeons can lose track of tumors as they move in and out of the camera's field of view. Guidance systems can be used to alleviate these concerns by providing a fusion of video with other imaging modalities, such as intraoperative ultrasound (US), to aid the surgeon in locating tumors or other regions of interest. The clinical utility of these guidance systems depends on the registration of other surgical tools and devices with the guidance system, such as stereoscopic endoscopes and US transducers.

The registration between US images and video visualization remains a significant challenge. Typically, electromagnetic (EM) or optical navigational trackers[2,3] are used to provide real-time position and orientation about tools such as US probes and endoscopic cameras. These navigational trackers, usually track sensors or markers, relative to a separate base station placed within the surgical setting, adding complexity. However, this approach has serious limitations and is subject to error buildup from multiple tracking and calibration errors.

EM-based surgical navigation and tracking systems[3,4] are the most common choice for laparoscopic surgery, flexible endoscopy, and other minimally invasive procedures because a clear line of sight is not required between the base station and the attached sensors (see Fig. 1). However, there are several drawbacks associated with using an EM-based system. First, wired EM sensors must be placed on the US transducer before it can be

tracked by the surgical navigation system. This could decrease the surgeon's comfort while potentially increasing the cost associated with handling and sterilizing modified surgical tools. Second, a large intrusive EM base station must be placed in close proximity with the operating table, adding clutter to the surgical setting. Third, EM-based systems also suffer from magnetic field distortions when metallic objects are placed within its field. This drawback is particularly significant as the tracked sensors will have significant errors when this is the case.

Optical tracking systems such as those developed by Claron Technology Inc. (Toronto, Ontario, Canada) or Northern Digital Inc. (Waterloo, Ontario, Canada) avoid the field distortion problems associated with EM trackers and frequently do not require wired sensors. Studies have shown that optical trackers can detect their optical markers with submillimeter accuracy,[5,6] but line-of-sight restrictions often make them impractical for laparoscopic procedures. Markers can certainly be placed outside the body, but this will degrade the tracking accuracy for long and flexible tools inserted into the body.

Another drawback for both optical and EM-based systems is the need to acquire the transformation from the tool to the surgical navigation system indirectly, i.e., the transformation of interest is computed via a chain of transformations over several coordinate systems. This case applies to both EM and optical tracker-based systems. As an example, a transformation of interest would be one that goes from the EM base station's coordinate system to the US image's coordinate system. To acquire this transformation, one must obtain both the transformation from the EM base station to the EM marker and the transformation from the EM marker to the US image. The first transformation is given by the tracking information from the EM base station and the second must be obtained through a calibration

---

Address all correspondence to: Alexis Cheng, Johns Hopkins University, Department of Computer Science, 3400 North Charles Street, Baltimore, Maryland 21218. Tel: 410-516-4639; E-mail: acheng22@jhu.edu
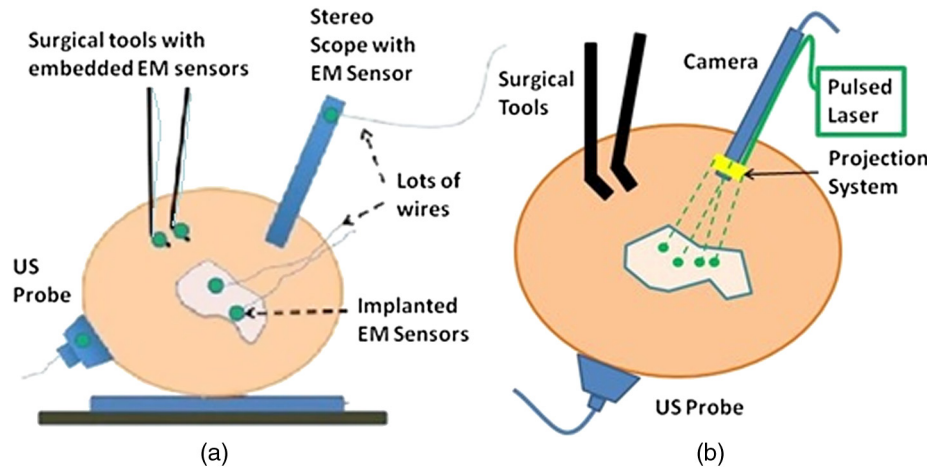
Fig. 1 (a) Standard electromagnetic-based navigation system and (b) photoacoustic (PA) navigation system.

process. Calibration is a topic where many authors have presented research to achieve better accuracy and lower errors.[7,8] Their results have shown that the calibration process dominates the overall error in the registration.[5–7] Other studies have presented overall registration errors of approximately 1.7 to 3 mm for artificial phantoms and 3 to 5 mm for tissue.[3,4,9,10]

Yip et al.[10] demonstrated a registration method that utilizes a tool at the air and tissue boundary. This tool had optical markers in the stereo camera (SC) space on one side and US-compatible fiducials on the other. A drawback of this method is that it requires a custom registration tool to be in direct contact with the tissue. Also, the US fiducials must be segmented from US B-mode images. This can be a difficult task as the speckle or structural information contained in US B-mode images may obscure the US fiducials.

Vyas et al.[11] demonstrated proof of concept for a direct registration method with the photoacoustic (PA) effect requiring a single transformation between the frames of interest as opposed to a chain of transformations. This method addresses each of the drawbacks above. Markers or sensors are not necessary to generate a coordinate transformation between the tracker frame and the US transducer frame, so the tools that surgeons use will remain the same. Previous work[12,13] showed that a pulsed laser source can effectively generate a PA signal in tissue, resulting in an acoustic wave that can be detected by conventional US transducers.[14,15] Only the US transducer needs to touch the surface. Each laser point projection was seen as a green spot in the SC space and as a PA signal in the US space. Segmentation of the PA signal is also simpler in a PA image than a US B-mode image because the laser spot was now the only acoustic source. Finally, the calibration process is unnecessary since the coordinate transformation from the SC frame to the US frame can be computed directly with the two three-dimensional (3-D) point sets based on rigid registration algorithms.[2,16]

Our work extends the earlier work of Vyas et al.[11] and our recent conference publication[17] to take a step toward realizing the system shown in Fig. 1(b). The main objective of our project is to establish a direct registration between US imaging and video. Several surgeries require real-time US including liver resections, partial nephrectomies, and prostatectomies, and real-time fusion of US and video is crucial to their success. Our approach is to create virtual fiducial landmarks, made of light, at the air–tissue interface. A projection system will be used to project these landmarks onto the surface of the organ

through air. At the air–tissue interface, these landmarks can be seen both in US with the PA effect and in video. We present a direct 3-D US to video registration method and demonstrate its feasibility on *ex vivo* tissue. Improving on the work of Vyas et al.,[11] we used a 3-D US transducer instead of a two-dimensional (2-D) US transducer to detect the PA signal. Using a 3-D transducer allows this registration method to function for a nonplanar set of 3-D points. This is a significant requirement because we aim to deploy this method in a laparoscopic environment and organ surfaces will rarely form a planar surface. We also improved significantly on the point-finding algorithms used by Vyas et al.[11] to find the PA signal in both SC images and US volumes. In addition to using a synthetic phantom with excellent light absorption characteristics, we also used resected *ex vivo* porcine liver, kidney, and fat each individually embedded in gelatin phantoms to demonstrate this method's feasibility for the eventual guidance of laparoscopic tumor resections and partial nephrectomies. These phantoms are representative of our proposed clinical scenario since the laser light will likewise only interact with the surface of the phantoms. The gelatin acts purely as a support material and does not affect the PA signal generation. This paper provides more detailed information about our processing methodology, a demonstration of the ability to generate a PA point signal, an analysis of the point localization errors (LEs) in various phantoms, and registration error results on two additional *ex vivo* tissue phantoms as an extension to our conference publication.[17]

This paper details the experimental setup, the processing methodology, and three key results: the ability to generate a PA point signal, a comparison of the point LEs in the SC and US domains, and registration error results. We will discuss the significance of our results, potential errors, and a detailed roadmap to eventual implementation in a surgical setting along with future directions.

2 Methods

In these experiments, we used a Q-switched neodymium-doped yttrium aluminum garnet (Nd:YAG) Brilliant (Quantel Laser, France) laser to generate a PA marker on various materials. We used different combinations of wavelengths (532 and 1064 nm) and energy densities ($\sim$6, 19, 64, 172, and 45 mJ/cm^2) with specifics for each experiment indicated in Sec. 3. These values do not represent the lowest possible energy

**Table 1** Observed laser energy densities in different scenarios.

	1	2	3	4	5
Phantom	Synthetic	Liver	Kidney	Fat	Fat
Wavelength (nm)	532	532	532	532	1064
Energy density (mJ/cm^2)	6	19	64	172	45
Maximum permissible exposure (mJ/cm^2)	19.5	19.5	19.5	19.5	97.5
Signal-to-noise ratio	3.5	2.7	4.1	3.6	3.8

necessary to generate a PA signal on our materials. We chose these values to give our PA images a sufficient signal-to-noise ratio (SNR) without having to average over multiple frames. The SNR for some sample images can be seen in Table 1. The images are first normalized and then the SNR is computed as the mean of the foreground divided by the standard deviation of the background, where the foreground and background are separated by a threshold described later. It should be noted that the maximum permissible exposure (MPE) is 19.5 mJ/cm^2 for 532 nm and 97.5 mJ/cm^2 for 1064 nm as calculated from the IEC 60825-1 laser safety standard[18] based on a 0.25 s exposure time, a 4 ns pulse width, and a frequency of 10 Hz. We used a SonixCEP US system and a 4DL14-5/38 US transducer developed by Ultrasonix Medical Corporation (Richmond, Canada) to scan the volume of interest. This US transducer has a motor that actuates a linear US array to move angularly around an internal pivot point. This US transducer has a bandwidth of 5 to 14 MHz and the linear array is approximately 38 mm in length. The Sonix DAQ device, developed by the University of Hong Kong and Ultrasonix, and the MUSiiC toolkit[19] are used to acquire prebeamformed radiofrequency (RF) data from the US machine. The k-wave toolbox[16] in MATLAB (Mathworks Inc. Natick, Massachusetts) is used to beamform and reconstruct PA images based on the prebeamformed RF data. For our SC setup, we used a custom system containing two CMLN-13S2C cameras (Point Grey Research, Richmond, Canada) to capture images at 18 Hz. A camera calibration process using the Camera Calibration Toolbox for MATLAB[20] generates a calibration file for our SC setup. These calibration files contain the SC setup intrinsic parameters to do 3-D triangulation. We created several phantoms for these experiments: a synthetic phantom made with plastisol and black dye, an *ex vivo* liver phantom made with gelatin and a freshly

resected porcine liver, an *ex vivo* kidney phantom made with gelatin and a freshly resected porcine kidney, and an *ex vivo* fat phantom made with gelatin and porcine fatback. The surface of the *ex vivo* tissue or fat is exposed from the gelatin.

Each of our experiments was split into a data collection phase, a data processing phase, and a registration phase. The data collection phase outputted SC image pairs, five images for each camera, and a 3-D RF US volume for each projected laser spot. The data processing phase used the data and generates a 3-D SC point set and 3-D US point set. Finally, the registration phase used the two point sets to compute a coordinate transformation from the SC frame to the US frame. Figure 2(a) shows the experimental setup using the porcine liver phantom and an overlay of a US image representation using the inverse of the computed transformation.

Figure 3(a) shows the workflow of the data collection phase. First, a laser spot is projected onto the exposed surface of the *ex vivo* tissue, *ex vivo* fat, or synthetic material. It is important to emphasize that most of the laser energy from these laser spots are absorbed at the surface of the phantom. There are inaccuracies in SC spot triangulation if the laser spot is projected at or near the tissue or fat gelatin interface because the laser spots become irregularly shaped when projected onto clear materials. When projecting onto the fat, there was a significant reflectance and saturation due to its color. To reduce these effects, we placed laser goggles in front of the cameras, acting as a filter that modulates light intensities with varying wavelength-dependent optical densities. Second, several images were taken with each camera. The laser spot projected onto the phantom must be visible in at least one image per camera for triangulation to be possible. Our cameras had a faster capture rate than our laser's repetition rate, so some of the frames were devoid of the laser signal. We exploit this during data processing. Steps 3
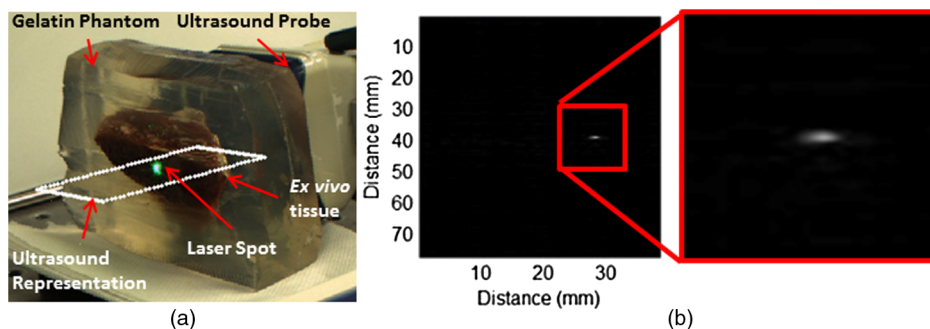
Fig. 2 (a) Experimental setup and video overlay, (b) PA signal within an ultrasound (US) image.
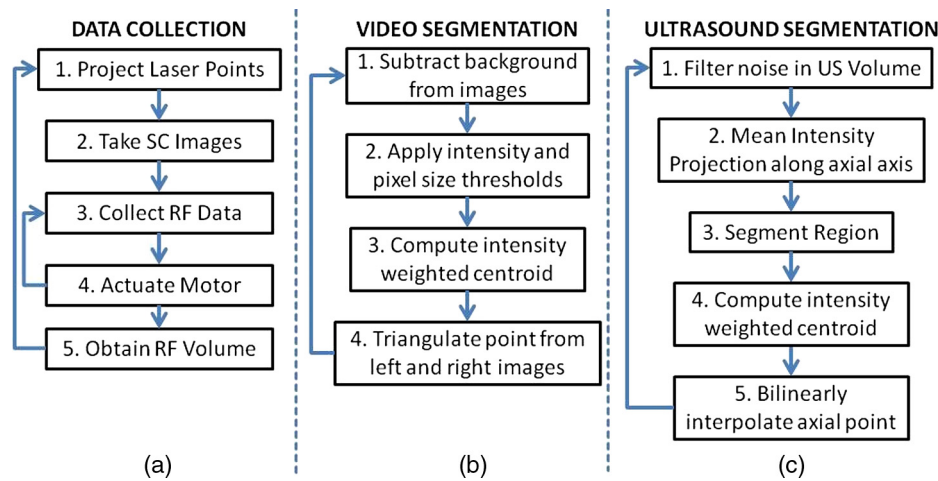
DATA COLLECTION

1. Project Laser Points
2. Take SC Images
3. Collect RF Data
4. Actuate Motor
5. Obtain RF Volume

(a)

VIDEO SEGMENTATION

1. Subtract background from images
2. Apply intensity and pixel size thresholds
3. Compute intensity weighted centroid
4. Triangulate point from left and right images

(b)

ULTRASOUND SEGMENTATION

1. Filter noise in US Volume
2. Mean Intensity Projection along axial axis
3. Segment Region
4. Compute intensity weighted centroid
5. Bilinearly interpolate axial point

(c)

Fig. 3 Workflow for (a) data collection, (b) SC segmentation, and (c) US segmentation.

and 4 show that the 3-D US transducer motor actuation and RF data were intermittently collected from the DAQ device to scan and acquire the RF data of the volume of interest. The motor step size was 0.49 deg. The volume's field of view varied among the experiments because the phantoms were of different sizes and several of them did not require as many slices to generate a volume that covered the entire phantom. The probe has a lateral length of 38.4 mm and the resulting PA images have a depth of 8 cm. A real-time implementation is feasible and an automatic process is in development. Currently, the manual process takes approximately 3 s/slice. This workflow was repeated for each of the laser spots.

From the data collection phase, we present the results showing that we can generate a PA point signal on the various phantoms using 532 nm, 1064 nm, or both. In each of these cases, the laser energy is mostly absorbed at the surface and the laser light does not interact with any medium other than air and the surface of the phantom. As we described, only a small amount of energy per pulse is required to generate a single PA spot and to localize it at the air–tissue interface. These are shown as a series of images similar to the one shown in Fig. 2(b). The size or location of the PA spot is unrelated between scenarios as parameters such as spot size or laser energy were different. Since we are

displaying a 2-D slice from a 3-D volume, it is also possible that the displayed image does not contain the centroid of the PA signal.

The data processing phase involved the segmentation of the SC images into 3-D SC points, the segmentation of the 3-D RF US volume data into 3-D US points, and the computation of the transformation from the SC frame to the US frame.

Figure 3(b) shows the workflow for SC segmentation. For each camera, we picked a SC image with the laser spot and without the laser spot. Next, the background images without the laser spot were subtracted from the images with the laser spot. As we can see from Fig. 4(a), the laser spot is nearly segmented from the scene. We used an iris to decrease the beam size of the laser and the reflection from the iris results in the laser spot in the lower left. To compensate for this spot, we neglected an appropriate border on each image. We then applied an intensity and pixel size thresholds such that the laser spot is segmented out. These thresholds were selected based on the laser beam diameter and the phantom's reflectance and are varied between the different scenarios. The thresholds were selected manually, but a method to automate threshold selection based on experimental parameters is in development. Next, we fitted an ellipse to the segmented region and computed the intensity-weighted centroid
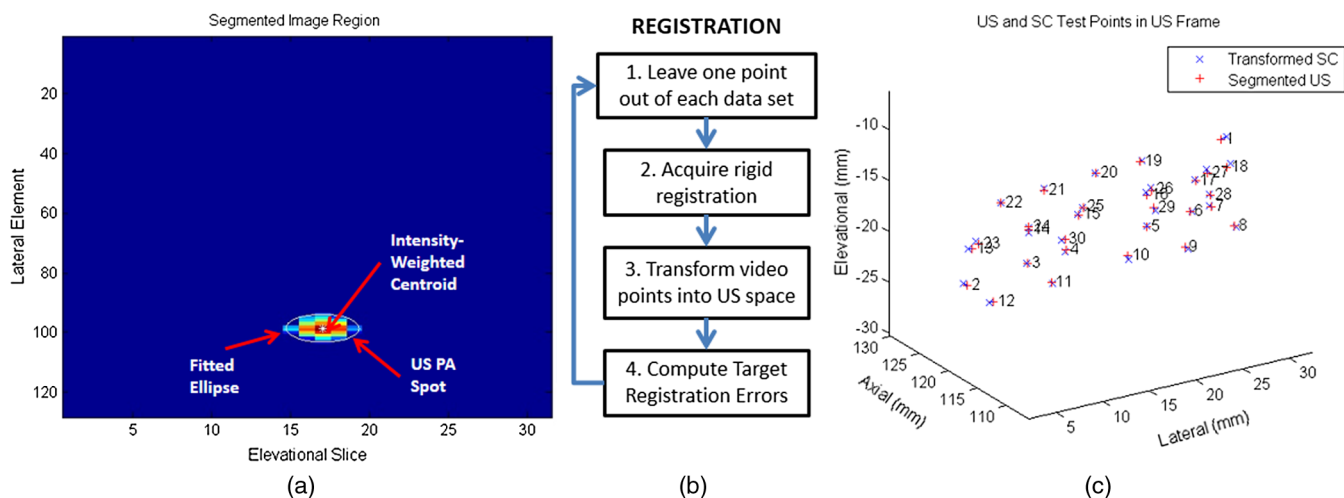

REGISTRATION

1. Leave one point out of each data set
2. Acquire rigid registration
3. Transform video points into US space
4. Compute Target Registration Errors

(a)

(b)

(c)

Fig. 4 (a) Resulting image of US segmentation workflow steps 2, 3, and 4, (b) workflow for transformation, and (c) US and SC point cloud registered together.

based on the pixels within the ellipse from the original image, resulting in the image shown in Fig. 4(b). Calibration files for our specific SC allowed us to triangulate the segmented point from each camera and to obtain a single 3-D point in the SC frame. In our current implementation, this step takes approximately 500 ms for each laser spot projection. This workflow was repeated for each laser spot projection.

The workflow for the segmentation of the 3-D RF US volume is shown in Fig. 3(c). First, for each slice of a 3-D RF US volume, the RF data were beamformed using the k-wave toolbox[16] in MATLAB. The dynamic range of the image was normalized with respect to the volume, and we applied a threshold to decrease the size of the PA signal seen in each volume. These thresholds ranged from 0.4 to 0.5. The size of the PA signal refers to the number of nonzero valued pixels, representing the PA signal after thresholding. A smaller size leads to a more compact representation of the PA signal, but it is also important to maintain the characteristic elliptical shape of the PA signal. Figure 2(b) shows a k-wave beamformed PA signal image. Beamforming requires approximately 140 ms for each PA image. Next, we projected the volume onto the lateral–elevational plane by taking the mean along each axial ray. An intensity and pixel size threshold were then applied to this image. These thresholds were selected in a similar fashion to the ones used for SC segmentation. An ellipse was fitted on the segmented region and an intensity-weighted centroid was computed resulting in lateral and elevational coordinates. Figure 5(a) is an example of this step showing the lateral–elevational image and the corresponding ellipse. As described earlier, the PA signal originated from the surface and any penetration into the surface. Since air cannot generate a PA signal in our setup, we conjecture that the high intensity pixels farthest away in the axial direction are from the surface. Thus, we obtained the axial coordinate corresponding with a lateral–elevational coordinate as the axial-most high intensity pixel. This step is particularly important because the penetration of the laser pulse was deeper for *ex vivo* tissue than the penetration for the synthetic phantom because the laser energy was not completely absorbed at the surface. We used bilinear interpolation to obtain axial coordinates between sampled points. These three coordinates were converted to 3-D US coordinates based on transducer specifications. The lateral coordinate combined with the lateral resolution of our US transducer results in the lateral coordinate in 3-D US space. The axial coordinate combined with the axial resolution represents a ray protruding from the US transducer, whereas the elevational coordinate relates to the angle in which this ray interacts with the first US image in the volume. Solving this geometry problem results in a 3-D US coordinate. The computation time for this process, not including beamforming, is correlated to the number of slices and field of view in each volume. Thus, the synthetic phantom required approximately 730 ms/volume and the other phantoms required approximately 570 ms. A 3-D RF US volume was acquired for each PA spot.

From the data processing phase, we present the localization accuracy of the US PA signal segmentation. In these results, we assume that the segmented SC points are the ground truth. For each US point set, we compute the distance between each pair of points. We compare the US distance of a particular pair of points with the distance between the same pair of SC points. The average of all differences for all point pairs is the recorded LE. These are reported in each of the lateral, axial, and elevational axes and the overall Euclidean distance for each experiment.

The transformation from the SC frame to the US frame was computed with the 3-D SC and 3-D US point sets as shown by the registration workflow in Fig. 5(b). Any registration method for computing the transformation between two 3-D point sets can be used. We used the coherent point drift algorithm[21] in our experiments. One of the main reasons for using coherent point drift was that it allows for the data points to be missing from either dataset. An assumption that we have made is that every laser spot will be visible in the SC images and every PA signal will be visible in the US volume. This assumption was valid for our experiments but may not hold in the surgical setting due to SC or US transducer movement. Some of the points may lie outside the field of view of either the SC or the US transducer. The coherent point drift registration algorithm allowed us to acquire a registration as long as there were enough corresponding points in the SC images and the US volume. More experiments will be necessary to determine how many points are enough and if there will be any corresponding tradeoff in accuracy.

In Fig. 5(c), each SC point was independently used as a test point, whereas the rest of the points in the point set were used to transform the SC test point into the US domain. The figure shows the test SC point alongside its corresponding US point
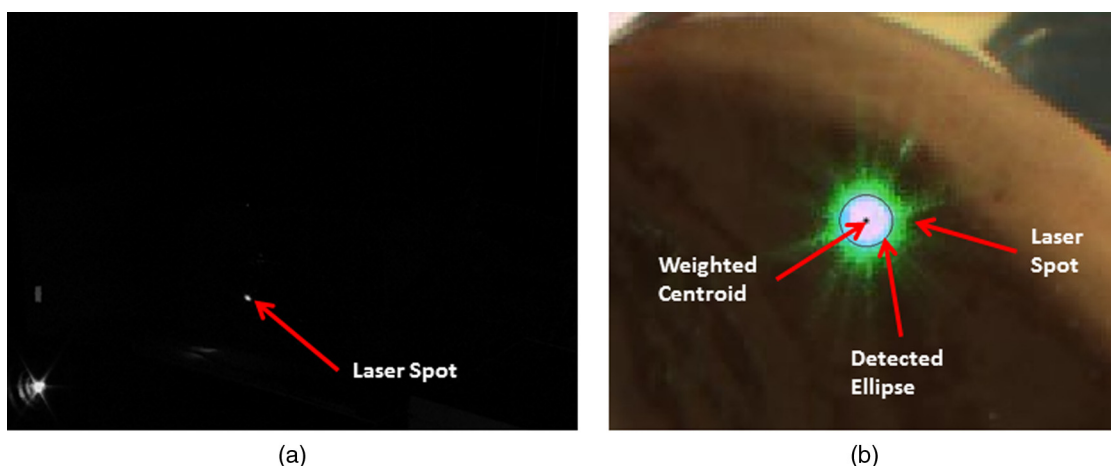

Fig. 5 Resulting images of video segmentation workflow for (a) step 1 and (b) step 3.

with a vector connecting the two. Our results from the registration workflow are shown as a series of target registration errors (TRE) computations in each of the lateral, axial, and elevational axes and the overall Euclidean distance for each experiment.

The transformation from the SC frame to the US frame was used to transform the 3-D SC points to the US frame for validation. The inverse transformation was used to display a representation of US image into the SC frame as shown in Fig. 2(a).

3 Results

Three sets of results from various points in our experiment are presented: the ability to generate a PA point signal, a comparison of the point LEs in the SC and US domains, and registration error results.

The first set of results indicates that we can generate a PA point signal on a variety of materials using various energy levels or wavelengths. Table 1 specifies each of the five scenarios that we tried. Figure 6 shows the US images displaying a PA signal corresponding with each of the scenarios in Table 1.

The second set of results includes the LEs of the PA points in scenarios 1 to 4 outlined in Table 2. Scenario 5 is not included because the 1064 nm wavelength light is invisible to the SC. This means that scenario 5 is not applicable to our experiments with the current design. The metric that we use is defined in Eq. (1). As mentioned previously, we compute the difference of the distance between a point pair in the US space versus the distance between a point pair in the SC space. This metric treats the SC points as the ground truth. The reported LE is the average of these differences over all point pairs for each experiment.

$$LE = |(\vec{SC}_i - \vec{SC}_j)| - |(\vec{US}_i - \vec{US}_j)| \qquad (1)$$

The registration results of our experiments on the synthetic phantom, the *ex vivo* liver phantom, the *ex vivo* kidney phantom, and the *ex vivo* fat phantom are validated using the TRE metric defined in Eq. (2). F_{SC_US} is the transformation from the SC frame to the US frame and is computed with all of the SC and US points except for one. The TRE is the difference between the actual US test point and the transformed SC test point in the US frame.
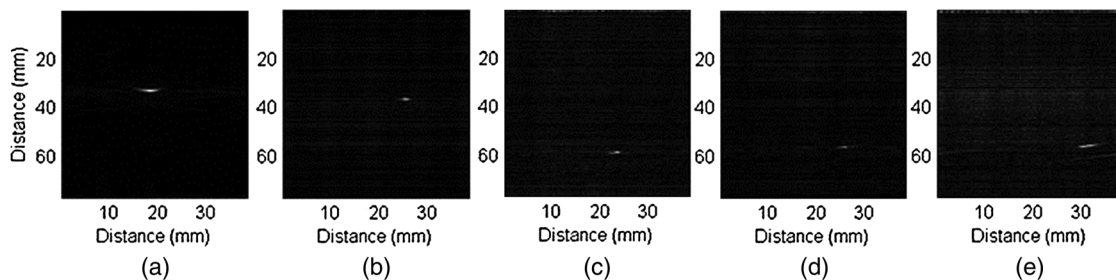
$$\vec{TRE} = \mathbf{F}_{SC\_US} * \vec{SC}_{test} - \vec{US}_{test} \qquad (2)$$

N is the number of points in a specific experiment and $N - 1$ points is used to compute the transformation from the SC frame to the US frame. The remaining point is used as a test point to compute the TRE. This computation is repeated with each of the N points as test points. Table 1 shows the average and standard deviation of the TRE results for the N cases in the synthetic phantom, the *ex vivo* liver phantom, the *ex vivo* kidney phantom, and the *ex vivo* fat phantom experiments, respectively. Figure 7 shows the TRE results in a box-whisker plot.

4 Discussion

As seen in Table 1, the energy densities used in the *ex vivo* liver and kidney experiments are close to or exceeding the MPE at that laser wavelength. These energy densities do not present a concern as we were not using the lowest possible energy density to generate the PA effect. Thus, energy density levels below the MPE threshold are quite feasible. Additionally, averaging multiple US PA images at a lower laser energy density can allow us to retain the same SNR at the expense of time. The energy density at 532 nm used in the *ex vivo* fat experiment does present a legitimate concern as it was an order of magnitude higher than the MPE. Fat is also the most likely material encountered in a realistic surgical setting as it covers the organ of interest and surgeons try to remove as little of it as possible. One situation where this method may benefit from an *in vivo* setup as opposed to an *ex vivo* setup is the presence of blood. At a wavelength of 532 nm, blood has a significantly higher absorption coefficient than fat.[22,23] It is therefore possible that the blood on the fat surface be used to generate the PA signals as opposed to the fat itself. This would also mean that a much lower laser energy density is necessary.

As seen in Table 1, the energy density at 1064 nm used in the *ex vivo* fat experiment was beneath the MPE threshold. While we avoid energy density level concerns, this presents a situation where the SC used must be receptive to 1064 nm light. This is usually not the case as cameras typically have a visible range of 400 to 900 nm. A possible solution is to project a continuous wave, low power, visible laser that is coincident to the 1064 nm
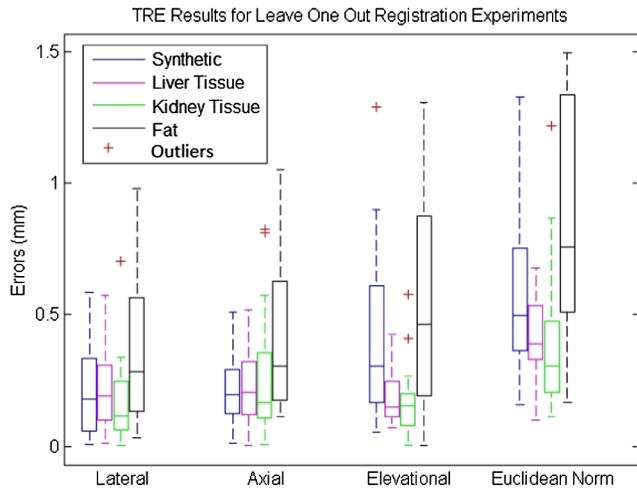

Fig. 6 Sample PA images for (a) scenario 1, (b) scenario 2, (c) scenario 3, (d) scenario 4, and (e) scenario 5.

Table 2 Average LE results for experiments.

	Synthetic phantom	*Ex vivo* liver phantom	*Ex vivo* kidney phantom	*Ex vivo* fat phantom
Number of point pairs	435	435	190	55
Average distance (mm)	0.46 ± 0.33	0.32 ± 0.23	0.19 ± 0.15	0.45 ± 0.39

Fig. 7 Box-whisker plot of TRE results for leave one out registration experiments.

laser. However, this solution presents a new source of error as it may be difficult to achieve coincidental points. Another possible solution is to use a wavelength that is within the visible range of typical cameras yet has a higher absorption coefficient for fat than that observed at a wavelength of 532 nm.

The results in Table 2 imply that the US PA spots are being localized fairly accurately. The average and standard deviation of the difference in distance between all point pairs in the US and SC domain are submillimeter for each experimental scenario. These errors are in line with the point source LEs for typical SC systems.[5,6] Future studies for determining the LE may require a more accurate representation of ground truth data. A possible study is to project the laser spots on specific known locations of the phantom.

At the level of error measurements shown in Table 3, it is likely that the calibration of the SC system is a significant contributor. Optical markers can be tracked at submillimeter accuracy,[5,6] so this error is usually negligible in comparison with the approximate 3 mm errors from calibration.[5,6,7] Since our results were 0.56, 0.42, 0.38, and 0.85 mm errors, respectively, the SC system's error became significant. We used a custom SC system, so its errors were also likely greater than a commercial SC system.

From the experimental results shown in Table 3 and Fig. 7, it can be seen that our system achieved submillimeter TRE measurements for each of our experimental scenarios with different phantoms. We wish to highlight that these results are significantly better than the overall registration errors of approximately 1.7 to 3 mm for artificial phantoms and approximately 3 to 5 mm for tissue[3,4,9,10] presented in literature. There are several differences in the results between each scenario. First, the synthetic phantom had a larger Euclidean error than the *ex vivo* liver and *ex vivo* kidney phantoms almost entirely due to the elevational error. This was likely due to the larger field of view in the synthetic phantom experiment as well as normal variation across experiments. More experiments must be performed to obtain an average error across multiple experiments. The *ex vivo* fat experiment had noticeably worse results in both the mean and standard deviation in each direction and the Euclidean norm. A possible reason for this is the light delivery system. To achieve the energy density necessary to generate a PA signal on fat, we needed to focus our laser beam. Our setup was such that the focusing of the laser beam was not uniform across all points. As mentioned before, laser goggles were also placed in front of the SC for this particular experiment. These two circumstances differ from the other scenarios and may have introduced inconsistencies in the SC or US point sets. Another possibility may be the smaller sample size of this experiment compared to the rest.

There are several considerations when discussing this system's deployment in our intended applications of laparoscopic tumor resections. The first is the placement of the transducer. In our experiments, we used a relatively large 3-D US transducer that would be difficult to place inside the body during a laparoscopic procedure. However, the transducer is often placed externally[3,10] in these procedures, so the size of the probe is not an issue. Naturally, there are disadvantages of placing the transducer externally and farther from the region or organ of interest. The SNR of US images degrades as the depth increases, which would likely lead to errors in localizing fiducials or, in our case, the PA signal. However, since the PA signal only has to travel in one direction, as opposed to traditional US, our PA images will have better quality than US images of equivalent depth.

Another issue with our 3-D US transducer was the acquisition speed. There are certain applications where an acquisition speed of a volume per several seconds is sufficient, but a real-time implementation would require a higher rate. We anticipate

Table 3 Average TRE results for leave one out registration experiments.

	Synthetic phantom	*Ex vivo* liver phantom	*Ex vivo* kidney phantom	*Ex vivo* fat phantom
Number of points (N)	30	30	20	11
Field of view (deg)	19.6	14.7	14.7	14.7
Lateral (mm)	0.21 ± 0.17	0.22 ± 0.16	0.17 ± 0.16	0.36 ± 0.28
Axial (mm)	0.21 ± 0.13	0.24 ± 0.15	0.26 ± 0.24	0.43 ± 0.30
Elevational (mm)	0.41 ± 0.31	0.18 ± 0.10	0.16 ± 0.14	0.54 ± 0.42
Euclidean norm (mm)	0.56 ± 0.28	0.42 ± 0.15	0.38 ± 0.28	0.85 ± 0.45
Registration runtime (ms)	163	150	190	173

using 2-D array US transducers for a real-time implementation. These transducers would provide acquisition rates of 50 to 1000 volumes/s.[24,25] These 2-D array transducers could also be fairly small and placed closer to the region of interest.

A third issue deals with the laser delivery system. As shown in our experimental setup, a laser would have to be fired at the organ in free space. This occurrence is unlikely in practical situations. We are developing a fiber delivery tool that will allow us to safely guide the laser beam into the patient's body. This tool will also be able to project concurrent laser spots, greatly enhancing our registration acquisition rate.

The computation times shown throughout Secs. 2 and 3 still require some optimization for this method to become real-time at a reasonable refresh rate. The most obvious areas for improvement are data collection and PA image beamforming. The data collection phase can be improved dramatically with two changes. First of all, automated data collection as opposed to manual data collection would theoretically bring the data collection to the laser system pulse rate, which in our case is 10 Hz. The second issue would be the laser delivery system described above. Processing a single volume as opposed to a volume for each PA signal will greatly decrease computation time. PA beamforming computational cost is similar to conventional and current B-mode beamforming methods. Therefore, we do not anticipate technical or computational challenges to the implementation of real-time PA beamforming methods.

There are several factors that will affect this system's errors as we move from a bench-top setup to *in vivo* experiments. When our SC system is replaced with a stereo endoscopic camera, the errors may increase because our SC system has a larger disparity due to the shorter distance between the two cameras in a stereo endoscopic camera. The disparity of a SC system directly affects the error in triangulating points found in each image into a 3-D point. Further work will be done to quantify the effects of this change. Also, the errors were reported based on surface points. Since the region of interest is often subsurface, our reported TRE will be biased for subsurface target errors. We believe that the bias will be fairly small since the PA spots are being detected in the same modality as any subsurface regions. Another factor is the effect of imaging a different medium using US. US images are generally reconstructed using a single speed of sound even though an image can contain multiple mediums with multiple speeds of sounds. There is significant variance in the speed of sounds in the phantoms that we used as *ex vivo* tissue and gelatin have different speed of sounds. This heterogeneity affects the axial scaling of the US image, but this is a problem that any US application must deal with.

5 Conclusion

We have proposed an innovative 3-D US-to-video direct registration medical tracking technology based on PA markers and demonstrated its feasibility on multiple *ex vivo* tissue phantoms. In this paper, we showed the ability to generate a PA signal on multiple *ex vivo* tissue phantoms in various scenarios and PA spot LEs rivaling point source LEs found in SC systems. The TRE results have been shown to have higher accuracy than state of the art surgical navigation systems. Future work will include the development of a fiber delivery tool, spot finding algorithms to support concurrent spot projection, and subsequent *in vivo* animal experiments. Integration of this direct

registration method into laparoscopic or robotic surgery environments will also be a point of emphasis.

References

1. Y. Wang, S. Butner, and A. Darzi, "The developing market for medical robotics," *Proc. IEEE* **94**(9), 1763–1771 (2006).
2. R. Taylor et al., *Computer Integrated Surgery*, MIT Press, Cambridge, Massachusetts (1996).
3. P. J. Stolka et al., "A 3D-elastography-guided system for laparoscopic partial nephrectomies," *Proc. SPIE* **7625**, 76251I (2010).
4. C. L. Cheung et al., "Fused video and ultrasound images for minimally invasive partial nephrectomy: a phantom study," *Med. Image. Comput. Comput. Assist. Interv.* **13**(3), 408–415 (2010).
5. N. Navab, M. Mitschke, and O. Schutz, "Camera-augmented mobile C-arm (CAMC) application: 3D reconstruction using low cost mobile C-arm," *Med. Image. Comput. Comput. Assist. Interv.* **1679**, 688–697 (1999).
6. A. Wiles, D. Thompson, and D. Frantz, "Accuracy assessment and interpretation for optical tracking systems," *Proc. SPIE* **5367**, 421–432 (2004).
7. E. Boctor et al., "A novel closed form solution for ultrasound calibration," in *Int. Symp. Biomed. Imag.*, pp. 527–530, IEEE, Arlington, (2004).
8. T. Poon and R. Rohling, "Comparison of calibration methods for spatial tracking of a 3-D ultrasound probe," *Ultrasound Med. Biol.* **31**(8), 1095–1108 (2005).
9. J. Leven et al., "DaVinci canvas: a telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability," *Med. Image. Comput. Comput. Assist. Interv.* **8**(1), 811–818 (2005).
10. M. C. Yip et al., "3D ultrasound to stereoscopic camera registration through an air-tissue boundary," *Med. Image. Comput. Comput. Assist. Interv.* **13**(2), 626–634 (2010).
11. S. Vyas et al., "Interoperative ultrasound to stereocamera registration using interventional photoacoustic imaging," *Proc. SPIE* **8316**, 83160S (2012).
12. R. Kolkman, W. Steenbergen, and T. van Leeuwen, "*In vivo* photoacoustic imaging of blood vessels with a pulsed laser diode," *Laser. Med. Sci.* **21**(3), 134–139 (2006).
13. N. Kuo et al., "Photoacoustic imaging of prostate brachytherapy seeds in *ex vivo* prostate," *Proc. SPIE* **7964**, 796409 (2011).
14. M. Xu and L. Wang, "Photoacoustic imaging in biomedicine," *Rev. Sci. Instrum.* **77**, 041101 (2006).
15. C. Hoelen et al., "Three-dimensional photoacoustic imaging of blood vessels in tissue," *Opt. Lett.* **23**(8), 648–650 (1998).
16. B. Treeby and B. Cox, "*k*-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields," *J. Biomed. Opt.* **15**(2), 021314 (2010).
17. A. Cheng et al., "Direct 3D ultrasound to video registration using photoacoustic effect," *Med. Image. Comput. Comput. Assist. Interv.* **2**, 552–559 (2012).
18. IEC 60825-1:1993+A1:1997+A2:2001: Safety of Laser Products—Part 1: Equipment Classification and Requirements, International Electrotechnical Commission, Geneva, 2001, "IEC safety standard for lasers," http://lpno.tnw.utwente.nl/safety/iec60825-1%7Bed1.2%7Den.pdf (4 June 2013).
19. H. J. Kang et al., "Software framework of a real-time pre-beamformed RF data acquisition of an ultrasound research scanner," *Proc. SPIE* **8320**, 83201F (2012).
20. J. Bouguet, "Camera calibration toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/ (4 June 2013).
21. A. Myronenko and X. Song, "Point-set registration: coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(12), 2262–2275 (2010).

22. S. Prahl, "Tabulated molar extinction coefficient for hemoglobin in water," *Molar Extinction Coefficient for Hemoglobin in Water*, 4, March 1998, Oregon Technology, 17 September 2012, 1998, http://omlc.ogi.edu/spectra/hemoglobin/summary.html (4 June 2013).

23. R. L. P. van Veen et al., "Determination of VIS- NIR absorption coefficients of mammalian fat, with time- and spatially resolved diffuse reflectance and transmission spectroscopy," in *Biomedical Topical Meeting*, p. SF4, Optical Society of America, Miami Beach, Florida (2004).

24. B. C. Byram et al., "3-D phantom and *in vivo* cardiac speckle tracking using a matrix array and raw echo data," *IEEE Trans. Ultrason. Ferroelectrics Freq. Contr.* **57**(4), 839–854 (2010).

25. M. A. Bell et al., "*In vivo* liver tracking with a high volume rate 4D ultrasound scanner and a 2D matrix array probe," *Phys. Med. Biol.* **57**(5), 1359–1374 (2012).