Concurrent Photoacoustic Markers for Direct three-dimensional Ultrasound to Video Registration

Alexis Cheng\textsuperscript{a}, Xiaoyu Guo\textsuperscript{b}, Hyun-Jae Kang\textsuperscript{a}, Behnoosh Tavakoli\textsuperscript{a}, Jin U. Kang\textsuperscript{b}, Russell H. Taylor\textsuperscript{a}, and Emad M. Boctor\textsuperscript{abc}

\textsuperscript{a}Department of Computer Science, Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218, USA;
\textsuperscript{b}Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218, USA;
\textsuperscript{c}Department of Radiation Oncology, Johns Hopkins Medical Institute, 1800 Orleans St., Baltimore MD 21287, USA

ABSTRACT
Fusion of video and other imaging modalities is common in modern surgical procedures to provide surgeons with additional information that can provide precise surgical guidance. An example of such uses interventional guidance equipment and surgical navigation systems to register the tools and devices used in surgery with each other. In this work, we focus explicitly on registering three-dimensional ultrasound with a stereocamera system. These surgical navigation systems often use optical or electromagnetic trackers. However, both of these tracking systems have various drawbacks leading to target registration errors of approximately 3mm. Previous work has shown that photoacoustic markers can be used to register three-dimensional ultrasound with video resulting in target registration errors which are much lower than the current state of the art. This work extends this idea by generating multiple photoacoustic markers concurrently as opposed to the sequential method used in the previous work. This development greatly enhances the acquisition time by a factor equal to the number of concurrently generated photoacoustic markers. This work is demonstrated on a synthetic phantom and an ex vivo porcine kidney phantom. The resulting target registration errors for these experiments ranged from 840 to 1360 \( \mu \text{m} \) and standard deviations from 370 to 640 \( \mu \text{m} \).

Keywords: Clinical Ultrasound, Video Registration, Photoacoustic Imaging, Planning and Image Guidance of Interventions

1. INTRODUCTION
Modern surgical procedures including open, laparoscopic, and robotic surgeries are often aided by interventional guidance systems.\textsuperscript{1} These systems are necessary because the surgical environment is constantly evolving and tumors or other regions of interest may be invisible to the naked eye, or move in and out of the camera’s field of view. Interventional guidance systems can alleviate these concerns by providing a fusion of video and other imaging modalities, such as interoperative ultrasound (US). This additional information can support surgeons in finding and tracking tumors or other regions of interest. These guidance systems are enabled by the registration between surgical tools and devices, such as stereoscopic endoscopes and US transducers.

This work focuses on the registration of US imaging and video. Many surgeries require real-time US imaging including liver resections, partial nephrectomies, and prostatectomies. Real-time fusion of US and video, is crucial to the success of these operations. Registration is required to generate this real-time fusion. The registration between US images and video is an active area of research and significant challenges remain. Typically, electromagnetic (EM) or optical navigational trackers\textsuperscript{2,3} are used to provide the real-time pose, position and orientation, of tools such as US transducers. These navigational trackers require a separate base station to track
the two main types of surgical navigation and tracking systems are based on EM or optical trackers. Each of these trackers have their respective advantages and disadvantages. The main advantage for EM-based surgical navigation systems is that a clear line of sight to the EM sensor is not necessary. This reduces the constraints on how the surgical field is setup. However, several drawbacks must be considered and accounted for. First, EM tracking systems require wired EM sensors to be placed on the tool to be tracked. This is disadvantageous as it clutters the surgical environment and modifies the tools, possible decreasing the surgeon’s comfort while potentially increasing handling and sterilizing costs. Second, EM tracking systems require a large and intrusive EM base station to be placed in close proximity to the tracked EM sensors. Generally, there is very limited space around the operating table, thus the surgeon must consider if such a system would be the most effective use of this limited resource. Finally, EM-based systems suffer from magnetic field distortions when metallic objects are placed within its field. This serves as one of the main limitations as it degrades the system’s accuracy, thereby decreasing the value that surgeons can derive from the system.

Optical tacking systems do not suffer from magnetic field distortion issues associated with EM trackers and often do not require wired sensors. While, optical tracking systems can detect optical markers with sub-millimeter accuracy, line of sight is required to track the optical markers. This places a restrictive constraint on the number and placement of other tools in the surgical field, making such systems often impractical for laparoscopic procedures. These concerns can be somewhat addressed by placing the optical markers outside the body, but the tracking accuracy of long and flexible tools will degrade as their tips are now much farther away from the optical markers.

One drawback that affects both EM and optical-based navigation systems is that the process to acquire the transformation registering surgical tools with the navigation system is an indirect process. This means that a chain of transformations linking the coordinate system of the surgical tool and the coordinate system of the navigation system must be computed or acquired. An example of the chain of transformations necessary to enable interventional US guidance can be seen in Figure 1. The desired transformation is composed by $F_{Opt\_M}$, the pose acquired from the tracking system, and $F_{M\_US}$, the transformation between the sensor or marker and the US image plane obtained by US calibration. By composing a chain of transformations, their respective errors are also composed and magnified. Thus, it is much more beneficial to have a method which can acquire the desired transformation directly. An example of a direct process using our method is shown in Figure 2. There is a single transformation registering the tracking system with the US image volume.
Surgical tools with embedded EM sensors

US Probe

Stereo Scope with EM Sensor

Lots of wires

Implanted EM Sensors

Surgical Tools

Camera

Pulsed Laser

Projection System

Figure 2. A clinical scenario A) using a conventional EM-based system B) using PA markers to acquire this video to three-dimensional US registration.

Another drawback of these navigation systems is specific to our use of interventional US guidance. US calibration is an active topic of research and many authors have presented methods to achieve better accuracy and lower errors.\textsuperscript{7, 8} Their results have shown that the overall registration error is dominated by the calibration process as its error is much larger than the error of the tracking systems. Overall registration errors of approximately 1.7 to 3mm for artificial phantoms and 3 to 5mm for tissue have been shown.\textsuperscript{3, 4, 9, 10}

Vyas et al.\textsuperscript{11} and Cheng et al.\textsuperscript{12, 13} demonstrated a direct three-dimensional US to video registration method using photoacoustic (PA) markers. This novel method used PA markers generated on an air-tissue interface, visible to both a stereocamera (SC) system and US, as fiducials to directly acquire the registration between video and three-dimensional US. Previous work\textsuperscript{14, 15} showed that a pulsed laser source is capable of generating a PA signal in tissue. The resulting acoustic wave from the PA signal can be detected by a conventional US transducer.\textsuperscript{16, 17} The laser source is also visible to the SC system, so the PA markers are also visible. This enables PA markers to be used as fiducials as the same point can be represented in both the SC system’s coordinate system and the US image’s coordinate system.

This method addresses the drawbacks present in EM and optical-based navigation systems, an example of which can be seen in figure 2A. First, this method does not require wired sensors so no modifications are made to the surgeon’s tools. This allows the tools to maintain their present handling and sterilizing procedures. Since there are no attached wired sensors, US calibration is also unnecessary with this method. This is a major advantage as the registration error can be much lower by avoiding the US calibration process.\textsuperscript{12, 13} While this method requires line of sight between the PA markers and the SC system, this requirement is less stringent than the requirement for optical trackers. The PA markers are projected onto the surface of a region of interest, so it will naturally be within the SC system’s field of view.

This work extends the earlier work of Vyas et al.\textsuperscript{11} and Cheng et al.,\textsuperscript{12, 13} and serves as another step towards realizing a practical clinical system shown in Figure 2B. The main contribution of this work is the development of a fiber delivery device, capable of projecting and generating concurrent PA markers. There are many advantages to using such a device in place of the prior methods of sequentially generating PA markers. Prior work used a free-space laser to generate the PA markers. There are safety concerns with this approach as the laser light is allowed to freely move outside of the body cavity. It would also be difficult to create a complex system of mirrors that would guide the laser light into the body cavity. Thus, a fiber delivery device is a logical progression as it allows the user to easily manipulate the laser source. The output of the fiber can now be inside the body cavity, eliminating some of the safety concerns. The ability to concurrently project PA markers significantly decreases the acquisition time for this method. When the PA markers were being generated sequentially, an US volume...
must be acquired for each PA marker. With this concurrent PA marker fiber delivery device, only a single US volume is necessary.

This work details the design and development of the fiber delivery device, enabling concurrent PA markers, the extension to the software to support concurrent spots, and experimental results on a synthetic phantom with excellent light absorption characteristics and on an *ex vivo* porcine kidney embedded in a gelatin phantom.

2. DEVICE DESIGN

The design of the fiber delivery device must meet several requirements. First of all, the input side must be fiber-coupled with the laser source. Standard optical fiber connectors can effectively couple the laser energy from the laser source into the optical fiber. These optical fiber connectors can be seen in Figure 3. The fiber must be able to carry and deliver enough energy to generate a PA marker on tissue. We will later show in our experiments that a bundle consisting of 200µm fibers is sufficient. The fiber must also be a fiber bundle since we want to have concurrent PA markers. This is only true because we use a single laser source. It is possible to use separate fibers if there are multiple laser sources. To achieve concurrent PA markers, the fiber output must be able to project spots of light onto some surface 10-20cm away. This distance is chosen because we wish to eventually develop these tools for laparoscopic use. We separate the individual fibers in the fiber bundle and lightly focus the output using optical lenses, allowing each of these outputs to resemble spots on the target surface. Two different approaches were considered: independent lenses and single lens.

2.1 Independent Lenses

The approach of using independent lenses means that each of the individual fibers had an attached lens. This setup can be seen in Figure 3. The fiber is fixed at a certain distance away from a bi-convex lens. The fiber has a numerical aperture of 0.22. The bi-convex lens has an aperture of 6mm and a focal length of 10mm. The bi-convex lens lightly focuses the output and a single PA spot can be projected onto a surface 10-20cm away. A custom holder was created using a three-dimensional printer to hold the bi-convex lens stationary relative to the fiber output.

This design has several advantages and disadvantages. First of all, there is no set pattern to the concurrent PA markers. Each of the fiber outputs can move independent of each other. This can be seen as both an advantage and a disadvantage. As an advantage, no set pattern means that the pattern can be adjusted to best fit the region of interest. There may be regions of the surface that should be avoided. For example, if a PA marker is generated onto an edge between two types of tissue, this may possibly affect the quality of the PA
marker seen in the US volume. However, the disadvantage is that there must be some method to reassemble the individual fibers into some pattern within the body cavity. This can be seen as added complexity, as these fibers must be somehow actuated and manipulated within the body cavity. Another advantage is that each individual fiber can be very small. If each fiber is introduced into a laparoscopic environment independently, it is likely that no additional port will be necessary. It may also be possible to attach each of these fibers onto another device or tool, such as the SC.

2.2 Single Lens

The other approach of using a single lens means that all of the fibers share a single lightly focusing lens. This setup can be seen in Figure 4. The fiber bundle is mechanically split by a fiber splitter into a set pattern. A convex lens is then attached to the end of the tool. This convex lens has an aperture of 25mm and a focal length of 50mm. This device has also been designed to be adjustable. The fiber splitter can be adjusted to generate a pattern from a small pattern to a large pattern. The distance between the convex lens and the fiber output can also be adjusted to change the ideal distance for PA marker projection. A closeup of this adjustable device can be seen in Figure 5.

This design also has several advantages and disadvantages. The projected pattern is relatively fixed, only differing in scale and skew parameters. This can act as additional information and aid the segmentation algorithms in finding all of the PA markers. A disadvantage of this approach is that the entire device is now a single large tool. This means that an additional port is necessary to introduce this tool into a laparoscopic environment. We decided to use this design in our experiments as its easily adjustable nature allowed us more testing opportunities throughout the design process.

3. METHODS

In these experiments, we used a Q-switched neodymium-doped yttrium aluminum garnet (Nd:YAG) Brilliant (Quantel Laser, France) laser to generate the PA marker. We used a wavelength of 532nm and an energy density of approximately 1.6mJ/cm$^2$ on the synthetic phantom and approximately 3.8mJ/cm$^2$ on the ex vivo kidney phantom. These values are below the maximum permissible exposure (MPE), 19.5mJ/cm$^2$, as calculated from the IEC 60825-1 laser safety standard based on a 0.25s exposure time, a 4ns 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 three-dimensional US transducer consists of a linear US array, motor actuated to move angularly around an internal pivot point. It has a bandwidth of 5
to 14MHz and the linear array is approximately 38.4mm. The Sonix DAQ device, developed by the University of Hong Kong and Ultrasonix, and the MUSiiC toolkit\textsuperscript{20} are used to acquire prebeamformed radio-frequency (RF) data from the US machine. The k-wave toolbox\textsuperscript{18} in MATLAB (Mathworks Inc. Natick, Massachusetts) is used to beamform and reconstruct PA images based on the prebeamformed RF data. The SC setup consists of two CMLN-13S2C cameras (Point Grey Research, Richmond, Canada) to capture images at 18Hz. The camera calibration process using the Camera Calibration Toolbox for MATLAB\textsuperscript{21} generates a calibration file for the SC setup, allowing us to perform three-dimensional triangulation.

These experiments were performed on two phantoms. The synthetic phantom is made with plastisol and black dye. The \textit{ex vivo} kidney phantom is made with gelatin and a freshly resected porcine kidney. The surface of the kidney is exposed from the gelatin such that there is an air-kidney interface. Figure 6 shows the \textit{ex vivo} kidney phantom during the experiment. The PA markers within figure 6 has been artificially marked for presentation.

The experiments can be separated into three phases: data collection, data processing, and registration. The data collection phase consists of collecting multiple pairs of SC images and a three-dimensional RF US volume.
The data processing phase will then process the data to generate a three-dimensional SC point set and a three-dimensional US point set. These two point sets are registered together in the registration phase to finally output the transformation registering the SC frame to the US frame. The new workflows can be seen in figure 7.

The details of these phases are similar to those in previous work\(^1\),\(^2\) except for several key differences. First of all, concurrent PA markers allows us to collect all of the necessary data in a single three-dimensional RF US volume and multiple pairs of SC images from the same instance. Previously, the data collection pipeline was repeated for each PA marker, but that is no longer necessary. This significantly decreases the data acquisition time and makes the assumption that the three-dimensional RF US volume is at a single time instant much more plausible.

The other main change is modifying both the video and US segmentation methods to segment multiple markers as opposed to a single one. The same pipeline modified to accept more than a single PA marker worked fairly well. However, there were some cases where some of the PA markers could not be automatically segmented from the three-dimensional RF US volume. The energy density of each PA marker was non-uniform, making it difficult to automatically select a threshold that would be ideal for all of the PA markers. A more robust segmentation method not based solely on intensity is in development, but this method is robust to missing PA markers because there is built-in redundancy in the PA markers.

The registration phase is unchanged from previous work.\(^1\),\(^2\) We note here that the coherent point drift\(^7\) algorithm is used, so point correspondence is not necessary. This is a tremendous advantage for this method, as the PA markers seen in the SC frame and seen in the US frame will not have inherent correspondence since they are being projected concurrently. They will be corresponding point sets, but their correspondence will be unknown unless established with another method. It is still possible to use point set registration methods requiring correspondence if this point correspondence is first established.

### 4. RESULTS

The registration results of our experiments on the synthetic phantom and the \textit{ex vivo} kidney phantom are validated using the target registration error (TRE) metric defined in Equation 1. \(F_{\text{SCUS}}\) is the transformation between the SC frame and the US frame 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. \(N\) is the number of points in the experiment and \(N-1\) points are used to compute \(F_{\text{SCUS}}\). This computation is repeated...
with each of the $N$ points as test points. The TRE results for the synthetic and \textit{ex vivo} kidney phantom are shown in Table 1.

$$T\hat{R}E = F_{SCUS} - U_{S_{test}}$$  \hspace{1cm} (1)

Table 1. TRE results for the synthetic and \textit{ex vivo} kidney phantom.

<table>
<thead>
<tr>
<th></th>
<th>Synthetic Phantom</th>
<th>Ex \textit{vivo} Kidney Phantom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral (mm)</td>
<td>0.38±0.28</td>
<td>0.48±0.45</td>
</tr>
<tr>
<td>Axial (mm)</td>
<td>0.8±0.10</td>
<td>0.46±0.34</td>
</tr>
<tr>
<td>Elevational (mm)</td>
<td>0.59±0.50</td>
<td>1.02±0.73</td>
</tr>
<tr>
<td>Euclidean Norm (mm)</td>
<td>0.84±0.37</td>
<td>1.36±0.64</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The experimental results in Table 1 show that this three-dimensional US to video registration method using PA markers has higher accuracy than state of the art surgical navigation systems. While these are good results, there is some concern that they are worse than the results shown previously.\cite{12, 13} There may be several explanations for this occurrence. First of all, the sample size of the point sets is very small. This is a result of using a single laser source to generate multiple PA markers. There is a limit to how many PA markers can be generated. There are several possible solutions. It is possible to project concurrent PA markers multiple times. This would maintain a short data acquisition time, while increasing the amount of data. Another solution is to use another laser source that can support more fibers. Second, the points are projected much closer together. This will cause any errors or uncertainties in the point segmentation to be magnified because their magnitude will be fairly large relative to the distance between the points.

There are also some considerations in moving this system to \textit{in vivo} experiments. Thus far, we have assumed that everything remains static during data acquisition. Evidently, this is not a valid assumption during an \textit{in vivo} experiment. While we have decreased the acquisition time significantly, it is still on the order of 3-10s. This much time is required because the US transducer that we are currently using is an actuated transducer. Thus, we need to collect data for each actuated motor position. A two-dimensional array transducer would be able to provide an entire volume, thus reducing the acquisition time to a level where the assumption that the surgical environment is static is valid.

6. CONCLUSION

We demonstrated an extension to an innovative three-dimensional US-to-video direct registration medical tracking technology based on PA markers. We demonstrated the feasibility of this method using concurrent PA markers on a synthetic phantom and an \textit{ex vivo} kidney phantom. We showed that this method has higher accuracy than state of the art surgical navigation systems. Future work will explore improving the robustness of the segmentation algorithms, \textit{in vivo} animal experiments, and integration of this registration method into laparoscopic or robotic surgery environments.

ACKNOWLEDGMENTS

Financial support was provided by Johns Hopkins University internal funds, NIBIB-NIH grant EB015638, and NSF grant IIS-1162095.
REFERENCES