Augmented Environments and Computer Aided Interventions Workshop, *MICCAI*, Boston, Sept. 2014, *Lecture Notes in Computer Science*, Vol. 8678

Towards video guidance for ultrasound, using a prior high-resolution 3D surface map of the external anatomy

Jihang Wang[1], Vikas Shivaprabhu[1], John Galeotti[1, 2], Samantha Horvath[2], Vijay Gorantla[3], George Stetten[1, 2]

[1] Department of Bioengineering, University of Pittsburgh
[2] Robotics Institute, Carnegie Mellon University
[3] Department of Reconstructive Surgery, University of Pittsburgh Medical Center

Abstract. We are developing techniques for guiding ultrasound probes and other clinical tools with respect to the exterior of the patient, using one or more video camera(s) mounted directly on the probe or tool. This paper reports on a new method of matching the real-time video image of the patient's exterior against a prior high-resolution surface map acquired with a multiple-camera imaging device used in reconstructive surgery. This surface map is rendered from multiple viewpoints in real-time to find the viewpoint that best matches the probe-mounted camera image, thus establishing the camera's pose relative to the anatomy. For ultrasound, this will permit the compilation of 3D ultrasound data as the probe is moved, as well as the comparison of a real-time ultrasound scan with previous scans from the same anatomical location, all without using external tracking devices. In a broader sense, tools that know where they are by looking at the patient's exterior could have an important beneficial impact on clinical medicine.

Keywords: ultrasound, tracking, anatomical coordinates, computer vision, guidance, ProbeSight.

1 Introduction

Ultrasound (US) is an extremely useful clinical imaging modality for monitoring a wide variety of anatomical and physiological characteristics. It has numerous advantages including low-cost, real-time operation, portability, and lack of ionizing radiation. Whereas other image modalities such as computed tomography (CT) or magnetic resonance imaging (MRI) provide innate 3D anatomical coordinates, US scans lack such contextual correlates due to changing probe location. The operator holding the US probe may not feel this limitation during the scan, because the patient's external anatomy is clearly visible to provide navigational context. However, when reviewing the US images later, difficulties may arise in accurately interpreting the anatomical location, for example, to reposition the probe at precisely the same location and orientation with respect to the patient as a previous scan, or even simply to understand the underlying anatomy. Besides the ambiguity introduced by the freely moving probe, variation in joint pose as well as compression of tissue by the probe create serious challenges to interpreting US images not experienced with CT or MRI.

These challenges for US stem from its lack of a stable coordinate system. When assembling 3D US data from multiple 2D scans, the typical approach is to track the US probe relative to an external optical or magnetic tracking system, with the patient kept immobile during the scan. The external coordinate system of the tracker must then be related to the patient's anatomy to establish a context for the US scans. Our present research aims to replace such external tracking systems with a self-contained guidance system based on one or more video cameras mounted on the US probe itself. The camera's view of the external anatomy can provide anatomical coordinates for the US probe as it scans the patient. We call this self-contained guidance system *ProbeSight*, since it provides the US probe with a visual capability analogous to that of the human operator, to see for itself where it is relative to the patient.

Although the present paper mainly concerns our progress in using video to determine probe pose relative to external anatomy, we begin in Section 2 by describing how we will use that pose information to reconstruct and interrogate a 3D US data set. Our present reconstruction of 3D US employs conventional optical tracking, which we will eventually replace with the integrated video-based navigation system described in Section 3. Since the first clinical application intended for our system is monitoring patients after hand transplants, the anatomical target for our initial tests is the human forearm.

2 Reconstructing 3D ultrasound data using external tracking

One use for ProbeSight is to provide anatomical coordinates for reconstructing a 3D US dataset and retrieving arbitrary slices from it. A number of other researchers have developed systems to determine the US probe location using either optical or magnetic tracking systems, e.g. [1]. We have implemented a similar system, in which the location and orientation of the probe is determined by an external fixed optical tracking system (MicronTracker Sx60, Claron Technology) with a marker mounted on the US probe (Fig. 1A). We use this tracking system to reconstruct a 3D US dataset. For each B-scan, individual 2D images are stamped with the time of acquisition and the location of images obtained from the tracking system. A 3D volume is then reconstructed by placing the 2D images within a 3D space based on the tracking information (Fig. 1B). When the 2D images are consolidated into a 3D space, a particular voxel in the 3D volume may either be intersected by pixels from more than one 2D image or may not be intersected by any scan. As suggested in [2], the former problem, known as *bin-filling*, can be solved by combining data in the overlapping pixels (compounding). The latter problem, known as *hole-filling*, can be solved by inferring values for the missing data, using information of the voxel's neighbors (interpolation). We employ bin-filling and hole-filling techniques described in [3] and [4]. In order to correctly localize the data captured, temporal and spatial calibrations are required. We employ the method described in [5] for temporal calibration, to find the latency between the image acquisition and the tracking system. Spatial calibration finds the transformation between pixels in the 2D image and the location of the tracked marker on the probe in 3D space. We use an established N-wire phantom developed specifically for this purpose [6]. Our algorithms for calibration, image acquisition, and volume reconstruction are based on the Public Software Library for Ultrasound (PLUS) toolkit [7]. Reconstructed image slices that correspond to the current location of the US probe may then be retrieved from previously stored US data. To illustrate this, we use an US phantom containing tubing to simulate vasculature (Blue Phantom, Inc.). The phantom is tracked with reference markers. Fig 1D shows an image slice retrieved from a reconstructed 3D US volume of the phantom corresponding to the live US image seen in Fig 1C. The quality of the reconstructed slice suffers from the problems outlined above.

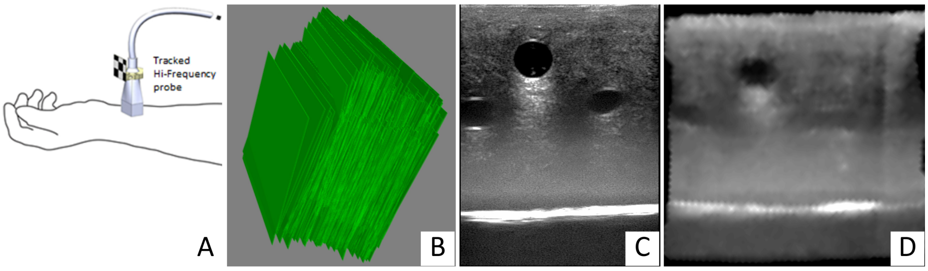


Fig 1. **(A)** US probe tracked by markers mounted on it. **(B)** 3D model of individual 2D US images displayed in 3D space based on the recorded location and orientation of the US probe. **(C)** Live US image. **(D)** US image slice extracted from a previously reconstructed US volume corresponding to the live US image.

We have several reasons to want to replace external tracking systems in our application. Although they work well in a controlled environment, optical tracking demands continuous line of sight and magnetic tracking is unpredictable near ferromagnetic materials. Neither technology is generally as accurate as the vendors claim. Furthermore, an external tracking system restricts the portability of the US scanner, one of its great advantages in the hospital. Finally, the location of the patient and the particular anatomical target being scanned must be independently determined. ProbeSight addresses all of these problems.

3 Probe-mounted video cameras to replace the external tracker

Attaching a video camera directly to the US probe theoretically permits determining the probe's pose relative to the patient's anatomy without any external tracking equipment. In [8] this approach was used to permit graphical overlays in the video image to show possible entry points for needle biopsy in the plane of the US scan. In [9] stereo cameras were mounted on the US probe to determine needle location relative to the probe. The US probe location relative to the patient's anatomy or US phantom has been determined by putting passive optical markers on the skin or phantom surface [10][11][12]. Such artificial surface markers can be problematic during clinical procedures, especially if they are to remain from one scan to the next. They may also influence the passage of US into patient and easily be smeared or distorted by the US gel. In our prior work we printed a checkerboard pattern on tracing paper and laid it upon a flat US phantom saturated with gel. The saturated tracing paper does not significantly interfere with the passage of US into the phantom, while remaining visible to stereo cameras mounted on the ultrasound probe, which determine the 3D location of the surface using stereo disparity [13].

We now propose to eliminate the optical trackers entirely and track natural skin features directly. The difficulties in applying computer vision algorithms to the unadulterated skin are significant. Hairless skin may contain only sparse features, hindering standard computer vision algorithms, such as stereo matching for determining depth, especially those algorithms operating without prior knowledge. We address this by providing detailed prior information in the form of a surface map, which we can match against images from the camera mounted on the ultrasound probe, close to the skin, where it can capture details such as pores and creases.

3.1 Using a high-resolution multi-camera surface map as prior information

We can greatly facilitate the determination of the probe-mounted camera's pose relative to the anatomy by supplying, beforehand, a detailed map of the entire anatomical terrain. Reconstructive surgeons already have devices with this ability. For example, the VECTRA M3 Imaging System (Canfield Imaging Systems) uses an array of three pairs of high-resolution cameras to acquire pre-operative images of anatomical structures for surgical planning. Over the course of several minutes, the system computes a detailed 3D surface map of the anatomy, including color texture with sufficient resolution to see pores and creases in the skin. Armed with such a prior scan, our probe-mounted camera can simply search for the matching portion of anatomy, much as a traveler might navigate with Google's Street View to stand in front of the correct house. We can render projections of the pre-acquired 3D surface map from any viewpoint to find a particular camera's actual viewpoint, and thus know the camera's pose relative to the external anatomy. Examples of projections of the surface map acquired from a subject's arm, rendered from three different viewpoints, are seen in Figure 2. Rather than match individual features, or small patches, as typical in stereo disparity routines, we will compare two entire images, a more robust proposition for computing camera pose.

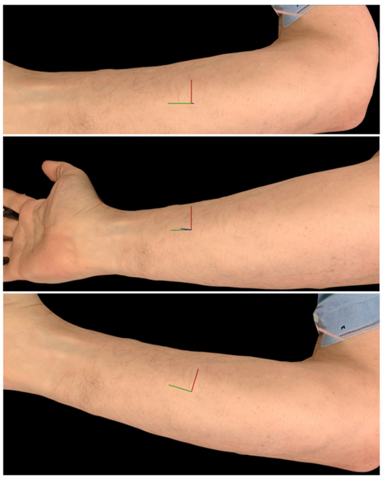


Fig. 2. Rendering a surface map from different viewpoints.

3.2 Rendering the surface map as seen by a real camera

Rendering a 3D surface map to yield the particular 2D projection that would be seen by the probe-mounted camera requires more effort than typical of graphical rendering for entertainment or visualization purposes. The actual optics of the particular camera must be accurately modeled, including focal length, distortion, and location of entrance pupil. In addition, the simulated lighting applied during the rendering process should match the lighting during the acquisition of the surface map. We discuss each of these issues next.

The 3D surface map data from the VECTRA imaging system consists of a tessellated point cloud, with every vertex assigned a color from the high-resolution camera array. We render this with OpenGL using a diffuse lighting model similar to the VECTRA scan's uniform lighting condition in the room where the VECTRA scanned the patient. Beyond the simple pinhole camera model used by OpenGL, however, we must model the optical parameters of our particular video camera.

Distortions are inherent in any real lens design, and we model them as separate polynomial expansions in the radial and tangential directions. Radial distortion arises because the lens behaves differently at the center of the image than at the periphery, resulting in "barrel" or "pincushion" distortion. We characterize this by a Taylor series expansion around *r*, the distance from the image center. For typical optical lenses, we generally require only the first 2 terms, which are conventionally termed k_1 and k_2. For highly distorted optics such as fish-eye lenses it may be necessary to use a third radial distortion term k_3 [14]. Location (x, y) on the image sensor will thus be corrected according to the following equations:

$$x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \qquad y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (1)$$

Tangential geometric distortion arises from imprecision during the manufacture of the camera resulting in the lens not being exactly parallel to the imaging plane. This can be minimally characterized by two additional parameters, p_1 and p_2, as follows [15]:

$$x_{corrected} = x + [2p_1 y + p_2(r^2 + 2x^2)] \qquad y_{corrected} = y + [2p_2 x + p_1(r^2 + 2y^2)] \quad (2)$$

All of these parameters can be estimated for a given camera, by applying existing routines in the Open-source Computer Vision (OpenCV) library to images taken by the camera of a standardized printed checkerboard pattern.

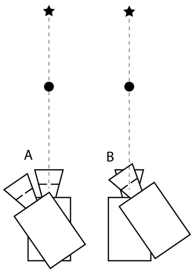


Fig. 3. Rotation to find entrance pupil.

In addition to calibrating the camera, we also perform a physical alignment along the camera's depth axis in order to align the physical axis of rotation with the camera's entrance pupil. The entrance pupil is the point about which the camera can be rotated without changing the relative pixel alignment between the objects at different distances. It is essential to know the entrance pupil's physical location to determine the camera's location, and thus the US probe's location. Our strategy to find the entrance pupil is to point the camera toward two objects that are positioned to perfectly overlap when the camera is facing directly toward them. The camera is then rotated. If the entrance pupil is not the same as the rotation point, the two objects will no longer overlap when the camera is rotated (Fig. 3A). We then move the camera on a slider until we find the location about which, when the camera is rotated, the two objects always overlap (Fig. 3B).

With the above parameters determined and the focal length provided by the camera-lens manufacturer, it is possible to render the 3D surface map simulating the image that would be seen by the camera from any possible point of view.

3.3 2D matching

Once the surface map has been rendered from a given viewpoint, it must be compared with the real-time image from the probe-mounted camera. A number of appropriate metrics are available, including normalized correlation. This method, however, does not perform well when pixels in the foreground (surface anatomy) are combined with significant regions of background that do not match. Segmenting foreground objects from the background by thresholding depth in the surface map can improve the matching result. However, we have found that mutual information [16][17], based on the joint distribution between the two images, is more reliable for our purposes, since it can accommodate differences in background without prior segmentation. In particular, we use *normalized mutual information* as described in [18].

3.4 Finding the best match

Given a metric for matching the rendering of the surface map to the real-time camera image, we can theoretically find the best match among all the possible camera viewpoints. However, the search space is very large, encompassing 6 degrees of freedom (DOFs): 3 translations and 3 rotations, and performing this search in real time presents challenges beyond the present paper. For now, we only demonstrate the accuracy of our projection method and the specificity of our matching process.

4 Validation

To validate our projection and matching methods, we used a textured phantom, in the form of a model dinosaur, roughly the same size as a human arm, viewed at a distance of approximately 20 cm. Note that in the eventual clinical system we expect the camera mounted on the US probe to be closer to the patient's skin, where finer details will be visible. We established ground truth for the location of the camera with respect to the dinosaur phantom using the same Micron optical tracking system described in Section 2. Markers for the tracking system were attached to a video camera (Prosilica GT1290C, Allied Vision Technologies) and the transformation between these markers and the cameras viewpoint determined, including entrance pupil. No US probe was included at this point. Five fiducials (small white dots) were painted on the surface of the phantom and a preassembled marker probe (Micron 950-MT-tool-B20) used to locate them in 3D space with the Micron tracker. These fiducials were also identified in the pre-acquired surface map using software for manually interrogating 3D data (MeshLab). Given this ground truth, we were able to predict the correct projection of the surface map to render, to match the actual image from the camera. An example of such a match is shown in Figure 4.

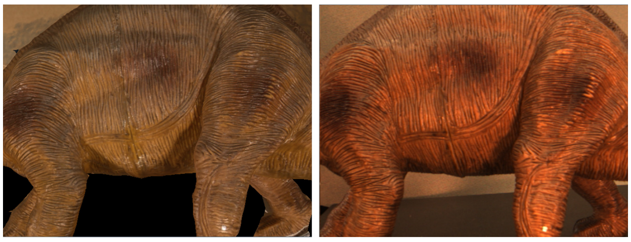
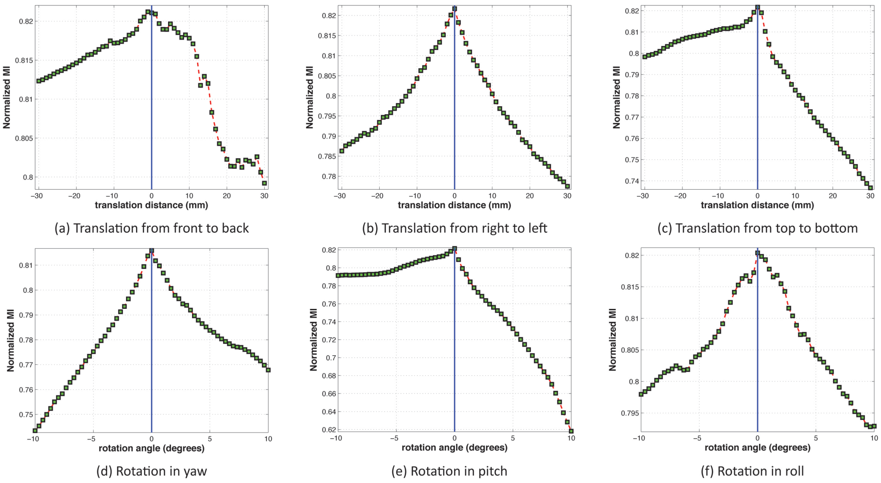


Fig. 4. Dinosaur model: Pre-acquired surface map (left), rendered from the correct viewpoint, matching the actual camera image (right). White dot fiducial on back leg.

We tested the accuracy of our ground truth and the suitability of our metric by deviating along each of the six DOFs from the ground truth viewpoint, projecting the surface map from each new viewpoint and applying the metric between it and the actual camera image. With an image size of 640 × 480 pixels, a displacement of 1 pixel corresponds to approximately 1 mm at a range of 20 cm to the phantom. Results are shown in Figure 5, with ground truth marked by the vertical line in the middle of the range for each DOF. Along each DOF, the normalized mutual information metric shows a clear maximum at the ground truth viewpoint established by the tracking apparatus, and diminishes nearly monotonically as one moves away from the optimal pose.



(a) Translation from front to back

(b) Translation from right to left

(c) Translation from top to bottom

(d) Rotation in yaw

(e) Rotation in pitch

(f) Rotation in roll

Fig. 5. Normalized mutual information for deviations from ground truth in each DOF.

5 Discussion

We have established that we can project a previously acquired surface map of a phantom to match the view through an actual camera, and that mutual information is an effective metric to determine which viewpoint best matches the camera image.

Our next step is to develop efficient optimization routines to search the 6 DOFs for the maximum metric value, so that (1) an initial match can be obtained without the use of fiducial markers, and (2) an optimal match can be maintained in real time with further motion of the camera. The challenge of efficiently searching the 6-dimensional space needs to be addressed, and motion simultaneously in multiple DOFs must be accommodated.

Once a fully self-contained camera navigation system is functional and validated on the phantom, we will move to human subjects, beginning with the arm, to determine the robustness of the system on skin. We expect the finer detail of an arm compared to the dinosaur phantom to be visible given the closer range of the camera mounted on an actual ultrasound probe. We will still use the optical tracking apparatus for validation purposes, although it will eventually not be needed in the clinical system. With the camera attached to an actual US probe, we will adapt the US processing described in Section 2 so that ProbeSight provides the pose relative to the anatomy.

A major source of error in the matching process is the difference in lighting between the projected surface map and actual camera image. These differences may be reduced by controlling the lighting conditions during the pre-operative scan as well as during the actual scan, and by simulating correct lighting conditions when rendering the surface map using OpenGL lighting models. Attaching a lighting source to the probe itself is a possibility, especially given that ultrasound scans are often performed in a darkened or shadowy environment.

Tissue deformation is a major concern for any clinically practical ProbeSight system. We hope to be able to determine deformation by how far the location of the probe tip is computed to be beneath surface of the pre-acquired model. We also plan to include deformable registration, to accommodate deformation in the 2D matching process. Finally, we expect to incorporate analysis of the ultrasound data itself to detect in-slice deformation, such as demonstrated in [19].

The major contribution described in this paper is the use of a previously acquired high-resolution 3D surface map, against which a real-time camera image can be matched to provide anatomical coordinates for an ultrasound probe or surgical tool to which the camera is mounted. Such a technology could enable safe, economical, non-invasive, reliable and reproducible 3D visualization of intricate anatomy, providing spatial orientation and precise localization of structures such as vessels, nerves, tendons, muscle and bone without the limitations and risks of CT angiography, intravascular ultrasound or MRI. The applications of such a technology could span screening, diagnostic, therapeutic, interventional and management strategies, in a wide array of medical and surgical indications.

Acknowledgments. This work was funded by NIH R01 grant 1R01EY021641, National Library of Medicine contract HHSN27620100058OP, and Peer Reviewed Medical Research Program (PR130773, HRPO Log No. A-18237) award from the U.S. Department of Defense.

References

1. Flaccavento, G., Lawrence, P., and Rohling, R., "Patient and probe tracking during freehand ultrasound," *MICCAI*, (2004).
2. Rohling, R., Gee, A., and Berman, L., "A comparison of freehand three-dimensional ultrasound reconstruction techniques," *Medical Image Analysis*, vol. 3, no. 4, (1999).
3. Dewi, D., Wilkinson, M., Mengko, T., Purnama, I., van Ooijen, P., Veldhuizen, A., et al., "3D Ultrasound Reconstruction of Spinal Images using an Improved Olympic Hole-Filling Method," *ICICI-BME* (2009).
4. Gobbi, D. and Peters, T., "Interactive intra-operative 3D ultrasound reconstruction and visualization." *MICCAI* (2002).
5. Rousseau, F., Hellier, P., and Barillot, C., "A novel temporal calibration method for 3-D ultrasound," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1108-1112,(2006).
6. Chen, T., Thurston, A., Ellis, R., and Abolmaesumi, P., "A real-time freehand ultrasound calibration system with automatic accuracy feedback and control," *Ultrasound in Medicine & Biology*, vol. 35, pp. 79-93 (2009).
7. Lasso, A., Heffter, T., Pinter, C., Ungi, T., and Fichtinger, G., "Implementation of the PLUS open-source toolkit for translational research of ultrasound-guided intervention systems," The MIDAS Journal - Systems and Architectures for Computer Assisted Interventions Workshop, pp. 1-12, *MICCAI* (2012).
8. Khamene, A., and Sauer, F., "Video-assistance for ultrasound guided needle biopsy". US Patent 6,612,991 (2002).
9. Chan, C., Lam, F., and Rohling, R., "A needle tracking device for ultrasound guided percutaneous procedures," *Ultrasound in medicine & biology,* 31(11), (2005).
10. Rafii-Tari, H., Abolmaesumi, P., and Rohling, R., "Panorama ultrasound for guiding epidural anesthesia: a feasibility study," in *Information Processing in Computer-Assisted Interventions*, Berlin (2011).
11. Sun, S., and Anthony, B., "Freehand 3D ultrasound volume imaging using a miniature-mobile 6-DOF camera tracking system," *9th IEEE International Symposium*, (2012).
12. Sun, S., Gilbertson, M., and Anthony, B., "6-DOF probe tracking via skin mapping for freehand 3D ultrasound," *10th IEEE International Symposium,* (2013).
13. Wang, J., Horvath, S., Stetten, G., Siegel, M., Galeotti, J., “Real-Time Registration of Video with Ultrasound using Stereo Disparity,” SPIE Medical Imaging, San Diego, CA. (2012).
14. Fryer J., and Brown, D., "Lens distortion for close-range photogrammetry," Photogrammetric engineering and remote sensing, vol. 52(1), pp. 51-58, (1986).
15. Brown, D., "Decentering Distortion of Lenses," Photometric Engineering, vol. 32(3), p. 444–462, (1966).
16. Viola, P., & Wells III, W. M., “Alignment by maximization of mutual information. International journal of computer vision,” 24(2), 137-154, (1997).
17. Li, W. “Mutual information functions versus correlation functions,” Journal of statistical physics, 60(5-6), 823-837, (1990).
18. Studholme, C., Hill, D., Hawkes, D., “An overlap invariant entropy measure of 3D medical image anlignment,” Pattern Recognition, (32) 71-86 (1999).

19. Krupa, A., Fichtinger, G., Hager, G., "Real-time tissue tracking with B-mode ultrasound using speckle and visual servoing," Proceedings of the 10th international conference on Medical Image Computing and Computer-Assisted Intervention, pp. 1-8 (2007).