FREEHAND 3D ULTRASOUND VOLUME IMAGING USING A MINIATURE-MOBILE 6-DOF CAMERA TRACKING SYSTEM

Shih-Yu Sun and Brian W. Anthony

Massachusetts Institute of Technology
Laboratory for Manufacturing and Productivity
Medical Electronics Device Realization Center
Cambridge, MA, USA

ABSTRACT

Spatial registration of 2D ultrasound (US) scans and 3D volume reconstruction are very useful tools for clinical diagnosis. However, the existing methods for freehand US scan registration tend to be hindered by physical constraints of the equipment and patient motion artifacts. This paper describes a low-cost and unobtrusive method for spatially registering 2D US scans with six degrees of freedom (6 DoF). This method uses a lightweight camera mounted on the US probe to track skin features of the patient body for probe motion recovery, which is performed efficiently by approximating the skin surface as a planar structure. As probe position and motion are determined with respect to the body, this method is robust to patient motion and could potentially facilitate longitudinal studies. In this paper, the performance of 3D volume reconstructions in linear, tilt, and rotational scanning is quantitatively examined through *in-vitro* experiments.

Index Terms— freehand 3D ultrasound, registration, planar homography, camera tracking

1. INTRODUCTION

Three-dimensional ultrasound (3D US) has been shown to provide useful information in many clinical applications [1]. Currently 3D US involves either fixed or freehand scanning. Fixed scanning is performed locally by using a mechanically actuated probe or a 2D probe array, and hence the acquired volume is generally limited by the form factor of the probe. On the other hand, freehand scanning has been growing in popularity, since it allows volume acquisition of large areas, and the probe hardware is relatively inexpensive.

Spatial registration of acquired 2D US scans is one of the essential components of freehand 3D US, in which the scans are aligned accurately in a common coordinate system. Generally, this process requires determining 6-DoF spatial rigid transformations of the probe. The most popular methods of probe tracking involve the use of an external electromagnetic [2] or optical tracker [3]. However, these methods register the US scans with respect to a common world coordinate system that is independent of the body, and thus the reconstructed 3D volume could be distorted as a result of patient motion. Also, the equipment required to implement these methods is often expensive and bulky, making the use of such methods both economically and physically constraining.

Methods have been developed to make probe tracking systems less obtrusive, sometimes involving the aid of inertial measurement units (e.g. gyroscopes and accelerometers) for providing measurements in additional DoF or for refining estimates. One popular method uses the level of ultrasound speckle dissimilarity, or decorrelation, between neighboring scans to estimate the relative probe motion [4]. Attempts have also been made to use optical mice for tracking skin features, close to the skin surface, and estimating 2D probe movement [5, 6].

In this paper, a low-cost and unobtrusive method for 6-DoF probe tracking in freehand 3D US is described. This method estimates probe position and orientation with respect to the body by tracking skin features, at a small distance from the skin surface, with a small camera rigidly mounted on the probe. The method is robust to patient motion and could potentially facilitate longitudinal studies. The main advantage of this method is that it provides 6-DoF probe motion and position estimates by using a camera alone, as opposed to 2-DoF by optical mice. This method does not require additional sensors for estimates in the complete 6-DoF, although other sensors could be incorporated in a multiple-estimator framework.

2. METHOD

In this section, the proposed method is described. A small camera is rigidly mounted on the US probe. At each acquisition of a 2D US scan, an image of skin features is recorded. The skin surface under the whole 3D US scan is approximated as a planar structure, and a series of planar homographies are computed from the camera images. The probe pose with

This work is partially sponsored by Singapore-MIT Alliance (SMA) and GE Healthcare.

Initialization

1. Estimate K through camera calibration
2. Mount the camera onto the probe, and perform ultrasound calibration
3. Stick a square with known dimensions on skin surface

3D ultrasound scan

1. At each acquisition of an ultrasound scan, capture the image of skin features.
2. In a camera image where the square sticker is fully visible, denoted by p, mark the four corners and compute the world-to-image homography matrix H_w^p.
3. Compute H_{i-1}^i for each image i and then H_w^i following equation (3)
4. Compute the camera pose R and $\mathbf{t}$ for each camera image following equation (2)
5. Convert the camera poses to those of 2D ultrasound scans, from which the 3D volume is reconstructed.

Fig. 1. Method summary

which each scan is acquired can then be determined. The procedure is detailed in the following subsections and summarized in Fig. 1.

2.1. World-to-Image Planar Homographies

For a projective camera [7], the relationship between the homogeneous world coordinates $[XYZ1]^T$ and the corresponding image coordinates $[x\,y\,1]^T$ can be expressed as follows up to a scaling factor:

$$[x\,y\,1]^T = \mathrm{K}\,[\mathrm{R}|\mathbf{t}]\,[XYZ1]^T = \mathrm{K}\,[\mathbf{r_1}\ \mathbf{r_2}\ \mathbf{r_3}|\mathbf{t}]\,[XYZ1]^T \tag{1}$$

where superscript T denotes matrix transpose and K is the 3-by-3 projection matrix that incorporates the intrinsic parameters of the camera. The rotation matrix R and the translation vector $\mathbf{t}$ describe the geometric relationship between the world and camera coordinate systems; $\mathbf{r_1}$, $\mathbf{r_2}$ and $\mathbf{r_3}$ are the column vectors of R.

For points lying on a planar structure in the scene, the expression in (1) can be simplified by defining the plane as $Z = 0$:

$$[x\,y\,1]^T = \mathrm{K}\,[\mathbf{r_1}\ \mathbf{r_2}\ \mathbf{r_3}|\mathbf{t}]\,[XY0\,1]^T = \mathrm{K}\,[\mathbf{r_1}\ \mathbf{r_2}|\mathbf{t}]\,[XY1]^T \tag{2}$$

Therefore, the world coordinates of points on a plane, $\mathbf{X} \equiv [XY1]^T$, and the corresponding image coordinates, $\mathbf{x} \equiv [x\,y\,1]^T$, can be related by a 3-by-3 matrix $\mathrm{H} \equiv \mathrm{K}\,[\mathbf{r_1}\ \mathbf{r_2}|\mathbf{t}]$, usually called the planar homography matrix. Actually, the image coordinates in different perspectives of a planar structure are also related by a 3-by-3 homography matrix. With at least four correspondences, a least-square solution of the homography matrix could be found by the Direct Linear Transformation (DLT) algorithm, which could then be used as an initial estimate in an algorithm for iterative refinement [7].

2.2. Recovering Camera Rotation and Translation

From the discussion in Section 2.1, it can be seen that as long as K is known and the world-to-image homography matrix is determined, the column vectors $\mathbf{r_1}$, $\mathbf{r_2}$, and $\mathbf{t}$ can be found. The remaining vector $\mathbf{r_3}$ is then the cross product of $\mathbf{r_1}$ and $\mathbf{r_2}$, since R is a rotation matrix. In other words, for each image i, $i = 1, 2, 3, \ldots$, the camera position and orientation in the world coordinate system can be recovered, as long as K and H_w^i, the homography matrix from the world to image i coordinates, are found.

The determination of correspondences for computing H_w^i usually requires defining and calibrating against a geometric pattern with known physical dimensions in the image i, such as squares. Nevertheless, once one of these homographies is determined, H_w^1 for instance, H_w^i for $i = 2, 3, \ldots$ can be efficiently computed in an iterative manner [8] by finding H_1^2, H_2^3, $\mathrm{H}_3^4, \ldots, \mathrm{H}_{i-1}^i$:

$$\mathrm{H}_w^i = \mathrm{H}_{i-1}^i \ldots \mathrm{H}_2^3\,\mathrm{H}_1^2\,\mathrm{H}_w^1, \tag{3}$$

where H_{i-1}^i denotes the planar homography matrix from image $i-1$ to the next image, for which the determination of correspondences can be performed automatically. Note that although in this example, image 1 is used to build direct connection to the world coordinates, an arbitrary selection out of the image series could work as well, as long as the known geometric pattern is fully visible in that image.

As planar homographies are used for correlating feature points in the camera images, this method is particularly suitable for scanning body parts that are locally planar, such as the abdomen and breast. Under circumstances where this planarity assumption is severely violated, it is possible to perform general motion tracking by using epipolar geometry in this framework, but this tends to be more computationally intensive [9].

2.3. Implementation

In this implementation, a small camera with an adjustable focal length was mounted on the ultrasound probe and focused at the skin surface from about 2 cm above. A probe housing was designed to rigidly hold the camera and lighting source, as shown in Fig. 2. The camera was mounted about 5 cm from the probe such that local surface deformation due to probe compression is mostly avoided in the field of view. Image resolution is 800×600, which approximately corresponds to a 16.5 mm $\times$ 12.5 mm field of view. A typical example of skin features captured by the camera is shown in Fig. 3(a). Through camera calibration [10], the projection matrix K was estimated and radial lens distortion was corrected.

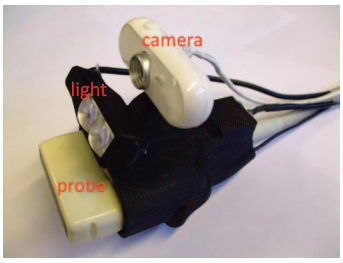
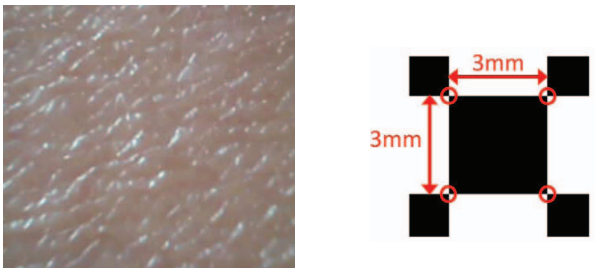


Fig. 2. The ultrasound probe with a housing for the camera and lighting source.



(a) skin features (b) square sticker

Fig. 3. Typical skin features under 5× magnification (a) and the square sticker with known dimensions (b): the four corners (marked by red circles) are located with subpixel accuracy in camera images.

The world coordinate system was established by affixing a 3 mm × 3 mm square, as shown in Fig. 3(b), to the skin surface. The four corners are located in camera images and used for calibration. The four smaller squares around the main square were designed to emphasize the corners and improve the accuracy of subpixel positioning.

The correspondences between consecutive images were extracted through scale-invariant feature transform (SIFT), which is relatively robust to affine distortion of features and illumination variation [11]. Following the Gold Standard Algorithm described in [7], the homography matrix relating these correspondences was estimated by minimizing reprojection errors, in which the rotation matrix R was parameterized by the Euler axis and angle representation. The transformation from camera coordinates to ultrasound image coordinates was found using a calibration process which uses an agar phantom with a surface optical target and a buried ultrasound target placed in known position and orientation with respect to each other [12].

3. RESULTS

3.1. Experiment Setup

We performed experiments on an agar phantom with "skin" surface texture, similar to the one used in [6]. Graphite powder was added to create scattering characteristics. A cylinder made from pure agar, whose volume is 1.26 mL, was embedded in the phantom for 3D US scanning. The dimensions of the phantom are shown in Fig. 4. The pattern shown in Fig. 3(b) was placed on the phantom surface for the establishment of a world coordinate system. 2D ultrasound scans were acquired by using the Terason t3000 system. At each ultrasound scan acquisition, a camera image of the phantom surface was recorded.

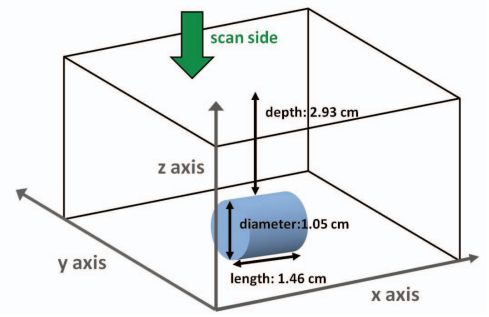


Fig. 4. Dimensions of the phantom

We tested three freehand scanning types on this phantom: linear, tilt, and rotational scanning. In linear scanning, the ultrasound scan plane remains approximately perpendicular to the phantom surface and the probe moves along the x axis. In tilt and rotational scanning, only pitch and yaw of the probe is varied, respectively, as illustrated in Fig. 5(a).

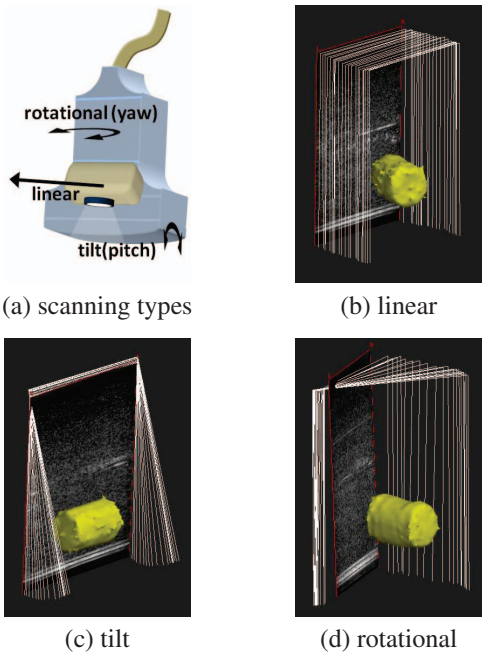
3.2. Experimental Data

After the ultrasound scans were acquired and their poses were estimated from the camera images, 3D volume reconstruction was performed using Stradwin [13]. Outlines of the cylinder were segmented in each scan by thresholding. The volume was then reconstructed and surface rendered. Typical scanning trajectories and reconstructed cylinders are shown in Fig. 5 (b)-(d).

An average of twenty 2D scans were acquired for each 3D volume reconstruction. Each type of scanning (linear, tilt, rotational) was performed five times. The volumes of the reconstructed cylinders were estimated in Stradwin. Fig. 6 shows the mean volume errors for each scanning type. The error bar represents one standard deviation. It can be seen that linear scanning generally gave volume estimates with less than 3% error. The errors in tilt and rotational scanning were about 3.5 times larger than in linear scanning.

4. DISCUSSION

Linear scanning gave the smallest volume error. This observation is not surprising, since recovery of rotation from camera images is a process that is relatively sensitive to errors in fea-



(a) scanning types (b) linear

(c) tilt (d) rotational

Fig. 5. (a) Illustration of the three scanning types: linear, tilt, and rotational scanning. (b)-(d) Examples of the trajectories and reconstructed cylinders are shown.

ture extraction. Pitch or yaw variation can be misinterpreted as translation given a limited camera field of view, especially when there is a lack of local features or camera defocusing in tilt scanning. For example, this phenomenon can be seen from the z-axis misalignment in the 2D scans in Fig. 5(c) and the cylinder shape distortion in Fig. 5(d), and was found to result in volume underestimation as shown in Fig. 6.

A major source of error in this method is drift - relying solely on skin features and transformation back to the camera calibration image once the calibration landmark is no longer in the camera's field of view. Drift can be reduced by registering the square landmark whenever it is fully visible, or by building a global map of the skin features. It is also possible to place multiple landmarks with known spatial configuration in the scanned region in order to register the camera globally and correct drift.

For this feature-based method, extraction of skin features is essential. When water is used as the ultrasound coupling medium, skin features could generally be detected and sometimes even emphasized by specularity. However, it was found that the use of ultrasound gel as a coupling medium makes the features less recognizable, or causes feature distortion that significantly varies with camera perspectives. Under circumstances where the use of ultrasound gel is necessary, one could move the probe in a way to avoid capturing skin features through the gel. The development of methods to accurately recover camera motion in the presence of gel remains part of future work.

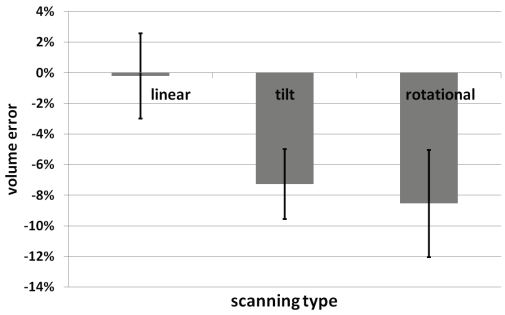


Fig. 6. Comparison of mean volume errors.

5. ACKNOWLEDGEMENT

The authors thank Matthew Gilbertson and Javier Ramos for their help with hardware design, and Kai E. Thomenius for his helpful discussion. The authors also thank Terason for donating the t3000 ultrasound imaging system and providing technical support.

6. REFERENCES

[1] R. W. Prager, U. Z. Ijaz, A. H. Gee, and G. M. Treece, "Three-dimensional ultrasound imaging," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 2, pp. 193–223, 2010.

[2] E. M Boctor, G. Fichtinger, R. H Taylor, and M. A Choti, "Tracked 3D ultrasound in radio-frequency liver ablation," in *Proc. SPIE Vol*, vol. 5035, pp. 174–182.

[3] G. M Treece, A. H Gee, R. W Prager, C. J.C Cash, and L. H Berman, "High-definition freehand 3-D ultrasound," *Ultrasound in Medicine & Biology*, vol. 29, no. 4, pp. 529–546, 2003.

[4] T. A. Tuthill, J. F. Krücker, J. B. Fowlkes, and P. L. Carson, "Automated three-dimensional US frame positioning computed from elevational speckle decorrelation.," *Radiology*, vol. 209, no. 2, pp. 575, 1998.

[5] P. J Stolka, H. J Kang, M. Choti, and E. M Boctor, "Multi-DoF probe trajectory reconstruction with local sensors for 2D-to-3D ultrasound," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, 2010, pp. 316–319.

[6] C. Poulsen, P. C Pedersen, and T. L Szabo, "An optical registration method for 3D ultrasound freehand scanning," in *Proceedings of the IEEE Ultrasonics Symposium*, 2005, pp. 1236–1240.

[7] R. Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2008.

[8] G. Simon, A. W Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," in *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, 2000, pp. 120–128.

[9] D. Nistér, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756–770, 2004.

[10] Jean-Yves Bouguet, "Camera Calibration Toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/.

[11] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] P. W Hsu, R. W Prager, A. H Gee, and G. M Treece, "Freehand 3D ultrasound calibration: a review," *Advanced Imaging in Biology and Medicine*, pp. 47–84, 2009.

[13] Graham Treece, Richard Prager, and Andrew Gee, "The Stradwin 3D Ultrasound Acquisition and Visualisation System," http://mi.eng.cam.ac.uk/˜rwp/stradwin/.