A Probe-Camera System for 3D Ultrasound Image Reconstruction

Koichi Ito[1], Kouya Yodokawa[1], Takafumi Aoki[1], Jun Ohmiya[2] and Satoshi Kondo[2]

1 Graduate School of Information Sciences, Tohoku University,
6-6-05, Aramaki Aza Aoba, Sendai, 980-8579, Japan.
2 Konica Minolta, Inc.,
12, Sakura Machi, Takatsuki, 569-8503, Japan.

Abstract. This paper proposes a probe-camera system for 3D ultrasound (US) image reconstruction with probe-camera calibration and probe localization methods. The probe-camera calibration method employs an existing US phantom for convenience with a simple procedure. The probe localization method employs structure from motion (SfM) to estimate the camera motion. SfM is used to reconstruct 3D point clouds from multiple-view images and simultaneously estimate each camera position. Through experiments using the developed system, we demonstrate that the proposed method exhibits good performance to reconstruct 3D US volume.

Keywords: 3D ultrasound, probe localization, structure from motion

1 Introduction

Ultrasound imaging has three key advantages: (i) high spatial resolution, (ii) real-time imaging and (iii) non-invasiveness. US imaging is useful in point-of-care due to such key advantages. Recently, three-dimensional (3D) US has attracted much attention as a valuable imaging tool for a diagnostic procedure. This paper explores 3D US imaging in the point of care. 3D US is acquired by sweeping a US probe around the area of interests and integrating a set of US images to reconstruct 3D volume data. Among 3D US acquisition protocols, we focus on the freehand protocol [2] because of its cost-effectiveness and flexibility. The quality of 3D volume data significantly depends on the accuracy of probe localization.

There are some approaches to estimate the motion of US probes. The first approach is to use electromagnetic (EM) device to know the accurate position of US probes [13, 5]. The accuracy of probe localization is high, while the special devices are required resulting in increasing cost. EM devices also have to be attached on the US probe resulting in interfering the smooth manipulation and EM devices are sensitive to ferromagnetic materials. The second approach is to use an optical tracker to measure the position of US probes [3, 15]. The accuracy of probe localization is relatively high, while the optical sensors are required resulting in increasing cost. The third approach is to use markers to estimate

the motion of US probes [6, 12, 10]. The motion of the US probe is estimated by detecting markers from the captured video sequence. The cost of the system is cheaper than the first and second approaches, since only a camera is required, while markers must be attached on the skin surface, resulting in decreasing the flexibility and acceptability.

Sun et al. [16] proposed a markerless freehand 3D US method. The motion of the US probe is estimated only from a video sequence of skin patterns captured by a low-cost camera using simultaneous localization and mapping (SLAM). This method is cost-effective compared with other methods, but the accuracy has to be improved, since the cumulative estimation error of the probe motion is about 10mm for the probe travel distance of 100mm. Ito et al. [8] proposed a 3D US imaging method using structure from motion (SfM). SfM [17, 4] is one of 3D reconstruction methods in the field of computer vision and is used to reconstruct the sparse 3D point clouds from multiple-view images and simultaneously estimate each camera position. To apply SfM to a video sequence, an accurate method is required to track features between adjacent frames. The motion estimation error is about 2mm for the probe travel distance of 200mm.

The above methods [16, 8] need the geometric relationship between the US probe and the camera so as to estimate the accurate US probe motion. Sun et al. [16] manually measured the position of the US probe and the camera. Ito et al. [8] assumed the translational displacement of the US probe and the camera. In practice, there is a complex movement of the US probe, since the human body consists of soft, curved and complex structures. Calibration between the US probe and the camera is indispensable for reconstructing accurate 3D US volume data. To address the above problem, this paper develops an Ultrasound Probe-Camera System (UPCS) for reconstructing 3D volume data and proposes a calibration method for UPCS. Through experiments using the developed UPCS, we demonstrate that the proposed method exhibits good performance to reconstruct 3D US volume.

2 US Probe-Camera System (UPCS)

UPCS consists of an ultrasound diagnostic system and a camera. US images are acquired by SONIMAGE HS1 (Konica Minolta, Inc.) with L18-4 linear probe (center frequency: 10MHz) as shown in Fig. 1 (a), where the field of view (FOV) of US images is 40×38mm, the frame rate is 30fps and the recording time is 10 seconds, i.e, 300 frames. A camera is C920 (Logicool, Webcam C920), where the image size is 640×480 pixels and the frame rate is 30fps. The camera is attached on the US probe as shown in Fig. 1 (b).

3 US Probe-Camera Calibration

This section describes the proposed calibration method between the US probe and the camera. Fig. 2 (a) shows a flow diagram of the proposed calibration method.

Fig. 1. Developed UPCS for 3D US volume reconstruction: (a) US diagnostic system, (b) camera attached on a US probe and (c) checker pattern used for camera calibration.

First, camera calibration is performed to obtain the intrinsic parameters of the camera, i.e., focal length, image center and lens distortion. We employ the camera calibration toolbox for MATLAB[1]. Next, probe localization is performed using the US phantom, which is used to evaluate the spatial resolution, contrast, and geometry of the US probe. Fig. 2 (b) shows the US phantom used in the experiment, RMI 403GS LE (Gammex, Inc.). This phantom is filled with water-based gels with the appearance of human tissue and includes 8 nylon wires of 0.1 mm in diameter with a 2cm interval. The US probe is located perpendicular to the horizontal plane of the phantom so that 8 wires are on the straight line as shown in Fig. 2 (c). Then, the extrinsic parameters, i.e., rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$, are estimated. Correspondence between the US probe and the camera cannot be used for calibration, since the US probe observes the inside of the phantom, while the camera observes the surface of the phantom. We estimate the geometric relationship between the camera and the checker pattern and between the US probe and the checker pattern, putting a checker pattern on the surface of the phantom.

We assume that the center of the probe coordinate system is the center of the contact area between the US probe and the phantom. Let one of corner points of the checker pattern be the center of the world coordinate system, $\boldsymbol{M}_w$. The rotation matrix $\boldsymbol{R}_p$ and the translation vector $\boldsymbol{t}_p$ from the probe center to $\boldsymbol{M}_w$ are estimated. Note that $\boldsymbol{R}_p$ is an identity matrix, since the US probe is located perpendicular to the horizontal plane of the phantom as mentioned above. Hence, all we have to do is to measure the translational displacement $\boldsymbol{t}_p$ between the probe center to $\boldsymbol{M}_w$. The rotation matrix $\boldsymbol{R}_c$ and $\boldsymbol{t}_c$ from the camera center to $\boldsymbol{M}_w$ can be estimated using the same approach of camera calibration.

The geometric relationship between 3D point $\boldsymbol{M}_c$ in the camera coordinate system and 3D point $\boldsymbol{M}_p$ in the probe coordinate system is defined by

$$\boldsymbol{M}_c = \boldsymbol{R}_c \boldsymbol{M}_w + \boldsymbol{t}_c, \tag{1}$$

$$\boldsymbol{M}_p = \boldsymbol{R}_p \boldsymbol{M}_w + \boldsymbol{t}_p. \tag{2}$$

Eq. (1) is modified as

$$\boldsymbol{M}_w = \boldsymbol{R}_c^T \boldsymbol{M}_c - \boldsymbol{R}_c^T \boldsymbol{t}_c. \tag{3}$$

[1] Camera Calibration Toolbox for MATLAB: `http://www.vision.caltech.edu/bouguetj/calib_doc/`

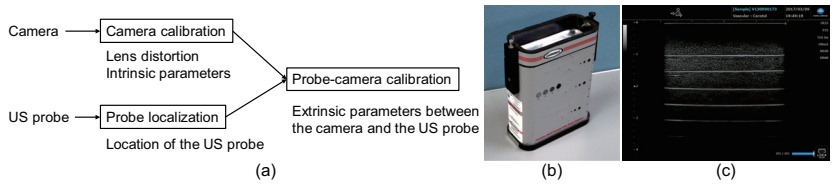

Fig. 2. Calibration between the US probe and the camera: (a) flow diagram of probe-camera calibration, (b) phantom used in the calibration and (c) US image of the phantom.

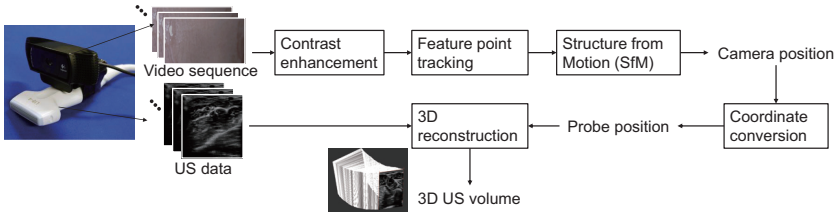

Fig. 3. Flow diagram of 3D US volume reconstruction.

Substituting this equation into Eq. (2), we obtain

$$\boldsymbol{M}_p = \boldsymbol{R}_p \boldsymbol{R}_c^T \boldsymbol{M}_c - \boldsymbol{R}_p \boldsymbol{R}_c^T \boldsymbol{t}_c + \boldsymbol{t}_p \tag{4}$$

$$= \boldsymbol{R}\boldsymbol{M}_c + \boldsymbol{t}, \tag{5}$$

where $\boldsymbol{R} = \boldsymbol{R}_p \boldsymbol{R}_c^T$ and $\boldsymbol{t} = -\boldsymbol{R}_p \boldsymbol{R}_c^T \boldsymbol{t}_c + \boldsymbol{t}_p$. The camera motion can be converted into the probe motion using Eq. (5). Note that Eq. (5) can be derived by estimating only the extrinsic parameter of the camera, if the probe center is set to $\boldsymbol{M}_w$.

4 3D US Volume Reconstruction

This section describes 3D US volume reconstruction from a US video sequence. This paper employs the similar method proposed by Ito et al [8]. Fig. 3 shows a flow diagram of the proposed method, which consists of 5 steps: (i) contrast enhancement, (ii) feature point tracking, (iii) Structure from Motion (SfM), (iv) coordinate conversion and (v) 3D reconstruction. The following is the brief description for each step.

4.1 Contrast Enhancement

Contrast enhancement is applied to captured camera images, since skin texture may not be observed due to ultrasound gels. We employ the method implemented

in MATLAB, which maps the intensity values in the input image I to new values in the output image J such that 1% of data is saturated at low and high intensities of I.

4.2 Feature Point Tracking

This step is to detect feature points from the first frame, track them in the subsequent frame and then obtain the correspondence between adjacent frames. We employ the similar approach proposed by Ishii et al [7]. Let the image frame be $f_i(n_1, n_2)$, where (n_1, n_2) is the pixel coordinate, i indicates the frame index $(0 \leq i \leq N)$ and N is the number of frames. First, feature points are detected from image $f_i(n_1, n_2)$ using the corner detection method proposed by Shi et al [14]. We introduce parameter D to control the density of extracted feature points so as to obtain the stable localization result. If feature points are extracted from the area within $\pm D$ pixels centered on a feature point, these points are removed. Next, we find the corresponding points in subsequent image $f_{i+1}(n_1, n_2)$ from the extracted feature points in current image $f_i(n_1, n_2)$ to track the feature points. We employ a correspondence matching method using Phase-Only Correlation (POC) [18] to find accurate corresponding points. The use of POC makes it possible to obtain the translational displacement with sub-pixel accuracy between small image blocks and evaluate the similarity between images according to the correlation peak of the POC function. If the corresponding point pair has a low correlation value of the POC function, we eliminate it as an outlier. Then, we extract feature points on $f_{i+1}(n_1, n_2)$ from the area without the feature point tracked from $f_i(n_1, n_2)$. By repeating the above processes until the last frame $f_N(n_1, n_2)$, we can obtain a set of tracked feature points through a video sequence.

4.3 Structure from Motion (SfM)

This step is to estimate the rigid-body camera motion, i.e., rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$, using SfM [17, 4]. SfM repeats the linear solution and nonlinear optimization by sequentially adding images. The extrinsic camera parameters $\boldsymbol{R}$ and $\boldsymbol{t}$ of i-th image $f_i(n_1, n_2)$ are estimated in the linear solution using the method proposed by Kneip et al. [9] from the geometric relationship between the reconstructed 3D points and the coordinates of tracked feature points. Note that the extrinsic camera parameters $\boldsymbol{R}$ and $\boldsymbol{t}$ of the first two frames are estimated with the normalized five-point algorithm [11]. We also employ random sample consensus (RANSAC) [1] to robustly estimate the parameters in the first two frames. The 3D points of tracked feature points are obtained using the estimated extrinsic parameters in the i-th image $f_i(n_1, n_2)$ according to triangulation. The reconstructed 3D points and estimated camera parameters are optimized by minimizing reprojection error using bundle adjustments [17, 4]. The reprojection error is defined by the Euclidean distance $||\boldsymbol{m} - \boldsymbol{m}_{\mathrm{rep}}||^2$, where $\boldsymbol{m} = (u, v)$ is a feature point and $\boldsymbol{m}_{\mathrm{rep}}$ is a point obtained by projecting a 3D point $\boldsymbol{M} = (X, Y, Z)$ onto the image using a projection matrix of a camera. We employ

global and local bundle adjustments [17] depending on the target range in this paper. Finally, we obtain the camera motion represented by rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ for each frame and sparse 3D point clouds. The resultant camera motion corresponds to the location of the US probe.

4.4 Coordinate Conversion

The coordinate $\boldsymbol{M}_c$ of estimated camera motion is converted into the coordinate $\boldsymbol{M}_p$ of the probe coordinate system using Eq. (5).

4.5 3D US Volume Reconstruction

This step is to reconstruct 3D US volume from a set of US images. We use Stradwin[2] to reconstruct 3D US volumes from a set of US images and their location in the 3D space obtained by the proposed method.

5 Experiments and Discussion

We evaluate the performance of the proposed method using the dataset acquired by the developed system as shown in Fig. 1. Our dataset consists of camera images and US images acquired from 2 volunteers. We scanned the area around the arm and the thigh. The travel distance of the US probe is about 200mm for the arm and 100mm for the thigh, respectively. The accuracy of US probe localization is evaluated by the accuracy of 3D point clouds reconstructed by SfM, since the accuracy of the camera motion is equivalent to that of 3D point clouds. A 3D mesh model of each target is measured with the laser scanner (Konica Minolta, Inc., VIVID910) for quantitative performance evaluation. The accuracy of 3D reconstruction is evaluated by comparing the reconstructed 3D point clouds and the ground-truth mesh model using the iterative closest point algorithm [19].

Fig. 4 shows the ground-truth 3D model, reconstructed 3D point clouds and estimated camera position and reconstructed 3D US volume for the arm. The shape of reconstructed 3D points is almost the same as the ground-truth 3D model. We observe that the camera is moved straight from the elbow to the wrist. The Root Mean Square (RMS) of reconstruction error is about 2mm for the probe travel distance of 200mm. In the conventional method [16], the cumulative error of the probe localization is about 10mm for the probe travel distance of 100mm.

The thigh area is used to confirm the effectiveness of probe-camera calibration, since this area has the curved shape. Fig.5 shows reconstructed 3D point clouds and camera position and reconstructed 3D US volume for the thigh. The 3D US volume without probe-camera calibration is warped out of shape as shown in Fig.5 (b), while the 3D US volume with probe-camera calibration represents the curved shape of the thigh as shown in Fig.5 (c).

[2] Stradwin: `http://mi.eng.cam.ac.uk/~rwp/stradwin`

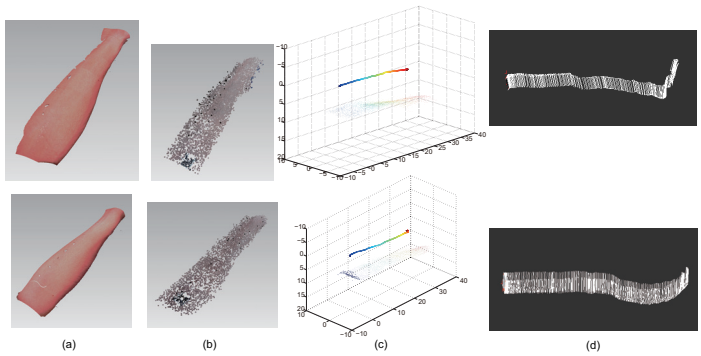

Fig. 4. Experimental results for arm area: (a) ground-truth 3D model, (b) reconstructed 3D point clouds, (c) 3D point clouds and estimated camera position and (d) reconstructed 3D US volume.

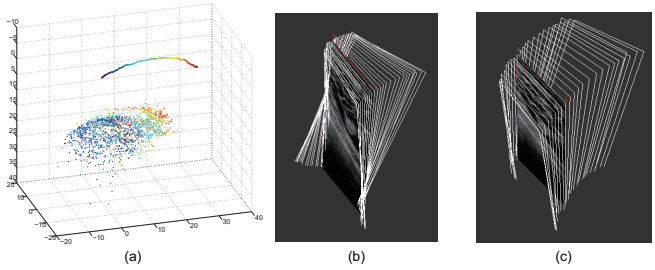

Fig. 5. Experimental results for thigh area: (a) 3D point clouds and camera position, (b) reconstructed 3D US volume without probe-camera calibration and (c) reconstructed 3D US volume with probe-camera calibration.

6 Conclusion

This paper proposed the probe-camera system for 3D US image reconstruction. This paper also proposed the simple method for probe-camera calibration and US probe localization method based on structure from motion. Through a set of experiments, we demonstrated that the location of the US probe can be estimated with about 2mm error for the probe travel distance of 200mm. We expect that the use of the proposed system makes it possible to enhance the effectiveness of US imaging in point-of-care, since 3D US volume can be obtained using a US probe with a camera.

References

1. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm.

ACM 24(6), 381–395 (1981)
2. Gee, A., Prager, R., Treece, G., Berman, L.: Engineering a freehand 3D ultrasound system. Pattern Recognition Letters 24(4–5), 757–777 (2003)
3. Goldsmith, A., Pedersen, P., Szabo, T.: An inertial-optical tracking system for portable, quantitative, 3D ultrasound. IEEE Int'l Ultrasonics Symp. Proc. pp. 45–49 (2008)
4. Hartley, R., Zisserman, A.: Multiple View Geometry. Cambridge University Press (2004)
5. Hastenteufel, M., Vetter, M., Meinzer, H.P., Wolf, I.: Effect of 3D ultrasound probes on the accuracy of electromagnetic tracking systems. Ultrasound in Med. & Biol. 32(9), 1359–1368 (2006)
6. Horvath, S., Galeotti, J., Wang, B., Perich, M., Wang, J., Siegel, M., Vescovi, P., Stetten, G.: Towards an ultrasound probe with vision: Structured light to determine surface orientation. LNCS 7264 (AE-CAI 2011) pp. 58–64 (2012)
7. Ishii, J., Sakai, S., Ito, K., Aoki, T., Yanagi, T., Ando, T.: 3D reconstruction of urban environments using in-vehicle fisheye camera. Proc. IEEE Int'l Conf. Image Processing pp. 2145–2148 (Sep 2013)
8. Ito, S., Ito, K., Aoki, T., Ohmiya, J., Kondo, S.: Probe localization using structure from motion for 3d ultrasound image reconstruction. Proc. Int'l Conf. Medical Imaging pp. 68–71 (2017)
9. Kneip, L., Scaramuzza, D., Siegwart, R.: A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. Proc. Int'l Conf. Computer Vision and Pattern Recognition pp. 2969–2976 (2011)
10. Lange, T., Kraft, S., Eulenstein, S., Lamecker, H., Schlag, P.: Automatic calibration of 3D ultrasound probes. Proc. Bildverarbeitung für die Medizin 2011 pp. 169–173 (2011)
11. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Analysis and Machine Intelligence 26(6), 756–770 (2004)
12. Rafii-Tari, H., Abolmaesumi, P., Rohling, R.: Panorama ultrasound for guiding epidural anesthesia: A feasibility study. LNCS 6689 (IPCAI 2011) pp. 179–189 (2011)
13. Rousseau, F., Hellier, P., Barillot, C.: A fully automatic calibration procedure for freehand 3D ultrasound. Proc. IEEE Int'l Symp. Biomedical Imaging pp. 985–988 (2002)
14. Shi, J., Tomasi, C.: Good features to track. Proc. Int'l Conf. Computer Vision and Pattern Recognition pp. 593–600 (1994)
15. Stolka, P., Kang, H., Choti, M., Boctor, E.: Multi-DoF probe trajectory reconstruction with local sensors for 2D-to-3D ultrasound. Proc. IEEE Int'l Symp. Biomedical Imaging pp. 316–319 (2010)
16. Sun, S.Y., Gilbertson, M., Anthony, B.: Probe localization for freehand 3D ultrasound by tracking skin features. LNCS 8674 (MICCAI 2014) pp. 365–372 (2014)
17. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer-Verlag New York Inc. (2010)
18. Takita, K., Muquit, M.A., Aoki, T., Higuchi, T.: A sub-pixel correspondence search for computer vision applications. IEICE Trans. Fundamentals E87-A(8), 1913–1923 (Aug 2004)
19. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int'l J. Computer Vision 13(2), 119–152 (1994)