Autonomous image-based ultrasound probe positioning via deep learning

Grzegorz Toporek, Haibo Wang, Marcin Balicki and Hua Xie

May 8, 2018

Autonomous image-based ultrasound probe positioning via deep learning

G. Toporek, H. Wang, M. Balicki, H. Xie

Philips Research North America, Cambridge, MA, USA
grzegorz.toporek@philips.com

INTRODUCTION

Although ultrasound (US) is a widely used, non-invasive, and radiation-free imaging modality, manual adjustment of the US probe can be cumbersome and time consuming. An autonomous US scanning device could not only reduce dependence on sonographer's skills and experience but also improve workflow efficiency especially during interventional procedures, such as fine needle aspiration biopsy of the thyroid. Robot-assisted ultrasound imaging has also potential to improve patient care in rural and underserved areas. There are many previous efforts in this direction but none is fully automatic or sufficiently accurate [1], [2].

In this work, as an initial small step towards operator-independent US imaging workflow solution, we developed and evaluated a robot-assisted fully autonomous ultrasound (RAFAUS) probe positioning system. Desired motion of the probe toward the target view is directly derived from anatomical features implicitly extracted via deep neural network; thus, making this technique (a) invariant to anatomical differences, (b) decoupled from the robotic system, (c) registration-free, and (d) independent from any external tracking technologies.

MATERIALS AND METHODS

A 36-weeks fetal US training phantom (CIRS) is used to mimic patient anatomy. Images of the anatomy are captured using an US transducer (X5-1, EPIQ 7, Philips). Probe is rigidly attached to the end effector of a commercial robotic manipulator (Universal Robotics), see Fig. 1a. During the operation of the system user chooses any target view with important anatomical structurers according to a scanning protocol. As soon as the sufficient acoustic coupling is provided velocity of the end effector towards target view is derived from predictions made by a convolutional neural network (CNN), therefore precise calibration between transducer and holder is not necessary. Because prediction accuracy improves proportionally with the distance to the target view (see Fig. 3a), the robot system control loop continuously updates velocity based on the target position estimates from the CNN (see Fig. 1b).

During the development of RAFAUS, an arbitrary reference view (see Fig. 2) is chosen and the entire phantom placed in water bath is automatically scanned from different views according to a pre-defined acquisition scheme. Each US image is labelled with a relative position of the transducer with respect to the reference view and stored as unique data points in the database. The data is divided into two separate sets:

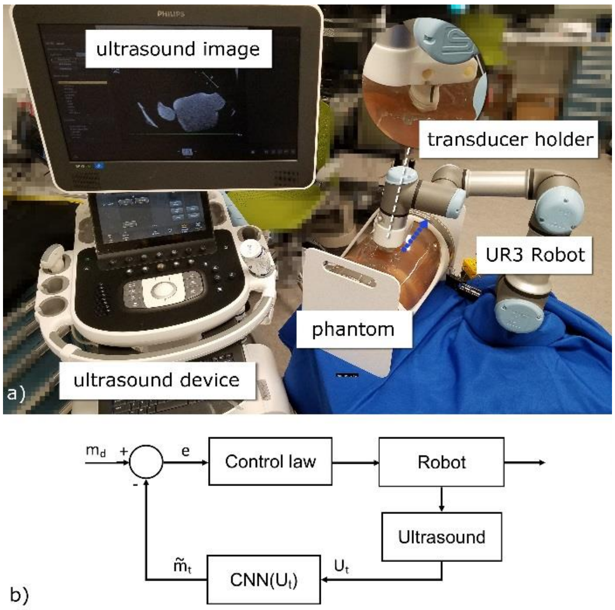

Fig. 1: a) Components of RAFAUS system; US transducer is mounted to serial robot manipulator using custom-made holder and is positioned above US training phantom. Images are sent to processing server with a CNN predicting the motion of the probe (blue arrow) towards the target view in the US probe coordinate frame; b) Schematic diagram of a robot control loop in which the desired Cartesian velocity $\tilde{m}_t$ of the end effector is derived from the prediction provided by a CNN using current US image U_t.

(a) development dataset (38,900 frames) for training the weights of the CNN (of which 15% are randomly chosen for validation), and (b) test dataset (5,400 frames) consisting of data points the model was not trained on (to optimize for generalizability).

In this study, we used a 18-layer CNN with linear residual connections similar to the network described in [3]. In contrast to [3] we increased the size of the last fully connected layer to 2048 and replaced softmax with two 4-dimensional regression layers representing both translational (magnitude and unit direction vector) and rotational (quaternion) components of the rigid transformation. We trained this CNN to predict relative motion of the US transducer towards the reference view using the loss function introduced in [4]. Adaptive Moment Estimator (Adam) optimizer was used as optimization function with an initial learning rate of 0.001, and being decayed every 8,000 samples with an exponential rate of 0.5. Batch normalization, and pre-initialization with weights from the same network trained on ImageNet dataset were used. We used early stopping criteria, and a dropout probability of 0.4 before the last two regression layers to avoid overfitting. We set batch size to 64 and augmented the training data using scale and aspect ratio augmentation. Input images were center-cropped and normalized based on their mean values and standard deviations to the range of 0–1.

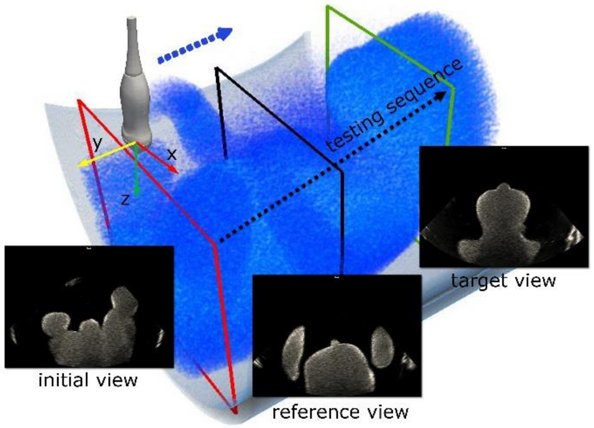

Fig. 2: Schematic overview of a testing sequence (5,400 frames), overlaid on top of a 3D reconstruction of the US phantom, on which prediction accuracy was evaluated. It starts at an initial view located near lower limbs (red outline) and finishes at a target view located at the fetus head (green outline). For the sake of clarity, reference view that was used during the development of the deep learning model is shown as a black outline. Additionally, a predicted motion of the probe (blue dashed line) is depicted in the coordinate system of the ultrasound transducer. This coordinate system is used to calculate errors in the result section (see Fig. 3a-b).

RESULTS

We evaluated the accuracy of RAFAUS using a challenging motion sequence (5,400 frames, see Fig. 2). The average translational accuracy our system achieved was 2.38 ± 2.73 mm, 3.11 ± 3.08 mm, and 1.09 ± 1.29 mm along x, y, and z axis respectively (see Fig. 3b). The overall translational accuracy increased when the distance to the target position decreased (see Fig. 3a). The average prediction inaccuracy measured around 90 mm from the target view was significantly higher (p-value < 0.0001, unpaired, two-tailed t test) than at 20 mm, 4.67 ± 2.8 mm and 1.41 ± 0.88 mm respectively. The average rotational accuracy was 0.68 ± 0.58°, 0.45 ± 0.26°, and 1.72 ± 1.03° around Z, Y, and X axis respectively (see Fig. 3c). We evaluated robustness of the RAFAUS in challenging imaging conditions (e.g. compromised acoustic windows and externally induced phantom motion); quantitative evaluation is not provided in the paper for the sake of brevity.

DISCUSSION

In this work we demonstrated that US transducer can be autonomously positioned with respect to a target view, at an average translational accuracy of 1.41 ± 0.88 mm. The image-based positional feedback using deep learning continuously estimates target position as robot converges on the target view. Naturally, the target position estimate accuracy increased when probe was closer to the target, thus suggesting potential for incorporating a series of former predictions into the deep learning model, for instance, by using temporally recurrent layers, such as Long Short-term Memory (LSTM) units. A clinical system will require safety measures, such as precise regulation of patient-contact forces, collision monitoring, and adjustment of the acoustic coupling. Force control methods will definitely add to the safety by maintaining constant contact and optimal force for image acquisition. The main limitation of this study is the usage of a single phantom for both training and testing. The phantom is static and absent of strong acoustic artifacts, e.g. reverberation artifacts, which are normally encountered in a clinical scan.

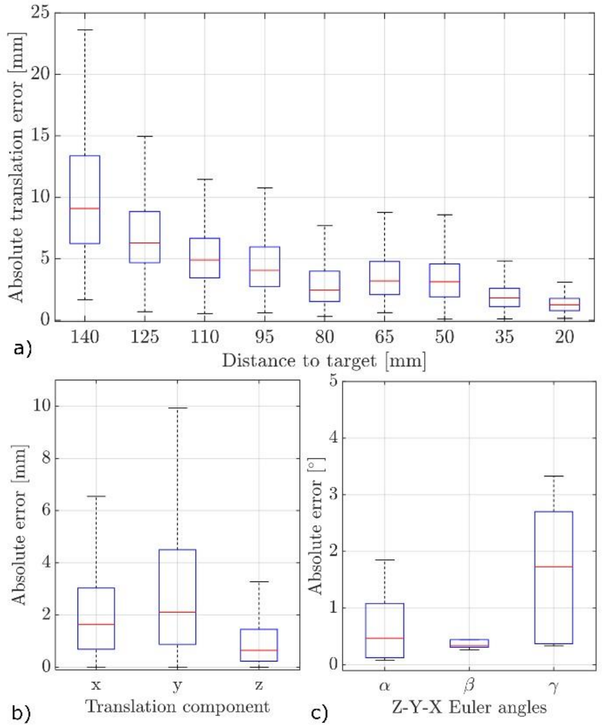

Fig. 3: Boxplots of the pose prediction errors calculated on the testing sequence (see Fig. 2) consisting of 5,400 frames; boxplot (a) shows a prediction accuracy at various distances from the target view along the testing sequence; an overall system accuracy separated into (x, y, z) translational components (b) as well as rotation angles (c) around each axis (Z, Y, X) are also presented.

REFERENCES

[1] A. Safwan, B. Mustafa, S. Member, T. Ishii, Y. Matsunaga, R. Nakadate, H. Ishii, K. Ogawa, A. Saito, M. Sugawara, K. Niki, and A. Takanishi, “Development of Robotic System for Autonomous Liver Screening Using Ultrasound Scanning Device,” *Proceeding IEEE Int. Conf. Robot. Biomimetics*, no. December, 2013.

[2] K. Liang, A. J. Rogers, E. D. Light, von D. Allmen, and S. W. Smith, “3D Ultrasound Guidance of Autonomous Robotic Breast Biopsy: Feasibility Study,” *Ultrasound Med Biol*, vol. 36, no. 1, pp. 173–177, 2010.

[3] K. He, X. Zhang, S. Ren, “Deep Residual Learning for Image Recognition,” *arXiv:1512.03385v1*, 2015.

[4] A. Kendall and K. College, “PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization,” *Comput. Vis. Pattern Recognit.*, 2016.