Deep Learning-Based 3D Freehand Ultrasound Reconstruction with Inertial Measurement Units

Raphael Prevost **Mehrdad Salehi** **Simon Jagoda** **Navneet Kumar**
Alexander Ladikos **Oliver Zettinig** **Wolfgang Wein**
ImFusion GmbH
Munich, Germany
prevost@imfusion.de

Julian Sprung **Robert Bauer**
piur Imaging GmbH
Vienna, Austria

Abstract

This work aims at reconstructing 3D ultrasound volumes from sequences of freehand images acquired with standard 2D probes without any expensive or cumbersome external tracking hardware. We extend our previous method based on deep learning to integrate and also learn from measurements of an inertial measurement unit (IMU). Our system is evaluated on a dataset of 600 in vivo ultrasound sweeps, yielding accurate reconstructions with a median normalized drift of 5.2% even on long sweeps with complex trajectories, hence paving the way towards translation into clinical routine.

This extended abstract is adapted from a journal paper to appear [4].

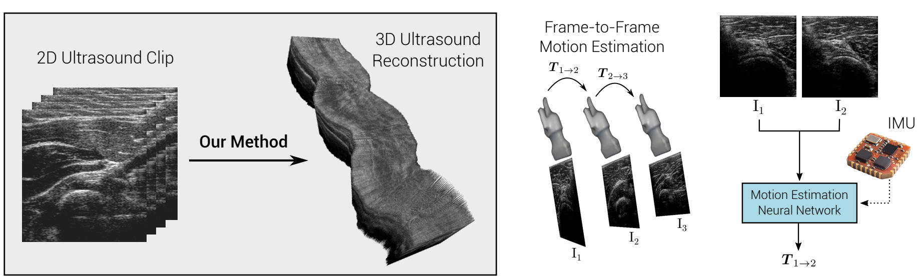

Figure 1: Our system turns 2D standard ultrasound clips into 3D volumes. The method is based on a frame-to-frame motion estimation performed by a neural network with the help of an IMU.

1 Introduction

Ultrasound imaging (US) combines a number of advantages that makes it a popular medical imaging modality: it is affordable, safe for both patients and clinicians, and is convenient to set up and use. Since a lot of clinical applications require three dimensional data (*e.g.* measurements, registration), a significant effort has been dedicated to the development of 3D US systems. This is usually done by acquiring a series of 2D images over the region of interest and then compounding them into an actual 3D volume. Such a solution requires the knowledge of the relative position from one image to the next, which is often achieved via special probes or external (electromagnetic or optical) tracking

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

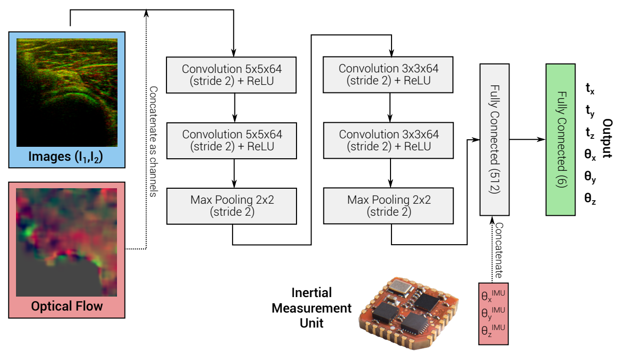

Figure 2: Architecture of our convolutional neural networks.

The main input is the pair of frames encoded as a 2-channel image while the output is a 6-dimensional vector representing the parameters of the transformation. The two other additional inputs are the optical flow vector field and the measures of an IMU.

systems. Our goal is instead to build a system capable of reconstructing 3D ultrasound volumes from a series of 2D images without any expensive or cumbersome hardware.

Such a system must therefore be able to estimate the motion of the probe between two successive frames by using solely the image content of those frames (see Figure 1). This has been an active topic of research since seminal papers introduced the *speckle decorrelation* approach [1, 2]. It relies on the fact that two successive frames are highly correlated and consists in estimating a displacement field by modeling the correlation of image patches as a function of their out-of-plane distance. Despite recent improvements, this algorithm unfortunately still has not proved itself accurate enough to be used in a clinical product. Recently, we proposed in [3] a completely different approach based on deep learning. By alleviating the necessity of designing a complex physical model of the image formation but instead relying on a statistical analysis of a large number of data, we were able to demonstrate significant improvements in terms of accuracy of the reconstruction.

However, while this method works well for ultrasound sweeps acquired with a simple linear motion, the accuracy decreases significantly whenever the probe motion contains out-of-plane rotational components. This is because the motion-from-image estimation problem is ill-posed and it is for instance very difficult (if not impossible) to understand in which direction an out-of-plane rotation happens. In order to accommodate more realistic motions, we therefore propose to augment our system with an inertial measurement unit (IMU), as suggested by [5]. Such chips have a very small footprint (less than 1cm) but contain a gyroscope that provides accurate estimates of the device orientation. We will show that integrating such an information in the deep learning-based method is straightforward and provides significant improvements.

2 Methods

Our method uses deep learning in an end-to-end approach. We design a neural network that takes as input a pair of successive frames, and is trained to output a vector of 6 parameters (3 translations, 3 rotations) that represent the relative probe motion between the two frames. Following [3], we also use a pre-computed optical flow between the two images as additional input channels, which is a valuable information about the in-plane motion.

The new information, namely the IMU orientation, can be integrated into such a network architecture by simply concatenating the three rotation angles θ_x^{IMU}, θ_y^{IMU}, θ_z^{IMU} measured by the IMU to the penultimate layer of the network, as shown in Figure 2.

Network Training We acquired a large set of ultrasound sweeps on a Cicada® research ultrasound machine (Cephasonics, Inc.) with a linear probe composed of 128 elements at a frequency of 5 MHz, onto which we mounted an Xsens MTi-3 (Xsens Technologies) IMU chip. In order to acquire a ground truth, the probe was also equipped with an optical target tracked by a Stryker NAV3™ (Stryker Co.) system originally intended for surgical navigation. This allowed us to record the position and the orientation of the probe during the acquisitions with a much higher accuracy and precision than standard electromagnetic systems. Using such data, the weights of the layers are then trained by minimizing the squared Euclidean distance between the network output and the ground truth.

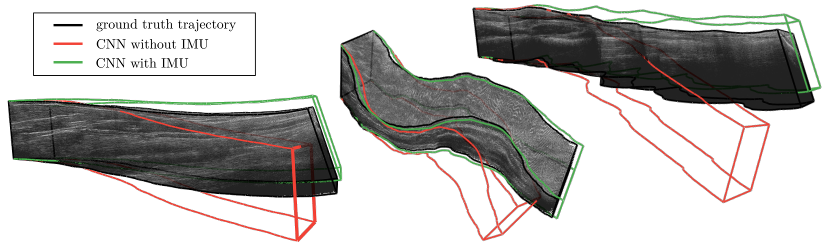

Figure 3: Comparison of the trajectories reconstructed with and without incorporation of the IMU information in the convolutional neural network on three sweeps with weaving and tilting motions.

Data		**Average absolute error [mm/°]**						**Final drift [mm]**		
Flow	**IMU**	t_x	t_y	t_z	θ_x	θ_y	θ_z	**min.**	**med.**	**max.**
×	✓	6.56	7.23	16.70	0.94	2.65	2.80	3.12	29.22	186.83
✓	×	8.89	6.61	5.73	5.21	7.38	4.01	3.22	27.34	139.02
✓	✓	2.75	2.41	4.36	0.19	0.21	0.13	0.76	10.42	35.22

Table 1: Error metrics depending on the input of the network (with or without optical flow channels and IMU information). The parameter-wise errors are computed for each frame of each sweep with respect to the absolute (not frame-to-frame) transformation. Final drift is the distance between the estimated center and the true center position of the last frames.

Trajectory Estimation The trajectory reconstruction algorithm then consists in feeding the successive pair of images to the neural network and chaining all the estimated transformations. The algorithm can therefore be executed in real-time on a standard computer.

3 Experiments and Results

We conducted several studies on a large dataset composed of more than 600 real ultrasound sweeps acquired on the forearms of 15 volunteers, with diverse trajectories that cover the potential motions that can occur during an actual clinical exam. Each sweep contains in average 500 frames and spans over 20 cm. This is, to the best of our knowledge, the largest and most clinically representative database used in a study on freehand 3D ultrasound reconstruction without external tracking.

We trained multiple networks with/without the optical flow channels and with/without the IMU measurements, following a two-fold cross-validation. The results are reported in Table 1 and clearly indicate the benefits of using an IMU to get an accurate reconstruction of complex trajectories. Qualitative results shown in Figure 3 also confirm our hypothesis. The median drift of the system is around 10 mm after approximately 20 cm, which represents a 5.2% drift and a 3.4% error on length measurements. The detail of all experiments will be available in our journal paper [4]. Finally, a video demonstrating our first prototype is available at `https://www.youtube.com/watch?v=TXNFbzYKRZY` and shows the potential of our system for clinical applications.

References

[1] Jian-Feng Chen, J. Brian Fowlkes, Paul L. Carson, and Jonathan M. Rubin. Determination of scan-plane motion using speckle decorrelation: Theoretical considerations and initial test. *International Journal of Imaging Systems and Technology*, 8(1):38–44, 1997.

[2] Prager et al. Sensorless freehand 3-d ultrasound using regression of the echo intensity. *Ultrasound in medicine & biology*, 29(3):437–446, 2003.

[3] Prevost et al. Deep learning for sensorless 3d freehand ultrasound imaging. In *MICCAI 2017 Proceedings, Part II*, pages 628–636, 2017.

[4] Prevost et al. 3d freehand ultrasound without external tracking using deep learning. *Medical Image Analysis - MICCAI 2017 Special Issue*, to appear.

[5] R.J. Housden, Andrew H. Gee, Richard W. Prager, and Graham M. Treece. Rotational motion in sensorless freehand three-dimensional ultrasound. *Ultrasonics*, 48(5):412 – 422, 2008.