Precise Ultrasound Bone Registration with Learning-Based Segmentation and Speed of Sound Calibration

Mehrdad Salehi[1,2(✉)], Raphael Prevost[1], José-Luis Moctezuma[3], Nassir Navab[2], and Wolfgang Wein[1]

[1] ImFusion GmbH, Munich, Germany
salehi@imfusion.de
[2] Computer Aided Medical Procedures (CAMP), TU Munich, Munich, Germany
[3] Stryker Leibinger GmbH & Co. KG, Freiburg, Germany

Abstract. Ultrasound imaging is increasingly used in navigated surgery and registration-based applications. However, spatial information quality in ultrasound is relatively inferior to other modalities. Main limiting factors for an accurate registration between ultrasound and other modalities are tissue deformation and speed of sound variation throughout the body. The bone surface in ultrasound is a landmark which is less affected by such geometric distortions. In this paper, we present a workflow to accurately register intra-operative ultrasound images to a reference pre-operative CT volume based on an automatic and real-time image processing pipeline. We show that a convolutional neural network is able to produce robust, accurate and fast bone segmentation of such ultrasound images. We also develop a dedicated method to perform online speed of sound calibration by focusing on the bone area and optimizing the appearance of steered compounded images. We provide extensive validation on both phantom and real cadaver data obtaining overall errors under one millimeter.

1 Introduction

Navigated surgery in the orthopedic domain often requires intra-operative registration of a bone surface to a pre-operative CT or MRI. Techniques established in clinical routine usually reconstruct solely the bone surface area exposed after the surgical incision with a tracked pointer. As an alternative, ultrasound (US) may be used in the operating theater to image and reconstruct a larger portion of the bone surface, which can henceforth be registered to a surface segmented from the pre-operative volume. However, inherent inaccuracies due to the physics of ultrasound such as speed of sound variations and refractions, as well as challenges for precise tracked ultrasound often limit the accuracy that may be achieved. Existing automated approaches usually employ custom image processing algorithms to detect the bone surface in individual ultrasound frames [1,2], followed

M. Salehi and R. Prevost contributed equally to this paper.

© Springer International Publishing AG 2017
M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 682–690, 2017.
DOI: 10.1007/978-3-319-66185-8_77

by a surface registration. These methods have severe limitations due to the high variability of bones appearance and shape.

In this paper, we propose a fast ultrasound bone surface detection algorithm using a fully convolutional neural network (FCNN) that is both more robust and efficient than previous methods. This real-time detection is then leveraged to build further contributions: (*i*) we developed a novel automatic speed of sound analysis and compensation method based on steered compound ultrasound imaging around the detected bone; (*ii*) we propose an automatic temporal calibration between tracking and ultrasound also based on the bone; (*iii*) we prove on both phantom and human cadaver studies that the aforementioned contributions allow us to design a system for CT-US registration with a sub-mm accuracy.

2 Methods

Bone Detection and Segmentation in US Images. Despite recent research [2,3], bone detection in US is still very challenging due to the variable appearance and the weak shape prior of bones. We propose here a simple method based on deep learning that is able to overcome those challenges and produce accurate bone probability maps in a more robust way than standard feature-based methods. First, we train a fully convolutional network [4] on a set of labeled images, where the bone area has been roughly drawn by several users. Our classification network is inspired the U-Net [5] and consists of a series of 3×3 convolutional layers with ReLU non-linearities and max-pooling layers, followed by deconvolutional layers and similar non-linearities. Its output is a fuzzy probability map with the same size as the input image.

From this bone probability map, we extract the bone surface for each scanline as suggested in [6], *i.e.* as the center pixel between the maximum gradient and the maximum intensity along the scanline. Some previous works use a more elaborate dynamic programming approach [3], however our experiments showed that the FCNN output was so reliable that simple thresholding and largest component analysis was enough to discard most outliers. Some results on very different images are shown in Fig. 1. Thanks to its simplicity, our method is thus able to run in real-time (30 images per second) on a standard computer, which enables us to leverage its results within dedicated online algorithms as detailed below.

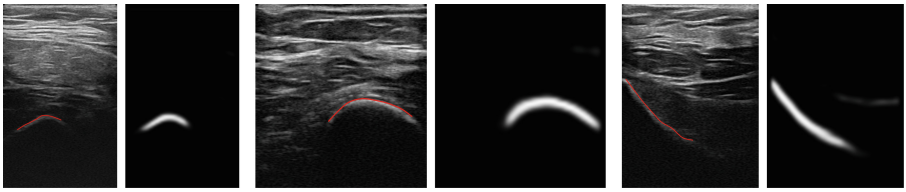

Fig. 1. Examples of automatic bone segmentations in various US images (different bones and acquisition settings), along with the neural network detection map.

Online Speed of Sound Calibration Using Bone Detection. In conventional delay-sum ultrasound beamforming, wrong speed of sound expands or compresses images along the direction of beams. This effect causes misalignment when imaging an object from different angles. We estimate the average speed of sound between the transducer and the bone surface by optimizing the appearance of super-imposed steered frames. Given two steered images I and J, we are interested in the speed of sound c which minimizes the following cost function: $f(I, J, c) = \frac{1}{|S|} \sum_{p \in S} |I_p^c - J_p^c|$, where S is the set of all pixels within the bone region of interest in image I^c; I_p^c and J_p^c are corresponding pixel intensities in the images after compounding with the speed of sound c.

A major obstacle when comparing steered ultrasound images is that the reflection from most tissue boundaries and objects depends on the insonification angle. Hence, to gain more consistency in the optimized value of sound speed, we had two considerations: (i) non-bone areas are masked out, (ii) instead of directly comparing two steered frames, which increases dissimilarities in point-spread-function, each of the left and right steered frames (I_l and I_r) are compared to the perpendicular image (I_m). The final estimation for the speed of sound is then defined as the minimum of the sum of $f(I_l, I_m, c)$ and $f(I_r, I_m, c)$ be computed by a simple exhaustive search in a few seconds. Figure 2 shows the benefits of correcting the speed of sound with our approach.

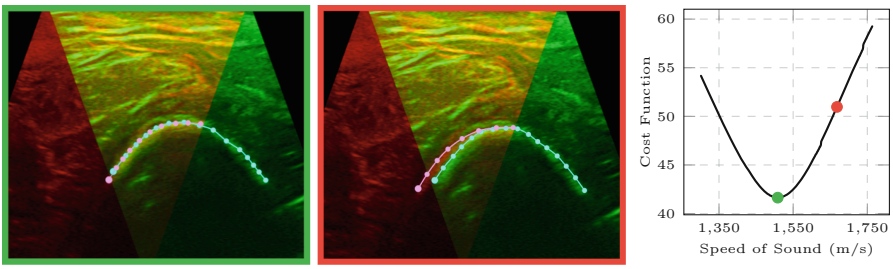

Fig. 2. Steered images compounded with correct speed of sound (left) and 10% error (middle) with their corresponding bone segmentation. Notice the better consistency on the left image. (Right) speed of sound calibration cost function of the two shown compounded images.

Comparing to other local speed of sound estimation methods [7,8], our approach is less general, but it achieves significant results and needs fewer ultrasound images, which makes it suitable for real-time applications. Furthermore, it leverages the bone detection to avoid being sensitive to tissue inconsistencies prevalent in *in-vivo* clinical data.

Automatic Temporal Calibration. Precise spatial and temporal calibration is necessary to figure out the relative transformation between tracking sensor and US image coordinates. We use an image-based spatial calibration similar

to [9], with extensions to optimize over multiple recordings and handle more arbitrary geometries. Still, the temporal synchronization shift parameter has to be optimized separately since it directly influences the sweep geometry and may create ambiguity. It is also affected by more ultrasound imaging settings, such as spatial compound frames and number of focal zones. Therefore, we have developed a dedicated automatic temporal calibration method exploiting the fact that the bone surface remains rigid while the soft tissue is compressed when pushing the probe. Since the error caused by wrong temporal calibration is detected easier in sweeps exhibiting fast motion, we used the following protocol:

1. An ultrasound sweep is recorded from a bone surface while the US probe is slowly pushed towards the bone and released several times.
2. Our proposed bone segmentation method is applied on all images and a 3D point cloud is extracted.
3. The whole point cloud is projected onto the average direction of ultrasound scanlines, as shown in Fig. 3. The optimal temporal lag is the value that minimizes the variance of the 1D coordinate of those projected points.

Figure 3 illustrates the principles of this method; the US sweep in the figure is expanded along the bone extent to better visualize the motion (sweeps used for calibration only move perpendicular to the bone).

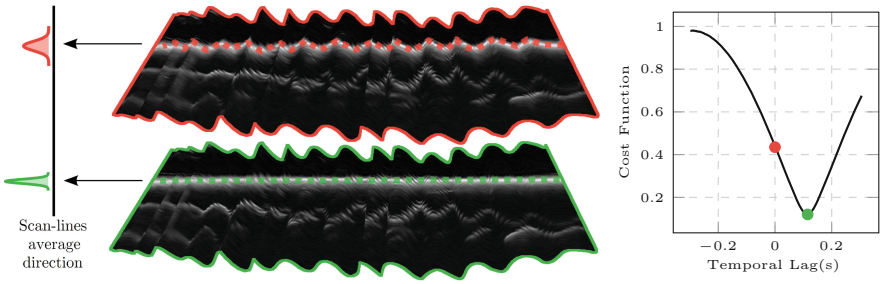

Fig. 3. Reconstruction of a sample sweep before (red) and after (green) optimization of the temporal lag. The cost function shows a strong global minimum in all cases. Standard deviation of repeated calibrations on human femur is 2 ms.

Registration to Pre-operative Data. Once we have properly calibrated the US system, we can address the actual registration. Assuming the availability of an accurate CT segmentation, we formulate our registration as a point-to-surface distance minimization problem where we minimize the sum of the absolute distance of all points extracted from the US sweeps to the pre-operative CT bone surface. This problem is solved via a global optimizer called *DiRect* [10]; in order to avoid local minima and allow for automatic initialization during surgery, the bounding search space was set to [−300 mm; 300 mm] for translations and

$[-100°; 100°]$ for rotations. For faster evaluation of this cost function, we precompute a signed distance transform from the CT segmentation. Once the global optimizer has found a suitable minimum, we finally refine the transformation estimate with a more local Nelder-Mead Simplex method, after removing the outliers (points further away than a distance of 5 mm after the first registration).

3 Experiments and Results

Bone Detection. In order to evaluate the bone detection method, we manually labeled an independent set of 1382 US images from different volunteers, different bones (femur, tibia, and pelvis) and various acquisition settings. Ultrasound images were recorded with different frame geometries, image-enhancement filters, brightness, and dynamic contrast to assure that US bone detection algorithm does not overfit to a specific bone appearance. Scan geometries consisted of linear, trapezoid, and steered compound images with 3 consecutive frames. We ran a 2-fold cross-validation to evaluate all machine learning methods. We compared our bone localization map based on deep learning to two other approaches: an implementation of the hand-crafted feature-based method proposed in [11] and a Random Forest similar to [3]. Results in Table 1 show the superiority of the deep learning method both in terms of precision and recall.

Table 1. Performance of the different bone localization methods. The two reported numbers are respectively the mean and median over the 1382 images.

	Feature-based [11]	Random forest	Neural network
Precision	0.50/0.58	0.76/0.83	0.87/0.91
Recall	0.39/0.42	0.83/0.93	0.87/0.94
Dice coefficient	0.44/0.49	0.79/0.88	0.87/0.92

Wire Phantom Calibration Validation. In order to demonstrate that the proposed speed of sound calibration method works in principle, several experiments were implemented on phantom data. A 4×4 wire grid was created on Perklab fCal-3.1 phantom and immersed in a water tank with 22.5 °C temperature. Based on the temperature and salinity of the water, the expected sound speed was 1490 m/s. Three steered US images with $-5°$, $0°$, and $+5°$ angles were recorded from the wire grid with imaging depth of 13 cm. Wires were positioned with 1 cm spacing at depth of 9 cm to 12 cm. Our system assumed a default speed of sound 1540 m/s during the recording. We ran our speed of sound calibration algorithm on those images (without the bone masking step) and obtained an estimated speed of 1497 m/s, representing an error of 0.47%, which means that the calibration method was able to compensate the initial error of 3.35%.

Femur Phantom Calibration Validation. In order to investigate effects of using the bone detection during the calibration, we performed several experiments using a Sawbones® femur phantom immersed in the same water tank. We recorded several US sweeps from different directions such that the extracted point cloud covers the femur's surface, including head, neck, and trochanters. For a large range of speeds of sound (1400–1600 m/s), we compounded the US images, extracted the bone point cloud, and registered it to the mesh model as described in Sect. 2. The final surface error is plotted in Fig. 4 as a function of the speed of sound. The average point-to-surface error between the mesh model and the extracted point cloud from US images was around 0.2 mm. One can see that the optimal speed of sound was reached at 1502 m/s, which has 0.8% deviation from the expected value in our water tank. Our online estimation based on the bone detection also agreed with those values and differed, on average, by $+0.37 \pm 0.61\%$ from the true speed of sound ($+5.5 \pm 9.1$ m/s). In comparison, using the whole image yielded a significantly different estimate, with an error of $+2.98 \pm 0.58\%$, and therefore a higher surface error. In general, the errors in our speed of sound estimation method are comparable to previous studies [7,8] but are obtained with a much simpler approach.

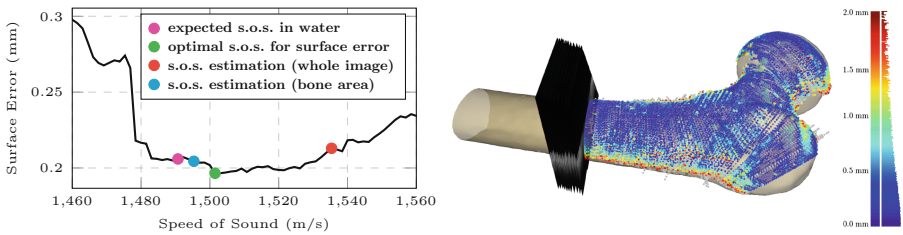

Fig. 4. Registration study on the bone phantom in water. (Left) the optimal surface agreement after registration is obtained for a speed of sound of 1502 m/s. (Right) visualization of the optimal registered point cloud, color-coded with the surface distance.

Cadaver Study. We finally performed a study of the overall system accuracy on two human cadavers. CT scans ($0.5 \times 0.5 \times 1$ mm resolution) of the two legs of both cadavers were acquired after implanting six multi-modal spherical fiducials into each of the bones of interest, namely pelvis, femur, and tibia. Manual CT bone segmentations were obtained by domain experts. However, in order to achieve a voxel-wise accurate bone surface, we had to refined all our segmentations using a 3D *guided filter*, recently introduced in the computer vision community [12] as a fast and precise method for image matting. We finally extracted the bone surface by running a marching cubes on this fuzzy map. In total, 142 tracked ultrasound sweeps were recorded by two orthopedic surgeons, with a linear 128-element probe at 7.5 MHz center frequency on a Cephasonics cQuest Cicada® system (0.2×0.08 mm resolution). They chose different US settings depending on the scanned bone, e.g. depth was between 4 and 7 cm. The Stryker Navigation

System III camera [13] was used with a reference tracking target fixed to the bone and another target on the ultrasound probe. Accuracy of the tracking targets were close to 0.2 mm.

In order to generate a ground truth registration between the CT and US images, we extracted the positions of the fiducials from the CT images and touched them just before the US acquisition with a dedicated tracked pointer. For each bone of each leg, we then rigidly registered the two sets of fiducials to obtain our ground truth (mean residual error 0.69 mm, median 0.28 mm).

After performing all calibration steps (temporal, spatial, and speed of sound), we found out that in average, tibia cases needed a −4% speed of sound correction compared to the system's default while femur and pelvis cases required a −1.5% correction. After such a compensation, we registered the extracted point clouds from multiple US sweeps to the CT surface and compared the obtained rigid transformation to the corresponding ground truth. We report the final accuracy metrics of the whole workflow in Table 2, summarized over all cases; the median case in terms of accuracy is shown in Fig. 5. For each case, on average, the US bone segmentation and registration took one minute on a standard computer.

Table 2. Median registration errors of all sweeps (resp. surface registration error, fiducial errors, and relative error for translations and rotations parameters). All errors are given in mm except rotations which are given in degrees. Translations errors depend on reference coordinate system (the bone mesh center was used here). Each *case* is a series of US sweeps (Avg. 5) acquired over a particular bone of one cadaver by one surgeon.

Bone type	Error surface	Error fid.	Error T1	Error T2	Error T3	Error R1	Error R2	Error R3
Femur (14 cases)	0.62	2.41	2.71	1.92	1.20	0.34	0.43	1.38
Tibia (5 cases)	0.50	2.12	3.91	1.01	1.48	0.16	0.30	2.02
Pelvis (9 cases)	0.60	2.76	2.66	3.27	2.46	0.92	0.64	1.12

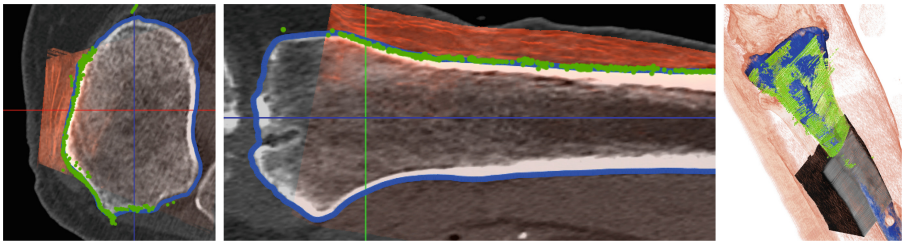

Fig. 5. Visualization of the median case (in terms of accuracy) of cadavers tibia. An ultrasound sweep has been superimposed in red on the CT image with the result of the point-to-surface registration

For all types of bones, we obtained a sub-mm average surface error, which shows that tracked US imaging enables to retrieve the bone shape very accurately, approaching the limit imposed by the 0.5 mm CT voxel spacing. The

fiducial errors were slightly above 2 mm. Because the fiducials were placed far apart (regularly 3 on the proximal and 3 on the distal end of the femur) but US sweeps were more locally confined, this constitutes a worst-case upper bound on the error. Defining local coordinate systems and target points for different implant placement or surgical procedures would result in lower errors, which is also independently confirmed by the small rotation errors.

4 Conclusion

We have developed a CNN-based ultrasound bone detection algorithm which yields complete surface coverage even under difficult imaging conditions, while offering real-time performance. We used it to power additional novel methods for speed of sound estimation and precise temporal calibration. This allowed us, to our knowledge for the first time, to put together an overall system for intra-operative bone surface registration in computer-aided orthopedic surgery applications that utilizes ultrasound and yields sub-mm surface error. It therefore has great clinical potential, especially in all navigated orthopedic scenarios where currently the bone surface has to be manually digitized with a pointer.

While the proposed speed of sound compensation method yields physically plausible results on phantom data, the improvements on our cadaver experiments are within the order of magnitude of the ground truth error; hence, future work is required to investigate its effects. It might also be beneficial to look into separate speed of sound compensation for fat and muscle layers, which can be achieved by pairing our method with a tissue classification algorithm.

References

1. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A., Rohling, R.: Automatic adaptive parameterization in local phase feature-based bone segmentation in ultrasound. Ultrasound Med. Biol. **37**(10), 1689–1703 (2011)
2. Ozdemir, F., Ozkan, E., Goksel, O.: Graphical modeling of ultrasound propagation in tissue for automatic bone segmentation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 256–264. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_30
3. Baka, N., Leenstra, S., Van Walsum, T.: Machine learning based bone segmentation in ultrasound. In: CSI MICCAI Workshop (2016)
4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of IEEE Conference: CVPR, pp. 3431–3440 (2015)
5. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
6. Jain, A.K., Taylor, R.H.: Understanding bone responses in b-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework. In: Proceedings of SPIE Medical Imaging 2004, pp. 131–142 (2004)

7. Jaeger, M., Held, G., Preisser, S., Peeters, S., Grnig, M., Frenz, M.: Image-based method for in-vivo freehand ultrasound calibration. In: Proceedings of SPIE, vol. 9040 (2014)
8. Shin, H.C., Prager, R., Gomersall, H., Kingsbury, N., Treece, G., Gee, A.: Estimation of speed of sound in dual-layered media using medical ultrasound image deconvolution. Ultrasonics **50**(7), 716–725 (2010)
9. Wein, W., Khamene, A.: Image-based method for in-vivo freehand ultrasound calibration. In: Proceedings of SPIE, Medical Imaging 2008, vol. 6920 (2008)
10. Jones, D.R.: Direct global optimization algorithm. In: Floudas, C.A., Pardalos, P.M. (eds.) Encyclopedia of Optimization, pp. 431–440. Springer, Heidelberg (2001). doi:10.1007/0-306-48332-7_93
11. Wein, W., Karamalis, A., Baumgartner, A., Navab, N.: Automatic bone detection and soft tissue aware ultrasound-CT registration for computer-aided orthopedic surgery. IJCARS **10**(6), 971–979 (2015)
12. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2013)
13. Elfring, R., de la Fuente, M., Radermacher, K.: Assessment of optical localizer accuracy for computer aided surgery systems. Comput. Aided Surg. **15**(1–3), 1–12 (2010)