

Skeleton Graph-based Ultrasound-CT Non-rigid Registration

Zhongliang Jiang¹, Xuesong Li¹, Chenyu Zhang¹, Yuan Bi¹, Walter Stechele², and Nassir Navab¹, *Fellow, IEEE*

Abstract—Autonomous ultrasound (US) scanning has attracted increased attention, and it has been seen as a potential solution to overcome the limitations of conventional US examinations, such as inter-operator variations. However, it is still challenging to autonomously and accurately transfer a planned scan trajectory on a generic atlas to the current setup for different patients, particularly for thorax applications with limited acoustic windows. To address this challenge, we proposed a skeleton graph-based non-rigid registration to adapt patient-specific properties using subcutaneous bone surface features rather than the skin surface. To this end, the self-organization mapping is successively used twice to unify the input point cloud and extract the key points, respectively. Afterward, the minimal spanning tree is employed to generate a tree graph to connect all extracted key points. To appropriately characterize the rib cartilage outline to match the source and target point cloud, the path extracted from the tree graph is optimized by maximally maintaining continuity throughout each rib. To validate the proposed approach, we manually extract the US cartilage point cloud from one volunteer and seven CT cartilage point clouds from different patients. The results demonstrate that the proposed graph-based registration is more effective and robust in adapting to the inter-patient variations than the ICP (distance error mean \pm SD: 5.0 \pm 1.9 mm vs 8.6 \pm 6.7 mm on seven CTs).

Index Terms—Medical robotics, robotic ultrasound, non-rigid registration, graph-based registration, ultrasound bone imaging, intercostal intervention, path transferring

I. INTRODUCTION

Medical ultrasound (US) is one of the most widely used imaging modalities in daily clinical practice. Since US imaging is real-time, non-ionizing radiation, and widely accessible, it is often used as the standard approach for internal organ examinations [1], [2] and imaging-guided intervention [3], particularly in the preliminary healthcare industry. Nevertheless, the conventional US examination is highly user-dependent, as the resulting image is sensitive to the acquisition parameters, i.e., contact force and probe orientation [4]. Besides the deep understanding of the physiological background, sonographers are required to simultaneously optimize the acquisition parameters toward high-quality images during US examinations, which often need years of systematic medical training [5].

Manuscript received: January, 22, 2023; Revised April, 9, 2023; Accepted May, 9, 2023. This paper was recommended for publication by Editor Jessica Burgner-Kahrs upon evaluation of the Associate Editor and Reviewers' comments. (Corresponding author: Zhongliang Jiang.)

¹Z. Jiang, X. Li, C. Zhang, Y. Bi and N. Navab are with the Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Germany. (z.l.jiang@tum.de)

²W. Stechele is with the Chair of Integrated Systems, Technical University of Munich, 80333 Munchen, Germany

The initial implementation was performed by C. Zhang in his thesis. Digital Object Identifier (DOI): see top of this page.

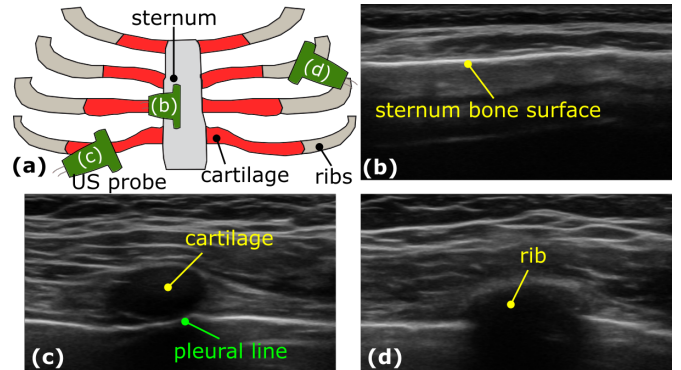


Fig. 1. Illustration of the anatomical feature of ribs on US images. (a) three types of thorax bones: sternum, rib and costal cartilage. (b), (c) and (d) are the representative B-mode images of sternum, cartilage and rib, respectively. Due to the low acoustic impedance of cartilage, subcutaneous pleural line can be found right under the cartilage from US images.

To address the shortage of experienced sonographers, in particular in rural areas, the robotic technique is employed to develop fully or semi-automatic US screening programs. Due to the superiority in terms of accuracy and repeatability, the robotic US system (RUSS) is considered as a promising solution to overcome the limitation of intra- and inter-operator variations. Pierrot *et al.* developed a robotic system to assist clinicians in moving a probe on patients' skin [6]. They presented in-vivo results for reconstructive surgery to demonstrate the feasibility of applying such a system in real scenarios. Gilbertson *et al.* employed compliant control strategies to maintain a constant contact force between the probe and the object [4]. To accurately control the acquisition parameters, Huang *et al.* employed a depth camera to compute the normal direction of the target surface [7]. Besides, Jiang *et al.* developed a mechanical model-based method to accurately position the probe in the normal direction of unknown constraint surfaces based on contact force [8] and real-time US images [9]. The camera-based approach has high time efficiency, while the force-based approach can achieve high accuracy. Therefore, the former is often used for real-time applications and the latter is more suitable for applications requiring high precision, such as CT-US registration for orthopedics applications [10].

To further enable autonomous scans, Huang *et al.* extract the abdominal area from RGB images based on the rule "Red>Green>Blue" and then automatically compute the multiple-line scanning trajectory to fully cover the extracted area [7]. Jiang *et al.* calculated the desired scanning direction of the target tubular structure in real-time based on B-mode images [11]. Yet, the proposed optimal control is

only validated for tubular structures. Considering the restricted acoustic windows for the organs covered by ribs, e.g., the liver and heart, Göbl *et al.* computed scan poses to maintain the coverage of the target on tomographic images and optimize the image quality simultaneously [12]. To transfer the planned trajectory on pre-operative images to the current setup, Hennesperger *et al.* proposed a skin surface-based registration approach to optimize the transformation between pre-operative images and the current setup captured using an RGB-D camera [13]. Recently, Jiang *et al.* extended this work by incorporating the non-rigid registration approach to bridge differences between pre-operative MR and individual patients, such as articulated motions in arm scanning [14]. Both of these approaches are proved to be robust for their specific applications (aorta and limb artery tree). However, since they only use the skin surface extracted from an external RGB-D camera without subcutaneous anatomical features, their use in thorax application is limited, such as the intercostal liver ablation [15]. For such a task, the probe should be accurately positioned between two ribs to guarantee the visibility of the target by avoiding acoustic shadows caused by bone structures (see Fig. 1). Without explicitly considering subcutaneous bone structures with high acoustic impedance, the shadows will hinder the visibility of the internal target of interest.

To address this challenge, this work presents a skeleton graph-based non-rigid registration between the cartilage point clouds extracted from a tomographic template and US images of patients. Compared to the existing studies using skin surface point clouds [13], [14], subcutaneous bone can better characterize the scene of thorax examinations, which often require accurate positioning of a probe in limited acoustic windows, i.e., between two ribs. Therefore, the proposed skeleton graph-based registration can more appropriately transfer a planned scanning path to the current setup for robotic scanning. Since there is no camera involved in our setup, the method does not require a hand-eye calibration process, which, as a side benefit, could bring more convenience in clinical practice. The main contributions are summarized as follows:

- A skeleton graph-based non-rigid registration is developed to enable the accurate transferring of scanning trajectories from tomographic templates to the current setup based on the invariant subcutaneous bone surfaces. This is particularly useful for thorax applications with limited acoustic windows.
- Automatically generating skeleton graphs from rib cartilage point clouds by successively using the self-organization map and minimal spanning tree algorithm; then, further optimizing the computed skeletons to match the rib cartilage outline by considering the anatomy's continuity.
- The visible cartilage anatomical features on both US and CT images are used as biometric landmarks to enable the registration between a generic CT template and tracked US images characterizing patient-specific properties.

Finally, the proposed graph-based registration is validated on the US data recorded from a volunteer and seven random CT images from a public dataset. The results demonstrate

that the proposed method outperforms the classical iterative closest point (ICP) algorithm in terms of adapting to inter-patient variations (distance error mean \pm SD: 5.0 ± 1.9 mm vs 8.6 ± 6.7 mm on seven CTs). More intuitive results can be found in the online video¹.

II. RELATED WORK

Due to the invariance of bone structures on images, it is often employed as the subcutaneous feature for the registration between pre-operative tomographic images (CT or MR) and inter-operative US images, particularly in orthopedic surgery. The existing approaches can be categorized into two groups: image-based approach [16] and bone surface-based approach [10], [17]. The image-based approaches often directly register US images to the preoperative tomographic images based on the similarity metrics such as LC² [16]. The 2D-3D registration [18], [19] has been an attractive direction recently, while it is an ill-posed problem and usually requires rich anatomical features to ensure registration accuracy, like brain images. The image-based approaches have the benefits that there is no need to do bone segmentation, while the result often suffers from the imaging noise [20].

In contrast, surface-based approaches [10], [17] require segmentation results of the subcutaneous anatomy. To segment bone surfaces from US images, various approaches have been proposed in the existing studies [10], [21]–[23]. Recently, Li *et al.* generated synthesized realistic US images from CT scans to facilitate the training of segmentation network [24]. Based on the segmentation results, 3D point clouds can be created by stacking the bone surface with proper tracking information. Based on the point clouds obtained from both modalities, the classical ICP algorithm [25] can be used to optimize the transformation matrix between two point clouds. Considering the point clouds may only be partially observed, Zhang *et al.* incorporate the partially reliable normal vectors to form the registration problem as a maximum likelihood estimation problem [26]. The experiments on a femur head showed that the method could robustly and accurately optimize the rigid matrix. But it is also worth noting that the performance may decrease if the target anatomy has limited geometry changes.

To consider the changes and deformations of targets, Chebrolu *et al.* segmented the plant leaves and stem, respectively, using a support vector machine and then developed a spatio-temporal non-rigid registration approach [27], [28]. In computer vision, the non-rigid ICP [29] and the Coherent Point Drift (CPD) algorithm [30] were proposed for non-rigid point set registration. Such methods often assume that the two inputs are intact and share the same region of interest (ROI), e.g., intact liver or kidney. Specific to the thoracic application with limited acoustic windows, it is hard to guarantee the selected rib ROIs from different patients are consistent with the templates because thoracic ribs cannot be completely scanned only from the front side without moving patients. Therefore, we only count the cartilage part to ensure consistent ROIs for different patients instead of the intact thoracic cage for non-rigid registration in this study. The cartilage's anatomical

¹Video: <https://www.youtube.com/watch?v=LkSHL7FJ8eU>

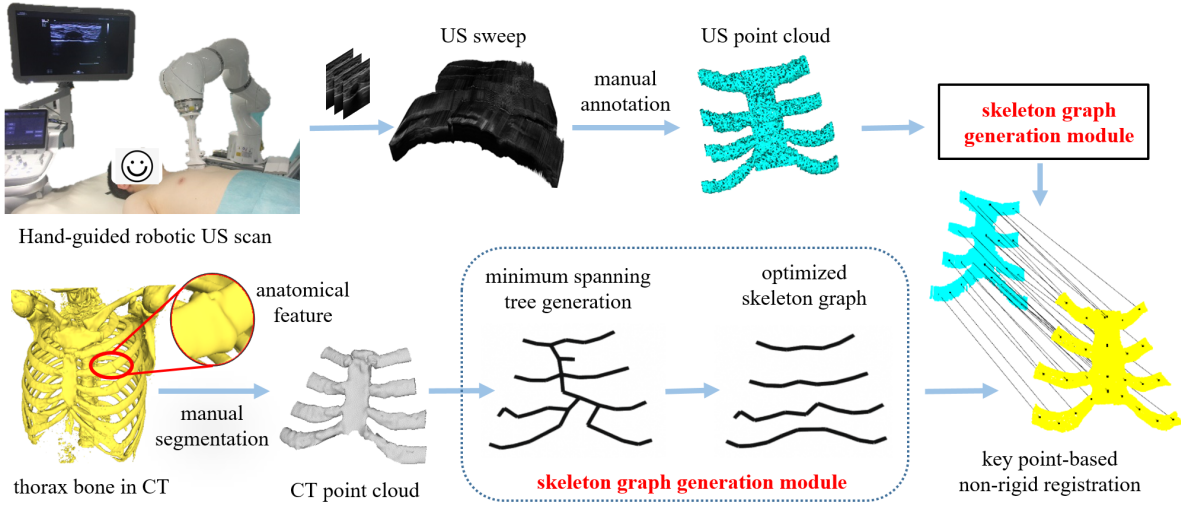


Fig. 2. The workflow of the proposed skeleton graph-based non-rigid registration framework.

features on both US and CT images are used as biometric landmarks to enable the registration of the same ROIs from CT templates and individual patients' US scans.

III. METHODS

The general aim of this work is to develop a skeleton graph-based non-rigid registration approach for autonomous robotic stenography based on the rib cartilage point clouds. The use of subcutaneous bone features enables appropriate transferring of US scanning trajectory from pre-operative tomographic images to the current setup; therefore, allowing autonomous scanning or navigation for thoracic applications with limited acoustic windows, such as liver ablation between ribs [15]. The overall workflow of the proposed framework is summarized in Fig. 2. We first manually annotate the rib cartilage in both US images and tomographic images based on its visible anatomical characteristic. Then, to appropriately characterize the rib cartilage outline, the self-organizing mapping and minimal spanning tree are successively implemented. Finally, the graph-based non-rigid registration is carried out by matching the key points on the extracted skeleton graph. The details of individual parts are described in the following sections.

A. Robotic US Calibration

To map US images' pixels into 3D space, spatial US calibration is needed. Besides, it is also important to have an additional temporal calibration between the US stream and robotic tracking data to properly stack the 2D images in 3D space, in particular for the sweep with multiple scan lines.

1) *Spatial Calibration*: Spatial calibration refers to the transformations between different coordinate frames. In this study, the involved frames are defined as follows: robotic base $\{b\}$, robotic end-effector $\{ee\}$, tool center point (TCP) $\{tcp\}$ and US image frame $\{us\}$. Similar to [11], the pixel scale factor is first computed based on the probe footprint length and image depth. Then the transformation from $\{ee\}$ to $\{b\}$ can be obtained based on the given kinematic model from the manufacturer. The final matrix ${}^b_{us}\mathbf{T}$ mapping the pixel position

from US image $\{us\}$ into the robotic base frame $\{b\}$ can be calculated as follows.

$${}^b_{us}\mathbf{T} = {}^b_{ee}\mathbf{T} {}^{ee}_{tcp}\mathbf{T} {}^{tcp}_{us}\mathbf{T} \quad (1)$$

where ${}^i_j\mathbf{T}$ is the matrix transferring position representations from frame $\{i\}$ to frame $\{j\}$.

2) *Temporal Calibration*: To obtain accurate 3D geometry of the scanned volume, temporal calibration is important for synchronizing the US and robotic tracking streams. Inspired by the temporal synchronization process proposed in [31], we repetitively move the robotic upward and downward on a gel phantom with a 3D-printed landmark inside. Then, an intensity threshold-based method is used to extract the landmark location from the recorded B-mode images. Since the image and tracking data are recorded at the same time, the repeated pixel location changes should have the same change tendency as the robotic position, e.g., the motion in the probe centerline direction. Thereby, the temporal difference can be computed by maximally matching the two curves with respect to time stamps. The temporal synchronization shift between the US and robotic tracking streams is 268 *ms* in our setup.

B. Cartilage Point Cloud Generation

The selection of consistent ROI is one of the practical problems that limit the application of non-rigid registration. Due to inter-patient variations, it is hard to accurately select the same ROIs from different patients in different imaging modalities. Inspired by the finding that the cartilage US images are different from ribs on B-mode images [32], where we can also observe pleural boundaries under rib cartilage (see Fig. 1), we can select common cartilage parts from different patients. Besides, there are also apparent anatomical features existing at the connection part between cartilage and ribs (see Fig. 2) in CT images. Therefore, we can get complete cartilage both in tomographic images and US images. Based on this, a non-rigid registration developed based on cartilage bones becomes feasible and such skeleton graph-based registration can enable

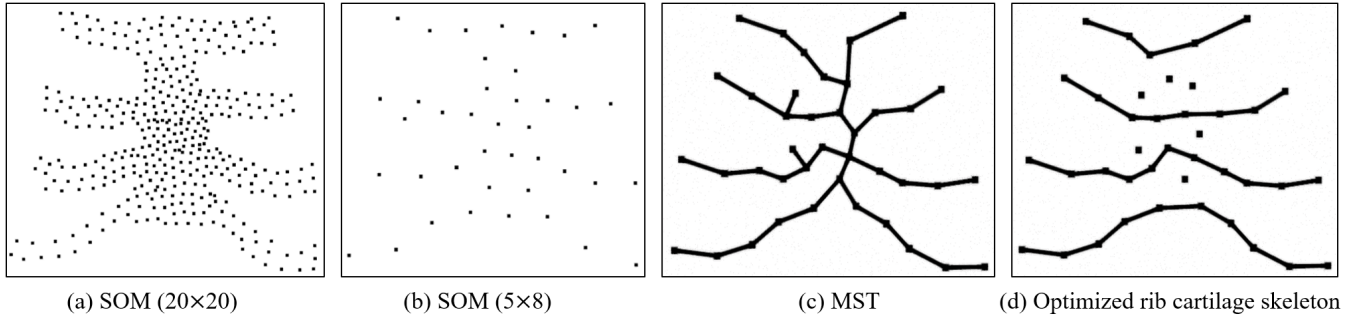


Fig. 3. Rib cartilage skeleton graph extraction and optimization from the US point cloud \mathbf{P}_{us} . (a) the key points \mathbf{P}_{us}^{som} extracted from original \mathbf{P}_{us} using a SOM with a 20×20 grid map. (b) the resulting key vertex points \mathbf{V}_p^{us} obtained from \mathbf{P}_{us}^{som} after applying a SOM with a 5×8 grid map. (c) the MST G_{tr} computed based on the key points extracted from \mathbf{V}_p^{us} . (d) the optimized four paths representing four rib cartilage.

transferring of a planned path from pre-operative images to the current setup of various patients.

1) *Point Clouds Generation from CT Template*: Seven chest templates are selected from a public human CT dataset². To extract the ribs from CT data, we first perform an intensity-based segmentation in an open-source software 3D Slicer³. A representative result of the chest mesh is depicted in Fig. 2 (bottom left part). Based on the anatomical features, we manually segment the cartilage sections from the 2-th, 3-th, 4-th and 5-th ribs in Meshlab⁴. To create uniformly distributed cartilage point clouds \mathbf{P}_{ct} from CT, the Poisson disc sampling algorithm is employed instead of the random sampling method.

2) *Point Clouds Generation from US Sweep*: To generate the subcutaneous bone surface point clouds, we first ask the volunteers to lie down on a flat operation bed as Fig. 2 (top left part). Then, a human operator manually defines a multiple-line US scanning to cover the upper part of the chest. The US acquisition took 106 s on average when the volunteer was under natural breathing. Based on the US calibration results, the 3D visualization of the recorded sweep can be seen in Fig 2. To further extract the cartilages point cloud \mathbf{P}_{us} , we first label the cartilage bone surface by manually annotating the bone surface on 2D B-mode images using ImFusionSuit (ImFusion AG, Germany). Afterward, the original B-mode images are replaced with the manually annotated binary masks of cartilage bone to form the 3D object and further compounded using a linear interpolation using PLUS [33]. Then, the same processes described in the previous subsection are performed in Meshlab to generate US cartilage point cloud (including sternum) \mathbf{P}_{us} (see Fig. 2).

C. Skeleton Graph Generation and Optimization

1) *Self-Organizing Map for Data Cleaning and Key Point Extraction*: The self-organizing map (SOM) algorithm [34] is proposed by Kohonen to represent high-dimensional features using a lower dimensional representation while maintaining the topological structure. It was used to extract plant's key points in [28]. Inspired by this work, we apply SOM to extract the key points from CT and/or US rib cartilage point clouds.

The SOM is an unsupervised machine learning technique trained using competitive learning, which ensures it can be used on unseen data. To extract the point cloud distributed uniformly in Euclidean space using the SOM algorithm, a grid map G_g with size of $m_g \times n_g$ is initialized. Each node of G_g contained a corresponding weight vector \mathbf{W}_s with the same dimension as input data \mathbf{P} . To update the weight vectors \mathbf{W}_s , the Euclidean distance between a random sample of the training input and all nodes of G_g are computed. The weight vector with the smallest value is called the best matching unit (BMU). The node close to the BMU will allow larger adjustments than the ones far away from BMU. Afterward, \mathbf{W}_s is updated as follows:

$$\mathbf{W}_s(i+1) = \mathbf{W}_s(i) + \theta_{(BMU,i)} \cdot l_r \cdot [\mathbf{P}(k) - \mathbf{W}_s(i)] \quad (2)$$

where i is the current iteration, $\theta_{(BMU,i)}$ is the updated restriction functions computed based on the grid distance between other nodes and the BMU, l_r is the learning rate, and $\mathbf{P}(k)$ is the k -th data in the input source point cloud.

In this study, the SOM is successively applied twice. The first SOM is used to get a uniformly distributed point cloud with a smaller size than the original input point cloud. Furthermore, this process can create point clouds (from CT, US, or different patients) including a unified number of points. Therefore, a unified input has the potential to benefit the generalization capability of the proposed graph-based registration against inter-patient variation. To preserve the character of the original skeleton point cloud, a relatively large number of map nodes ($400 = 20 \times 20$) are used to unify the size of input point clouds. The second SOM is used to automatically extract the key skeleton points to represent the rib cartilage point cloud. Since the point clouds are quasi-symmetrical about the sternum, the second map size is set around $m_g \times 2m_g$. Since the point cloud includes four cartilage, m_g is set to 4. By testing a few paired m_g and n_g around (4×8) , we empirically use 5×8 for the initial map of the second SOM based on the extracted cartilage skeleton profiles. The representative results of the two SOM results on a US rib cartilage point cloud are depicted in Fig. 3 (a) and (b), respectively.

2) *Minimum Spanning Tree Generation*: After obtaining a set of skeleton key points \mathbf{V}_p [Fig. 3 (b)], the minimum spanning tree (MST) is used to connect the skeleton key points

²CT dataset: <https://github.com/M3DV/RibSeg>

³3D Slicer: <https://www.slicer.org/>

⁴Meshlab: <https://www.meshlab.net/>

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

to form a representative tree characterizing the feature of rib cages. The MST is a graphical generation approach to connect \mathbf{V}_p by edges. The spanning tree means a connected graph including all \mathbf{V}_p , but does not have any closed loops. Here, Euclidean distance is employed as the edge weights for two arbitrary points. The weight matrix \mathbf{W} is defined as follows:

$$\mathbf{W}_{i,j} = \|\mathbf{V}_p(i) - \mathbf{V}_p(j)\| \quad (3)$$

where i and j are the i -th and j -th point in \mathbf{V}_p , respectively.

In order to find the tree graph with a minimum total weight for an undirected graph, Prim's algorithm [35] is used. This greedy algorithm builds MST G_{tr} step by step from an arbitrary point. In each step, G_{tr} is augmented by adding the cheapest possible edge connection from the tentative tree G_{tr} to the remaining points not yet in G_{tr} . A representative MST for the collected cartilage point cloud \mathbf{P}_{us} is depicted in Fig. 3 (c).

3) *Skeleton Tree Optimization*: It can be seen from Fig. 3 (c) that the generated MST G_{tr} connected all \mathbf{V}_p . Besides, the path connecting two arbitrary nodes in G_{tr} is unique. Since there are four ribs (2-th, 3-th, 4-th and 5-th) involved in our setup, we can obtain four representative and unique paths to represent the four ribs. To extract such paths, two endpoints of each rib need to be identified. To this end, the gift wrapping algorithm [36] is employed to extract the convex hull for the G_{tr} of cartilage point cloud \mathbf{P}_{us} and \mathbf{P}_{ct} . The eight endpoints $\mathbf{P}_{end} \in R^{8 \times 1 \times 3}$ of the four ribs are the key points forming the convex hull. To further understand the correspondence between the identified eight points in US data, we match \mathbf{P}_{end}^{us} to the \mathbf{P}_{end}^{ct} using the ICP algorithm. According to the ribs order information on CT template, \mathbf{P}_{end} can be grouped into the pairwise endpoints set $\mathbf{P}_{pair} \in R^{4 \times 2 \times 3}$.

Based on the ordered four pairs of endpoints \mathbf{P}_{pair} , four unique paths can be obtained from the MST tree by finding the shortest tree distance. Due to the fact that the human ribs are continuous, the direction variations between the vectors connecting two neighboring key points should be small. Therefore, four rib skeleton representations \mathbf{P}_{rs} are optimized as follows:

$$\mathbf{T}_{rs}^i = \begin{cases} [\mathbf{T}_{rs}^i, \mathbf{P}_{rs}^i(j)] & \text{if } \angle \theta \leq T_\theta \\ \mathbf{T}_{rs}^i & \text{otherwise} \end{cases} \quad (4)$$

where $i = \{2, 3, 4, 5\}$ represents the i -th rib cartilage skeleton, \mathbf{T}_{rs}^i is the optimized rib skeleton, which is initialized as $\mathbf{T}_{rs}^i(1) = \mathbf{P}_{rs}^i(1)$, $j = 3, \dots, N_i$ is the j -th point of i -th rib skeleton, $\angle \theta = \angle \left(\overrightarrow{\mathbf{T}_{rs}^i(j-2)\mathbf{T}_{rs}^i(j-1)}, \overrightarrow{\mathbf{T}_{rs}^i(j-1)\mathbf{P}_{rs}^i(j)} \right)$, T_θ is the angular threshold filter out unsuitable points. T_θ is empirically set to 60° based on the experimental results. The optimized \mathbf{T}_{rs} are visualized in Fig. 3 (d).

D. Skeleton Graph-based Non-rigid Registration

Based on the extracted skeleton in Fig. 3 (d), the four rib cartilage are characterized. To perform the graph-based registration between the US and CT cartilage point clouds (\mathbf{P}_{us} and \mathbf{P}_{ct}), the skeleton key points are fitted using a cubic polynomial function and then resampled uniformly for each

rib. Considering the variation in the length, different sampling points are used for different ribs, while the same number of key points will be used for the same rib (e.g., the second rib) in CT and US. The resampled points on the skeleton are computed as follows:

$$[X^i, Y^i, Z^i] = [X^i, f_y^i(X^i), f_z^i(X^i)] \quad (5)$$

where f_y^i and f_z^i are the fitted cubic polynomial functions with respect to X in Y and Z directions, respectively. $i = \{2, 3, 4, 5\}$ is the level of ribs, $X^i \in R^{N_i \times 1}$ is the N_i values uniformly sampled in X direction with equal interval. Based on the length of each rib cartilage, N_2, N_3, N_4 , and N_5 are empirically set 6, 6, 8 and 10, respectively.

After obtaining the re-sampled skeleton points, the graph-based non-rigid registration can be implemented by computing local transformations for individual points in point clouds. To this end, the closest N_r re-sampled skeleton key points to the specific point in \mathbf{P}_{ct} are automatically selected based on Euclidean distance. Then, a local transformation matrix ${}^{us}_{ct}\mathbf{T}$ mapping \mathbf{P}_{ct} to \mathbf{P}_{us} can be optimized by minimizing Eq. (6).

$$\min_{{}^{us}_{ct}\mathbf{T}} \frac{1}{N_r} \sum_{i=1}^{N_r} \|{}^{us}_{ct}\mathbf{T} P_i^{ct} - {}^{us}P_i\|^2 \quad (6)$$

where ${}^{us}P_i$ and ${}^{ct}P_i$ are the paired re-sampled skeleton key points of US and CT point clouds, respectively. The hyperparameter $N_r = 10$ is empirically determined based on the experimental results by making a compromise between anatomy continuity and local accuracy. A large N_r will reduce the non-rigid probability. An extreme case is that if we use all key points to calculate the transformation matrix, the method will be degraded to a rigid registration.

IV. RESULTS

A. Robotic US System

The presented RUSS consists of a robotic manipulator (LBR iiwa 7 R800, KUKA GmbH, Germany) and an ACUSON Juniper US System (Siemens Healthineers, Germany). A linear probe (12L3, Siemens, Germany) is attached to the robot flange by a 3D printed holder (see Fig. 2). The robot status is updated at 100 Hz. The US images are captured by a frame grabber (Epiphan video, Canada) in 30 fps. To properly visualize the bone structure in B-mode images, a default setting provided by the manufacturer is used in this study: MI: 1.13, TIS: 0.2 TIB: 0.2 DB: 60 dB. Since the ribs of interest are shallow, the imaging depth was set to 35 mm.

B. Performance of the Skeleton Graph-based Registration

To quantitatively evaluate the proposed skeleton graph-based non-rigid registration, seven human chest CT images are selected from a public dataset. Using multiple images is helpful in drawing the correct conclusion by ruling out occasional cases. Following the steps described in Sec. III-B, seven CT point clouds can be obtained to evaluate the registration performance in different cases. The US point cloud is obtained from a healthy volunteer (gender: male, year: 26, BMI: 19.8, height: 174 cm).

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

To further compare the performance of the proposed approach to existing ones, i.e., standard ICP, non-rigid ICP, and CPD algorithm, the computed registration errors on two representative CTs are depicted in Fig. 4. Since the ICP and non-rigid ICP are sensitive to the initial alignment, we manually adjust the source and target point clouds to be close and also align them in the right orientation in order to provide a fair comparison. The proposed skeleton graph-based approach does not require dedicated initialization. The final results are computed in terms of Euclidean distance, which is calculated between a point in the realigned \mathbf{P}_{ct} (after registration) and the closest point in \mathbf{P}_{us} . It can be seen from Fig. 4 that the proposed method can significantly outperform standard ICP and CPD in terms of distance error. The mean ($\pm SD$) error of standard ICP, CPD, and the proposed method are $7.9 \pm 4.8 \text{ mm}$, $7.5 \pm 6.0 \text{ mm}$, and $4.9 \pm 1.8 \text{ mm}$ on CT 1 and $7.7 \pm 4.9 \text{ mm}$, $6.9 \pm 6.0 \text{ mm}$, and $5.0 \pm 1.4 \text{ mm}$ on CT 2. To perform the t-test for statistical comparison, the Lilliefors test is employed to confirm that the estimated errors are normally distributed. The p -value for the results computed using the proposed method and the standard ICP and CPD are 3.4×10^{-8} , and 6.0×10^{-5} on CT 1, and 5.1×10^{-7} and 1.1×10^{-3} on CT 2, respectively. The rigid ICP algorithm performs worst because the inter-patient variations are not taken into consideration. Based on the data distributions, the results computed using CPD are more concentrated than the standard ICP. Additionally, the non-rigid ICP and the proposed method can achieve similar performance (CT 1: $5.4 \pm 3.2 \text{ mm}$ vs $4.9 \pm 1.8 \text{ mm}$, and CT 2: $4.4 \pm 2.2 \text{ mm}$ vs $5.0 \pm 1.4 \text{ mm}$). However, it is worth noting that the proposed method can limit the maximum error (10.2 mm vs 16.4 mm on CT 1 and 9.3 mm vs 13.5 mm on CT 2), and its performance is invariant to the initial alignments.

The Euclidean distance (ED) between two point sets can not prevent the self-deformation for non-rigid ICP. To ensure the original shape can be preserved after non-rigid registration, we compute the Hausdorff distance (HD) between the transferred CT and the US point clouds. The HD is often used to describe how far two unpaired subsets are, which is sensitive to the boundary shape. The statistical results in terms of ED and HD using seven independent CT data are summarized in TABLE I. It can be seen that both ED and HD errors achieved by the proposed method are smaller and more stable than others over different CTs. The small ED error and large HD results achieved by non-rigid ICP on CT 1, 2, and 4 indicate self-deformation. After applying the variable transformation matrix to individual CT points, multiple CT points are mapped to the same US points. The maximum HD computed by the proposed method is 13.7 mm , whereas the one computed by the non-rigid ICP is 69.4 mm . Therefore, we consider the proposed graph-based non-rigid registration is more effective and robust in adapting to inter-patient variations.

C. Validation of Transferring Trajectory from CT to US

In this section, we further validate whether the proposed graph-based registration can appropriately map a planned trajectory from generic template CT to the current setup. To

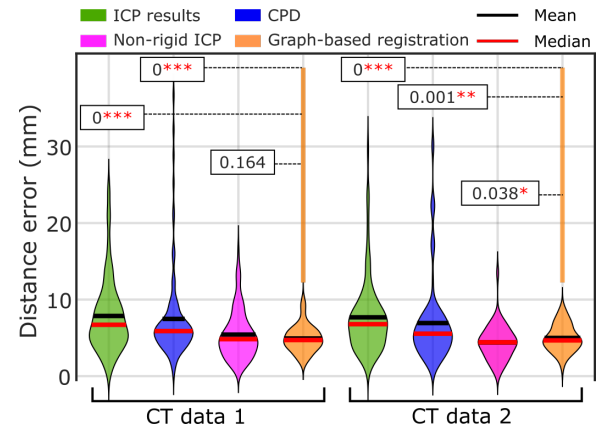


Fig. 4. The quantitative results of the existing standard ICP, CPD and non-rigid ICP algorithm and the proposed skeleton graph-based non-rigid registration method. The Euclidean distance is used as the validation metric. The p -values are computed between the results computed using the proposed and existing methods. * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$) mean there are significant differences.

this end, we manually determined ten scanning waypoints in both \mathbf{P}_{us} and \mathbf{P}_{ct} . The positions of the ten waypoints are defined by computing the average of the re-sampled key points [Eq. (5)] in two neighboring ribs. It can be seen from Fig. 5 (a), that the ten waypoints are located in the middle of two ribs. To avoid biased results, the ten points are defined in all three levels of the intercostal gaps, which are distributed on both sides. These ten points were only used for evaluating the trajectory transferring performance.

A representative result is intuitively described in Fig. 5 (a), where the red points are the transferred points from the CT to the US point cloud obtained from a volunteer. The transferred points are roughly distributed in the middle of ribs in US point cloud as expected. To further provide quantitative analysis on trajectory transferring performance, the statistical errors are summarized in Fig. 5 (b). The position error is represented by the distance difference between the transferred points from CT and the points defined on US point clouds directly.

The trajectory transferring results obtained by different methods are consistent between the two representative CT images. Compared to existing methods, the errors obtained by the proposed method are the smallest in both cases (mean \pm SD) $3.4 \pm 1.9 \text{ mm}$ and $3.1 \pm 1.5 \text{ mm}$ on CT 1 and 2. The results obtained by the standard ICP, non-rigid ICP and CPD are $5.6 \pm 3.0 \text{ mm}$, $5.0 \pm 2.5 \text{ mm}$, and $4.9 \pm 2.5 \text{ mm}$ on CT 1 and $5.2 \pm 3.0 \text{ mm}$, $4.9 \pm 2.7 \text{ mm}$, and $5.0 \pm 3.1 \text{ mm}$ on CT 2. It is also noteworthy that the standard ICP performs worst in terms of positional error in both cases. The two non-rigid methods achieve similar accuracy. Considering the gap between two ribs centerlines varies from 24.2 mm to 31.6 mm in our setup, which is more than six times larger than the positional error. Since the average intercostal space is around 15 mm [37] and the footprint's width of the used probe is around 10 mm , the tolerance of positioning error will be around 5 mm . Therefore, we consider the proposed graph-based approach can meet the clinical requirement and can be used to remap the planned scan trajectory to current setups for autonomous robotic scanning.

TABLE I
REGISTRATION RESULTS OF DIFFERENT APPROACHES (MEAN)

Methods	CT 1		CT 2		CT 3		CT 4		CT 5		CT 6		CT 7	
	ED	HD	ED	HD	ED	HD	ED	HD	ED	HD	ED	HD	ED	HD
ICP [25] ♣	7.9	59.3	7.7	32.5	7.1	31.2	6.2	15.1	16.3	68.8	9.8	43.0	5.3	21.6
CPD [30]	7.5	36.7	6.8	30.2	6.7	35.8	8.3	38.5	6.0	25.2	7.9	46.1	9.5	56.7
Nonrigid ICP [29] ♣	5.4	69.4	4.4	34.1	3.9	10.3	4.1	29.9	3.7	15.8	3.7	19.7	4.1	16.2
Our method	4.9	11.7	5.0	12.8	4.9	10.7	5.7	13.7	4.9	13.2	4.9	10.5	4.7	9.8

ED: Euclidean distance; HD: Hausdorff distance; ♣: sensitive to initial alignments; Unit: mm

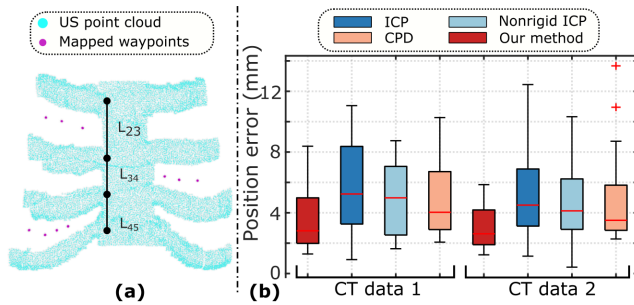


Fig. 5. Results of trajectory transferring from CT to US spaces. (a) Illustration of the results of 10 waypoints mapped from CT to US. (b) The statistical results of the mapping performance in terms of position errors obtained by various methods. The length of the neighbouring ribs are measured as $L_{23} = 31.6$ mm, $L_{34} = 26.1$ mm and $L_{45} = 24.2$ mm.

V. DISCUSSION

The proposed skeleton graph-based non-rigid registration has been validated on the cartilage point clouds obtained from a volunteer and seven CT images. The results demonstrate that the proposed approach can achieve better performance than the standard ICP, CPD [29] and nonrigid ICP [30] algorithms in adapting to inter-patient variations. However, there are still some limitations that need to be discussed to further inspire future research. First, the physiological motion, i.e., respiration, has not been explicitly considered, which will affect the accurate representation of the bone surface in 3D. To alleviate this issue, future work can try to estimate the respiration-induced motion using external tracking [38] or real-time B-model images [39]. Second, the skin surface hasn't been included in the registration approach, which can result in undesired contact conditions (too deep or with a gap) between the probe and the patient's skin. However, this negative influence could be reduced by using a compliant control [1], [4], [8] to maintain a contact force during scans. Besides, due to the physiological variations between males and females, the breast will result in significant deformation during the acquisition of US images for females. Such inevitable deformation will bring challenges in both acquiring accurate point clouds of the subcutaneous bone surface and appropriately pointing a probe to visualize the internal objects.

After validating the feasibility of explicitly using cartilage bone for trajectory mapping, future studies need to reduce the manual component in the current pipeline to improve the

level of autonomy. To autonomously plan a scanning path for covering the chest part, an external depth camera can be used to extract the object surface based on predefined rules [7], [14] or learning-based segmentation methods [40]. To ensure the coverage of the whole ROI, Tan *et al.* employed multiple cameras and fused the surface point clouds obtained from different views [41]. In addition, the US cartilage surface point was obtained by manual annotation based on the anatomical feature depicted in Fig. 1. Future studies need to further integrate advanced bone surface extraction methods [10], [21]–[23], [31] to improve the autonomy level. Furthermore, considering practical issues in real scenarios, potential tissue motions [40], [42] and deformations [43] should be properly compensated for to ensure acquisition robustness and anatomy accuracy.

VI. CONCLUSION

This work presents a skeleton graph-based non-rigid registration approach between tomographic templates and B-mode images of patients for autonomous transferring of US scan trajectory from CT to the current setup. Compared to the existing studies using skin surface point clouds [13], [14], subcutaneous bone surfaces are used to better characterize challenging scenes of thorax application with limited acoustic windows; therefore achieving bone-aware registration results. The proposed method has been validated on the US data recorded from a volunteer and seven CT volumes. The registration results shown in Fig. 4 and TABLE I demonstrate that the proposed method outperforms the standard ICP and CPD algorithms in terms of adapting to inter-patient variations. Although the non-rigid ICP method obtains similar ED results, the proposed method can achieve significantly lower HD results than non-rigid ICP. This indicates that the skeleton graph-based method can better preserve the topology structure of the anatomies, which can improve the robustness and time efficiency of the graph-based registration by leveraging prior skeleton knowledge. By explicitly considering subcutaneous bone structures, we believe this study can aid in the development of autonomous RUSS scanning or navigation for challenging thoracic applications with limited acoustic windows, e.g., liver ablation through intercostal space.

REFERENCES

- [1] M. K. Welleweerd, A. G. de Groot, V. Groenhuis, F. J. Siepel, and S. Stramigioli, "Out-of-plane corrections for autonomous robotic breast ultrasound acquisitions," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 515–12 521.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

- [2] Z. Jiang, F. Duellmer, and N. Navab, "Dopus-net: Quality-aware robotic ultrasound imaging based on doppler signal," *IEEE Transactions on Automation Science and Engineering*, pp. 1–1, 2023.
- [3] M. Napoli, G. Bacchini, L. Scarpati, G. Loizzo, and A. Zito, "Ultrasound guided interventional procedures on arteriovenous fistulae," *The Journal of Vascular Access*, vol. 22, no. 1_suppl, pp. 91–96, 2021.
- [4] M. W. Gilbertson and B. W. Anthony, "Force and position control system for freehand ultrasound," *IEEE Trans. Robot.*, vol. 31, no. 4, pp. 835–849, 2015.
- [5] M. A. Maraci, R. Napolitano, A. Papageorghiou, and J. A. Noble, "Searching for structures of interest in an ultrasound video sequence," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2014, pp. 133–140.
- [6] F. Pierrot, E. Dombre, E. Dégouange, L. Urbain, P. Caron, S. Boudet, J. Gariépy, and J.-L. Mégnien, "Hippocrate: A safe robot arm for medical applications with force feedback," *Med. Image Anal.*, vol. 3, no. 3, pp. 285–300, 1999.
- [7] Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Trans. Ind. Inform.*, vol. 15, no. 2, pp. 1173–1182, 2018.
- [8] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, "Automatic force-based probe positioning for precise robotic ultrasound acquisition," *IEEE Transactions on Industrial Electronics*, 2020.
- [9] Z. Jiang, M. Grimm, M. Zhou, J. Esteban, W. Simson, G. Zahnd, and N. Navab, "Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1342–1349, 2020.
- [10] W. Wein, A. Karamalis, A. Baumgartner, and N. Navab, "Automatic bone detection and soft tissue aware ultrasound-ct registration for computer-aided orthopedic surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 971–979, 2015.
- [11] Z. Jiang, Z. Li, M. Grimm, M. Zhou, M. Esposito, W. Wein, W. Stechele, T. Wendler, and N. Navab, "Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7064–7075, 2021.
- [12] R. Göbl, S. Virga, J. Rackerseder, B. Frisch, N. Navab, and C. Hennersperger, "Acoustic window planning for ultrasound acquisition," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 6, pp. 993–1001, 2017.
- [13] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, and N. Navab, "Towards MRI-based autonomous robotic us acquisitions: a first feasibility study," *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 538–548, 2016.
- [14] Z. Jiang, Y. Gao, L. Xie, and N. Navab, "Towards autonomous atlas-based ultrasound acquisitions in presence of articulated motion," *IEEE Robotics and Automation Letters*, 2022.
- [15] B. Quesson, M. Merle, M. O. Köhler, C. Mougenot, S. Roujol, B. D. De Senneville, and C. T. Moonen, "A method for mri guidance of intercostal high intensity focused ultrasound ablation in the liver," *Medical physics*, vol. 37, no. 6Part1, pp. 2533–2540, 2010.
- [16] W. Wein, A. Khamene, D.-A. Clevert, O. Kutter, and N. Navab, "Simulation and fully automatic multimodal registration of medical ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2007, pp. 136–143.
- [17] M. Ciganovic, F. Ozdemir, F. Pean, P. Fuernstahl, C. Tanner, and O. Goksel, "Registration of 3d freehand ultrasound to a bone model for orthopedic procedures of the forearm," *International journal of computer assisted radiology and surgery*, vol. 13, no. 6, pp. 827–836, 2018.
- [18] B. Fuerst, W. Wein, M. Müller, and N. Navab, "Automatic ultrasound-mri registration for neurosurgery using the 2d and 3d lc2 metric," *Medical image analysis*, vol. 18, no. 8, pp. 1312–1319, 2014.
- [19] M. Zheng, A. Mohamadi, T. Szabo, and B. Snyder, "Evaluation of in-vivo kinematics of cervical spines by co-registering dynamic ultrasound with mri," in *2017 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2017, pp. 1–4.
- [20] I. Hacıhaliloğlu, A. Rasoulouian, R. N. Rohling, and P. Abolmaesumi, "Local phase tensor features for 3-d ultrasound to statistical shape+pose spine model registration," *IEEE transactions on Medical Imaging*, vol. 33, no. 11, pp. 2167–2179, 2014.
- [21] I. Hacıhaliloğlu, R. Abugharbieh, A. J. Hodgson, and R. N. Rohling, "Bone surface localization in ultrasound using image phase-based features," *Ultrasound in medicine & biology*, vol. 35, no. 9, pp. 1475–1487, 2009.
- [22] P. Wang, M. Vives, V. M. Patel, and I. Hacıhaliloğlu, "Robust real-time bone surfaces segmentation from ultrasound using a local phase tensor-guided cnn," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 7, pp. 1127–1135, 2020.
- [23] A. Z. Alsinan, V. M. Patel, and I. Hacıhaliloğlu, "Bone shadow segmentation from ultrasound data for orthopedic surgery using gan," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 9, pp. 1477–1485, 2020.
- [24] A. Li, J. Han, Y. Zhao, K. Li, and L. Liu, "Enabling augmented segmentation and registration in ultrasound-guided spinal surgery via realistic ultrasound synthesis from diagnostic ct volume," *arXiv preprint arXiv:2301.01940*, 2023.
- [25] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.
- [26] Z. Zhang, Z. Min, A. Zhang, J. Wang, S. Song, and M. Q.-H. Meng, "Reliable hybrid mixture model for generalized point set registration," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [27] N. Chebroul, T. Läbe, and C. Stachniss, "Spatio-temporal non-rigid registration of 3d point clouds of plants," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3112–3118.
- [28] F. Magistri, N. Chebroul, and C. Stachniss, "Segmentation-based 4d registration of plants point clouds for phenotyping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2433–2439.
- [29] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid icp algorithms for surface registration," in *2007 IEEE conference on computer vision and pattern recognition*. IEEE, 2007, pp. 1–8.
- [30] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [31] M. Salehi, R. Prevost, J.-L. Moctezuma, N. Navab, and W. Wein, "Precise ultrasound bone registration with learning-based segmentation and speed of sound calibration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 682–690.
- [32] D. Van Tassel, L. E. McMahon, M. Riemann, K. Wong, and C. E. Barnes, "Dynamic ultrasound in the evaluation of patients with suspected slipping rib syndrome," *Skeletal Radiology*, vol. 48, no. 5, pp. 741–751, 2019.
- [33] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger, "Plus: open-source toolkit for ultrasound-guided intervention systems," *IEEE transactions on biomedical engineering*, vol. 61, no. 10, pp. 2527–2537, 2014.
- [34] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [35] R. C. Prim, "Shortest connection networks and some generalizations," *The Bell System Technical Journal*, vol. 36, no. 6, pp. 1389–1401, 1957.
- [36] F. P. Preparata and S. J. Hong, "Convex hulls of finite sets of points in two and three dimensions," *Communications of the ACM*, vol. 20, no. 2, pp. 87–93, 1977.
- [37] Y.-S. Kim, M. J. Park, H. Rhim, M. W. Lee, and H. K. Lim, "Sonographic analysis of the intercostal spaces for the application of high-intensity focused ultrasound therapy to the liver," *American Journal of Roentgenology*, vol. 203, no. 1, pp. 201–208, 2014.
- [38] Z. Jiang, L. Lei, Y. Sun, X. Qi, Y. Hu, B. Li, N. Navab, and J. Zhang, "Model-based compensation of moving tissue for state recognition in robotic-assisted pedicle drilling," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 463–473, 2020.
- [39] X. Dai, Y. Lei, J. Roper, Y. Chen, J. D. Bradley, W. J. Curran, T. Liu, and X. Yang, "Deep learning-based motion tracking using ultrasound images," *Medical Physics*, vol. 48, no. 12, pp. 7747–7756, 2021.
- [40] Z. Jiang, N. Danis, Y. Bi, M. Zhou, M. Kroenke, T. Wendler, and N. Navab, "Precise repositioning of robotic ultrasound: Improving registration-based motion compensation using ultrasound confidence optimization," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.
- [41] J. Tan, B. Li, Y. Leng, Y. Li, J. Peng, J. Wu, B. Luo, X. Chen, Y. Rong, and C. Fu, "Fully automatic dual-probe lung ultrasound scanning robot for screening triage," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2022.
- [42] Z. Jiang, H. Wang, Z. Li, M. Grimm, M. Zhou, U. Eck, S. V. Brecht, T. C. Lueth, T. Wendler, and N. Navab, "Motion-aware robotic 3d ultrasound," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12494–12500.
- [43] Z. Jiang, Y. Zhou, Y. Bi, M. Zhou, T. Wendler, and N. Navab, "Deformation-aware robotic 3d ultrasound," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7675–7682, 2021.