Precise Repositioning of Robotic Ultrasound: Improving Registration-Based Motion Compensation Using Ultrasound Confidence Optimization

Zhongliang Jiang, Member, IEEE, Nehil Danis, Yuan Bi, Graduate Student Member, IEEE, Mingchuan Zhou, Member, IEEE, Markus Kroenke, Fellow, IEEE, Thomas Wendler, and Nassir Navab, Fellow, IEEE

Abstract—Robotic ultrasound (US) imaging has been seen as a promising solution to overcome the limitations of free-hand US examinations, i.e., interoperator variability. However, the fact that robotic US systems (RUSSs) cannot react to subject movements during scans limits their clinical acceptance. Regarding human sonographers, they often react to patient movements by repositioning the probe or even restarting the acquisition, in particular for the scans of anatomies with long structures like limb arteries. To realize this characteristic, we proposed a vision-based system to monitor the subject’s movement and automatically update the scan trajectory thus seamlessly obtaining a complete 3-D image of the target anatomy. The motion monitoring module is developed using the segmented object masks from RGB images. Once the subject is moved, the robot will stop and recompute a suitable trajectory by registering the surface point clouds of the object obtained before and after the movement using the iterative closest point (ICP) algorithm. Afterward, to ensure optimal contact conditions after repositioning US probe, a confidence-based fine-tuning process is used to avoid potential gaps between the probe and contact surface. Finally, the whole system is validated on a human-like arm phantom with an uneven surface, while the object segmentation network is also validated on volunteers. The results demonstrate that the presented system can react to object movements and reliably provide accurate 3-D images.

Index Terms—Blood vessel visualization, medical robotics, robotic ultrasound (US), vision-based control.

I. INTRODUCTION

ULTRASOUND (US) imaging is a powerful and effective tool for the diagnosis of lesions and abnormalities of internal tissues and organs. US imaging is noninvasive, real-time, radiation-free, and widely available. In 2017, over 9.2 million US scans were performed in England. This number is twice and three times larger than the number of computer tomography (CT) scans and magnetic resonance imaging (MRI), respectively, during the same period [1]. Besides, US has been employed as the primary mechanism to diagnose peripheral artery disease (PAD) in clinical practice [2]. Regarding the femoral artery, Collins et al. [3] reported that US imaging has 80%-98% sensitivity in detecting arterial stenoses.

However, due to the inherited limitations of the US modality, the conventional 2-D B-mode images suffer from high intraoperator and interoperator variability. These variations result in inconsistent diagnosis results of examinations carried by sonographers with different level of experience, or even by the same operator at different times. To alleviate this problem, 3-D imaging is employed to intuitively characterize and quantify the abnormality, i.e., vascular stenosis [4], [5].
Compared with 2-D images, 3-D images are more reproducible because they do not require exactly repeating the probe orientation. Ottacher et al. [6] first introduced 3-D US image to accurately guide the process for insertion of pedicle screws.

To augment 2-D images into 3-D volumes, Hossack et al. [7] designed a probe with a 2-D arrangement of transducer elements instead of traditional 1-D arrays. The 2-D array allows to directly visualize the objects of interest in 3-D, while the cost of such systems is high. In addition, the size of the acoustic window is relatively small. In contrast, an optical or electromagnetic tracking system is used as an alternative way to obtain 3-D volumes by providing tracked 2-D images [8]. However, potential magnetic interference and occlusion between the optical camera and the markers limit the usability of such systems in real scenarios. To address this, Yang et al. [9] employed a robotic manipulator to automatically produce 3-D images of human spine.

To further address this challenge, robotic techniques are employed to develop robotic US systems (RUSSs) for accurate and reproducible US acquisitions. Gilbertson and Anthony [10] developed an impedance controller for a one degree of freedom (DoF) device to stabilize US imaging during scans. Huang et al. [11], [12] attached two thin strain sensors on the front face of the US probe to maintain a constant force during scans. Besides contact force, Jiang et al. [13] quantitatively optimize the orientation using both US imaging and estimated force at the end-effector to improve the imaging quality. Yet, this approach is only valid for a convex probe.

To further accurately orient both linear and convex probes to unknown constraint surfaces, Jiang et al. [14] further developed a mechanical model-based approach using the accurate force measurements from an external force sensor. In addition, Huang et al. [12] used a depth camera to compute the normal direction of the target surface for autonomous US scans. Although the accuracy of the camera-based approach is not as high as the force-based approach, the estimated orientation can be computed in real time. Additionally, benefited from force measurements, Sun et al. [15] and Jiang et al. [16] proposed approaches to further recover the force-induced deformation in 2-D and 3-D images, respectively.

To automatically perform the screening of aorta by RUSS, Virga et al. [17] computed a scan trajectory from a preoperative MR image and further transferred it to the robotic frame by registering the MR to the object surface captured by an RGB-D camera. Nevertheless, potential movements of objects during scans have not been considered, which may lead to failure of visualizing target anatomies. Considering the task to display long structures, e.g., a limb artery tree, sonographers even need to actively adjust the limb pose to overcome the limitation of robotic working space and to better visualize the target artery. Since the limbs are flexible, articulated motion will happen around the joints (see Fig. 1). However, the motion of separate parts, i.e., forearm, and upper arm, can be seen as rigid transformation. If the potential motion (articulated or rigid motion) is not considered during scans, the imaging quality will be significantly decayed in the presence of motion as Fig. 1(e). To address this, Jiang et al. [18] developed a marker-based method to monitor and compensate for potential motion during scanning. Yet, this approach relied on optical markers, which limits the usability of the approach in real scenarios. In addition, it is time consuming to carefully configure the markers on individual patients.

To further tackle this challenging, we proposed an surface registration-based motion-aware RUSS with the ability to reposition the probe for precise and complete 3-D images of target anatomies in presence of rigid motions during scanning. This system combines the advantages of free-hand US (flexibility) and RUSS (accuracy and stability). To enable the motion-aware ability, an RGB-D camera is employed to monitor and compensate for potential object movements. This technique is particularly important for the development of robotic screening of PAD on limb arteries, where sonographers usually need to actively adjust the limb position to avoid exceeding working space limits and to visualize the whole artery tree. The main contributions are summarized as follows.

1) A learning-based approach is developed to monitor the object’s movement and automatically update the initial trajectory based on the surface registration technique to accomplish the scan for a complete and accurate 3-D geometry. Compared with the state-of-the-art work [18], the proposed surface registration-based approach is more compatible with the current clinical practices without the requirements of time-consuming tuning of patient-specific configurations.

2) An online probe orientation adjustment method is developed based on the US confidence map [19] to optimize the contact condition between probe and object surface during scans. Besides the aim of guaranteeing the image quality, this adjustment also benefits the visibility of the target object in the imaging view against surface registration errors.

The segmentation approach are validated on both human arms and an arm phantom. In addition, the proposed motion-aware RUSS is fully validated on the human-like arm phantom with an uneven surface. The video can be publicly accessed.

II. System Overview

In this section, we first describe the overview of the proposed motion-aware RUSS for achieving an accurate and complete 3-D volume of target anatomies against potential (expected or unexpected) object movements during the scanning (see Section II-A). Then, the system calibration procedures and the compliant controller used to maintain a given force throughout US sweep are described in Sections II-B and II-C, respectively.

A. Workflow

To enable the motion-aware ability, a vision-based approach is developed in this study to detect the (expected or unexpected) movements and update the sweep trajectory based on the surface registration approach. The general pipeline
is described in Fig. 2. Compared with the marker-based approach [18], the passive markers are not needed anymore and a human-like arm phantom is used to mimic real scenarios.

1) **Hardware Setup:** The proposed motion-aware RUSS mainly consists of three components: a redundant robotic arm (LBR iiwa 14 R820, KUKA GmbH, Germany), a US machine (Cephasonics, USA), and an RGB-D camera (Azure Kinect, Microsoft Corporation, USA). The robotic arm has seven joints, and external joint torque sensors are integrated into all joints. This configuration enables the accurate control of contact force at the end-effector. To have an overview of the whole setup, the camera is fixed in front of the robotic arm phantom and volunteer’s arm. The resolutions of RGB camera and depth camera are set to 2048 × 1536 and 1024 × 1024, respectively.

The robot is controlled using a self-developed robot operating system (ROS) interface [20]. The control commands are updated at 100 Hz to guarantee real-time performance. A linear probe (CPLA12875, Cephasonics) is attached to the robotic flange using a custom-designed holder. The B-mode images are accessed via a USB interface (50 frames/s) and visualized in a software platform (ImFusion Suite, ImFusion GmbH, Germany). The detailed configurations of US acquisition are listed as follows: imaging depth: 55 mm, brightness: 67 dB, frequency: 7.6 MHz. To validate the proposed system, a commercial human-like arm phantom (PICC US Training Model, Skills Med Deutschland GmbH, Germany) is employed.

2) **Implementation Pipeline:** The motion-aware RUSS is realized by three main components: the sweep trajectory extraction and optimization module, the robotic scanning execution module, and the camera-based movement monitoring module. With the extraction of sweep trajectory, the artery of interest is manually segmented from preoperative CT data using an open-source software 3-D slicer.\(^2\) Then, the meshes of the arm surface and the target limb artery surface are generated and further used to create point clouds using Meshlab\(^3\) [see Fig. 2(a)]. To transfer the target artery position from CT data to the robotic base frame, a self-occlusion point cloud of the arm surface is generated based on the RGB-D camera placed in front of the target object. After the registration between the two surface point clouds, the artery trajectory can be transferred into the camera frame. Further utilizing the hand-eye calibration result, the artery trajectory can be transferred into the robotic base frame to guide the probe movement for autonomous scans.

Afterward, the RUSS starts executing US examination along the artery. To provide good imaging quality and guarantee patient safety during the scanning, a compliant controller is employed to maintain a constant force between the probe and the contact surface [20], [21]. In addition, to monitor the object movement, a neural network (UNet-VGG16 [22]) was trained to segment the arm from RGB images. The segmented results of images acquired at \(t_1\) and \(t_2\) are further used to compute the dice coefficient. Once the dice coefficient is smaller than a preset threshold, the system considers the object has moved, and a corresponding signal is emitted to the ROS master. After detecting such a signal, the robot stops, and the registration between the camera point clouds acquired at \(t_1\) and \(t_2\) is carried out. Based on the registration result, the remaining trajectory can be updated, and RUSS can continue the scan by automatically moving the probe to the breaking point, namely, the probe position when the movement happens. Benefited from this compensation, a complete and accurate 3-D compounding can still be achieved when the object moves during the scans.

**B. Coordinate System Transformation**

There are two different types of calibration procedures involved in this study: 1) US calibration and 2) hand-eye

\(^{2}\)https://www.slicer.org/  
\(^{3}\)https://www.meshlab.net/
calibration. The former one is used to transfer the pixel position in B-mode images to the robotic base frame while the latter one is used to transfer the position from the camera view to the robotic base frame to guide robotic movement. The involved coordinate systems are depicted in Fig. 3: 1) the robotic base frame \( \{b\} \); 2) the robotic flange frame \( \{f\} \); 3) the probe frame \( \{p\} \), 4) US imaging frame \( \{us\} \); 5) the camera frame \( \{c\} \); and 6) the CT atlas frame \( \{ct\} \).

Fig. 3. Illustration of involved coordinate frames. Red arrows, green arrows, and blue arrows represent \( x \)-, \( y \)-, and \( z \)-directions of corresponding coordinate systems, respectively.

To perform trajectory correction based on US imaging feedback, the transformation matrix \( T_{us} = \frac{b}{f} T_{p} \frac{p}{us} T \) can be calculated as follows:

\[
\begin{align*}
 T_{us} &= \frac{b}{f} T_{p} \frac{p}{us} T \\
 &= \begin{bmatrix}
 b \times f & f \times p & p \times us \\
 -L_p & 0 & 0 \\
 0 & -L_p & 0 \\
 0 & 0 & \varepsilon_0
\end{bmatrix}
\end{align*}
\]

where \( T_{us} \) is the transformation matrix used to transfer the position from frame \( \{i\} \) to frame \( \{j\} \). \( T_{us} \) represents the kinematic model of the robotic arm. This can be directly accessed by the application programming interface (API) function provided by the manufacturer. Besides, \( T_{us} \) depends on custom configuration. In this work, the rotation part of \( T_{us} \) is set to an identity matrix \( \mathbf{1} \times 3 \) and the translational part is obtained from the custom-designed 3-D holder model. Due to the characteristic of linear probe that US elements are physically distributed on the probe tip within a specific length \( L_p \), the horizontal mapping \( (L_p/W_{us}) \) between pixelwise positions and physical positions can be computed, where \( W_{us} \) is the pixelwise width of B-mode imaging. Similarly, the vertical mapping is computed by \( (D/H_{us}) \), where \( D \) and \( H_{us} \) are the depth setting and the pixelwise height of US images. The origin of frame \( \{us\} \) was set at the upper left of the B-mode image, while the origin of frame \( \{p\} \) was set in the middle of the probe tip (see Fig. 3). Thus, \( T_{us} \) is calculated as follows:

\[
\begin{align*}
 T_{us} &= \begin{bmatrix}
 0 & 0 & -1 & 0 \\
 -L_p & 0 & 0 & L_p \\
 0 & 0 & & 2 \\
 0 & D & 0 & \varepsilon_0 \\
 0 & 0 & 0 & 1
\end{bmatrix}
\end{align*}
\]

where \( L_p = 37.5 \) mm, \( D = 55 \) mm, and hyperparameter \( \varepsilon_0 \) is used to neutralize the indeterminacy of US elements configuration.

In addition, regarding the hand-eye calibration, similar implementation procedures are carried out as [18]. To compute the transformation matrix \( T \) between the base frame \( \{b\} \) and the camera frame \( \{c\} \), paired coordinate descriptions of the points (intersection points on two chessboards) in frame \( \{c\} \) and frame \( \{b\} \) are recorded. The coordinate representations in \( \{c\} \) are computed using OpenCV while the representations in \( \{b\} \) are obtained by manually moving a custom-designed pointer tool to the same intersection points. More implementation details can be found in [18].

C. Compliant Control Architecture

Since the acoustic impedance of air is much lower than human tissues (usually over 1000 times lower), the reflection coefficient at the air–skin interface is around one (i.e., almost all signal gets reflected). Thus, a certain contact force between a US probe and object surface is necessary to guarantee the visibility of target anatomies. To eliminate the negative influence of nonhomogeneous deformations in 3-D results, a compliant controller using the built-in joint torque sensors was employed [20], [21]. The Cartesian compliant controller law is described as follows:

\[
\tau = J^T [K_m (x_d - x_c) + F_d] + D(dm) + f_{dyn}(q, \dot{q}, \ddot{q}) \quad (3)
\]

where \( \tau \in \mathbb{R}^{7 \times 1} \) is the target torque for all joints, \( J^T \in \mathbb{R}^{7 \times 6} \) is the transposed Jacobian matrix, \( x_d \in \mathbb{R}^{6 \times 1} \) and \( F_d \in \mathbb{R}^{6 \times 1} \) are the desired Cartesian position and force, \( x_c \in \mathbb{R}^{6 \times 1} \) is current position, \( K_m \in \mathbb{R}^{6 \times 6} \) represents the Cartesian stiffness, \( D(dm) \in \mathbb{R}^{7 \times 1} \) is the damping term, and \( f_{dyn}(q, \dot{q}, \ddot{q}) \in \mathbb{R}^{7 \times 1} \) is the dynamic model of the robotic arm.

Considering the characteristic of US examinations, the constant force is maintained in the direction of the probe centerline by a one-DoF compliant controller, while the other five-DoF motion is controlled to accurately follow the planned sweep trajectory. These two types of controllers are fused by assigning varied stiffness values. High stiffness (translational DoF: 2000 N/m and rotational DoF: 200 Nm/m) are assigned to the DoFs controlled in position mode while the stiffness of the compliant DoF is set between [125, 500] N/m for different human tissues [20].

III. MOTION-AWARE ROBOTIC US SCANNING

This section describes the characteristic aspects of the proposed system. The method used to automatically create scan trajectory from a template image is described in Section III-A. In Section III-B, a confidence-based in-plane orientation optimization is proposed to optimize the contact condition between probe and object surface. Section III-C demonstrates a markerless motion compensation method designed based on surface point cloud registration and a fine adjustment for further eliminating the stitching gap in 3-D view.

A. Surface Registration and Trajectory Transfer

Since the object position varies in different examinations, a depth camera is employed to gain information about the current environmental setup. To properly execute US scan for the target artery, the annotated artery is required to be transferred
into camera frame \( \{ c \} \). To obtain such transformation \( \mathbf{T} \), the iterative closest point (ICP) algorithm is used to register the surface point cloud \( \mathbf{P}_{ct} \) obtained from a preoperative CT image to the live point cloud \( \mathbf{P}_c \) computed from the RGB-D camera. ICP proved to be sufficient for the targeted task in preliminary experiments. To achieve uniformly distributed \( \mathbf{P}_{ct} \), the Poisson disk sampling algorithm [23] is employed instead of the random sampling approach. Since ICP is sensitive to the initial result, the multiscale feature resistance approach is used to extract unique features from both point clouds for computing the initial transformation [24]. Afterward, the iterative algorithm is used to accurately align \( \mathbf{P}_{ct} \) to \( \mathbf{P}_c \).

To correctly and robustly align \( \mathbf{P}_{ct} \) to \( \mathbf{P}_c \), the distinctiveness of each feature descriptor is important, which will significantly affect the number of possible ambiguous correspondences between two point clouds. To this end, the multiscale feature resistance approach is employed, which computes the same class of feature descriptor at various scales and further refines the point pairs that were proved to be distinctive over multiple scales. The default scales are 0.5, 1.0, and 1.5. Then, the paired correspondences are further filtered in terms of consistency based on the distance threshold, which is empirically selected as 0.25 based on the final performance. The implementation is based on the Point Cloud Library (PCL)\(^4\) and more details can refer [24].

Since the artery is located inside the arm, the artery trajectory is required to be projected to the surface for generating an executable scanning trajectory. To alleviate potential negative influences of registration error between \( \mathbf{P}_{ct} \) and \( \mathbf{P}_c \) on creating scanning trajectory, the point cloud of the artery surface in CT \( \mathbf{P}_{art} \) is directly transferred into the camera frame. Then, \( \mathbf{P}_{ct} \) can be exactly projected onto the object surface captured by the camera. Based on the hand-eye calibration result described in Section II-B, the scanning trajectory is computed.

1) Artery Centerline Extraction: The artery surface point clouds have a tubular structure. Besides, regarding the geometry of limb arteries, the length along the vascular centerline is significantly longer than in other directions (namely, artery diameter). To extract the vascular centerline for generating scanning trajectory, the principal component analysis (PCA) is employed to define a new orthogonal coordinate system for optimally describing variance of the data. The first principal component is the direction making the projections achieve the largest variance, namely the eigenvector corresponding to the largest eigenvalue of the artery point cloud’s covariance matrix. To estimate the vascular centerline, the average value of all points inside a small artery segmentation is computed to an approximate local center point. A small distance interval \( d_{in} \) is used to generate multiple center points between the minimal and maximal values of the projections in the first principal component. The vascular centerline is approximated by connecting the center points successively.

2) Scanning Trajectory on Object Surface: After deriving the vascular centerline, an executable scanning trajectory on the object surface should be further generated. To avoid potential safety issues caused by improper scanning trajectories, e.g., inside of the arm, the scanning trajectory is generated as following steps. Since the robot and the experimental table are parallel to the ground, a continuous trajectory can be created by projecting the estimated vascular centerline onto the surface in the direction of \( Z_b \). To achieve this, the center points of the artery \((x, y, z)\) computed in Section III-A1 is replaced with \((x, y, z_{arm}^{\text{max}})\), where \( z_{arm}^{\text{max}} \) is the maximum value of the arm surface point cloud \( \mathbf{P}_{ct} \) in \( Z_b \) direction. Then \( K \)-nearest neighbors (KNN) approach is used to search for the \( K = 5 \) nearest points around \((x, y, z_{arm}^{\text{max}})\), individually, in terms of Euclidean distance. Then, the average of these \( K_M \) neighbors is used as key points of the scanning trajectory for RUSS.

To fully control the probe during the scanning, the probe orientation needs to be further determined. To improve the contrast of the resulting B-mode imaging, Jiang et al. [13] and Lhønsenka and Boezaart [25] suggested that the probe should be aligned in the normal direction of the contact surface \( \hat{n}_i \). Here, the normal direction is quickly estimated using computer vision techniques based on the neighboring points around the key points. Compared with the force-based method [13], the vision-based method is able to quickly compute all the normal directions along the whole trajectory rather than only the current contact point. This process is implemented using PCL. The estimated normal directions at various positions are shown in Fig. 4. To properly visualize the target vessel, the probe centerline is aligned to the computed normal direction and the long axis of the probe \((Y_p)\) is aligned to be orthogonal to the scanning path.

B. Confidence-Based Probe Orientation Correction

US confidence map was originally introduced by Karamalis et al. [19]. They provide a pixelwise metric to assess the image quality by computing the loss of emitted US signals by the transducer. Regarding the computed confidence maps of B-mode images, a probabilistic map \( C \in \mathbb{R}^2 \rightarrow [0, 1] \) is created (see Fig. 5). The top white pixels \( (1) \) represent the strongest signal and the bottom black pixels \( (0) \) represent no US signal arrived. US confidence map has been employed for different aims, e.g., optimizing contact force [17] and tracking target anatomy [26].

In this study, the computed scanning trajectory may not be perfect for scanning due to the error of both hand-eye calibration and surface registration (ICP). The use of a human-like arm phantom with an uneven surface aggravates the negative influence of such errors. More specifically, the US probe may not be able to fully contact the arm surface, which will result in shadow in the imaging (see Fig. 5). To correctly display the anatomy under the shadow, a confidence-based

\(^4\)https://pointclouds.org/
orientation correction approach is proposed. To this end, the confidence map is further transferred into a binary image by applying a probabilistic threshold. This emphasizes the difference between the well-contacted part and the noncontact part in the resulting images. Then, the weighted barycenter point \( \zeta_c \in \mathbb{R}^{2 \times 1} \) can be calculated as follows:

\[
\zeta_c = \frac{1}{N_c} \sum_{i=1}^{N_c} w_{i} C(h, w)[h, w]^T
\]

where \( C(h, w) \) is binary confidence value at pixel grid \((h, w)\), \( N_c = 550 \) and \( W_{as} = 375 \) are the image’s height and width in pixel. \( N_c = \sum_{(h, w)\in\Omega} C(h, w) \) is the accumulated confidence value over the entire image (\( \Omega \)).

Then, the in-plane rotational adjustment angle \( \theta_c \) can be calculated by connecting the barycenter and the top center point of image view (see Fig. 5). To ensure that the probe is fully in contact with the object surface, multiple in-plane orientation corrections may be required at one position. To reduce the time consumed at each position, the pose \( R_p \in \mathbb{R}^{3 \times 3} \) of the remaining \( N_{up} \) points in the trajectory are updated as follows:

\[
R_p^i(i, i + 1) = R_p^i(i_{c}, i) R_x(-\eta_i / \theta_c)
\]

where \( i_{c} \) represents the iteration of current position in the trajectory, \( i = 1, 2, \ldots, N_{up} \). \( R_x \in \mathbb{R}^{3 \times 3} \) is the rotation matrix around \( x \)-direction of frame coordinate system. \( \eta_i = (d^2_{N_{up}-i+1}/(\sum_{i=1}^{N_{up}} d^2_j)) \) is the weight coefficient used to update the next \( N_{up} \) positions.

\section{Movement Identification and Compensation}

\subsection{Movement Monitor System}

To effectively detect the potential movement (expected or unexpected) during the scanning, the UNet-VGG16 architecture [22], [27] is employed to segment the arm from RGB images. The U-Net was proposed by Ronneberger et al. [28] for segmentation tasks based on fully convolutional networks (FCNs). U-Net consists of an encoder and a decoder. To achieve accurate segmentation, skip connections are used to transfer the features from the encoder to the decoder. This is one of the most common network structures used for semantic segmentation. Compared to training the UNet encoder from scratch, VGG16 [29] is already trained on a large-scale ImageNet classification dataset (over 14 million images) [30], which has already learned enough features. Thereby, UNet-VGG16 replaces the UNet encoder with a well-trained VGG16 model to extract the image features, which can improve the overall performance of the network and as well as reduce the training time and training dataset size [27].

To train an UNet-VGG16 model to efficiently segment the arm in this setup, 270 images were collected by changing both the camera position and the environments, e.g., different backgrounds, occlusion caused by the robotic arm and US probe, and with/without humans in the field of view. Then, the arm was manually annotated on all images using LabelMe software [31]. To augment the data for better generalization, the images are randomly flipped and the color are also randomly modified by changing the related parameters, i.e., brightness and contrast. Finally, 3240 images are generated. The ratio between the training and the validation datasets is 8:2. The size of the input images is 512 \times 512. In addition, the pixelwise mean squared error (mse) \( L_{mse} \) is employed as the loss function to train the network

\[
L_{mse} = \frac{1}{N} \sum_{i=1}^{C_i} \sum_{j=1}^{H_i} \sum_{w_j} [y_c(i, j) - \hat{y}_c(i, j)]^2
\]

where \( N = C_i \cdot H_i \cdot W_j \); \( C_i, H_i, \) and \( W_j \) are the number of channels, height, and width, respectively. \( y \) and \( \hat{y} \) are the annotated image and output of the network.

To validate whether the proposed pipeline is compatible for real scenarios, the second UNet-VGG16 model was trained separately for human volunteers. To avoid articulated movements, only the forearm was selected as target object. The dataset of human arm was generated using the same manner as arm phantom data. Finally, 3560 labeled images from two healthy volunteers (BMI: 25.0 \pm 2.8, age: 28.5 \pm 1.5) were separated as 8:2 for training and validation.

After obtaining the binary masks using the trained model, the dice coefficient \( C_d \) between the segmented results from the current frame (\( i \)th) and the previous frame (\( i - j \)th) is calculated as follows:

\[
C_d = \frac{2 |S_i \cap S_{i-j}|}{|S_i| + |S_{i-j}|}
\]

where \( S_i \) is the segmented binary results of \( i \)th camera image.

\subsection{Movement Compensation}

If a movement happened during the scanning, the real-time dice coefficient decreases correspondingly. Thus, \( C_d \) is used as the metric to identify the object movement. Once \( C_d \) becomes smaller than a preset threshold \( T_{dice} \), a signal is emitted to inform the controller to stop scanning via ROS. Based on the experimental results, \( T_{dice} \) was empirically set as 0.95 to guarantee both the sensitivity and the robustness of motion detection results in our setup. To properly resume the sweep from the breaking point, the transformation between previous (before the movement) and current objects is calculated using ICP (see Fig. 2). Here, the point clouds of the arm surface are created based on the mask of RGB images. After properly assigning the depth information to the pixels located inside the bounding box of the mask, a raw point cloud can be generated. To extract the arm, the plane segmentation algorithm is applied to remove the points under the table surface and an additional depth filter is employed to remove the points higher than a certain value in \( Z_h \) direction of frame \( \{ h \} \). To validate whether the registration
results are good enough for continuing the scanning from the break point, an ArUco marker is rigidly placed on the table. To quantitatively access the registration results, the motion compensation error $e_{mc}$ is computed as follows:

$$e_{mc} = \left\| P'_m - (R_{mc} \cdot P_m + V_{mc}) \right\|$$

(8)

where $P_m \in \mathbb{R}^{3 \times 1}$ and $P'_m \in \mathbb{R}^{3 \times 1}$ are the positions of the ArUco marker before and after the movement, respectively, while $R_{mc} \in \mathbb{R}^{3 \times 3}$ and $V_{mc} \in \mathbb{R}^{3 \times 1}$ are the computed rotation matrix and translation vector.

The RUSS automatically resumes the scanning from the break point if $e_{mc}$ is small enough ($<1$ cm). Otherwise, the RUSS automatically ends the sweep. Compared with the state-of-the-art marker-based approach [18], only camera images make it more suitable for real clinical routines without the need to carefully configure the markers on various patients.

3) Fine Adjustment Algorithm: To well stitch the two or multiple sweeps when movement happens during scans, a two-step fine adjustment procedure is further developed to displaying a complete and continuous 3-D anatomy. To achieve this, the probe pose of the last frame in the first sweep (before movement) ($T_{be}^i \in \mathbb{R}^{4 \times 4}$) and the probe pose of the first frame in the second sweep (after movement) ($T_{bf}^i \in \mathbb{R}^{4 \times 4}$) are used. To overlap these two frames, the tracked probe pose ($T_{mc}^i \in \mathbb{R}^{4 \times 4}$) of the first sweep can be updated as follows:

$$\begin{bmatrix}
R_{mc}(i) \\
0 \\
1
\end{bmatrix} = \begin{bmatrix}
T_{af}^{-1} & T_{mc} & T_{be}^{-1}
\end{bmatrix} T_{mc} T_{be}(i)$$

where $T_{mc}$ is the homogeneous expression of $R_{mc}$ and $V_{mc}$ in (8).

To further achieve good continuity of vascular boundary in 3-D, an in-plane translational adjustment is carried out. The translational adjustment is computed based on the two centroids ($P_{be}$ and $P_{af}$) of the vessel of interest from the last frame in the first sweep and the first frame in the second sweep, respectively. The centroid is computed based on the binary mask computed using a well-trained U-Net model as [21]. To train the network, 3369 B-mode images were used. Since the batch size was set to ten to improve the generalizability, the batch normalization layers used in the original U-Net [28] are replaced by the group normalization approach. The training details are exactly the same as [21]: the ratio between training and validation dataset is 9:1, the learning rate is initialized as 0.001, the optimization method is ADAM optimizer, and the epoch is 10. Finally, the tracked probe pose of the sweep obtained before the movement is further updated as follows:

$$P_{mc}^i(i) = P_{mc}^i(i) + R_{mc} (P_{af}^i - P_{mc}^i).$$

(10)

IV. RESULTS

A. Segmentation Results on Phantoms and Human Arm

The UNet-VGG16 network parameters were optimized using Adam [32] on a single GPU (Nvidia GeForce GTX 1080). The learning rate was set to $10^{-3}$ at the beginning and reduced by a factor of 10 when the loss changes were less than 0.0001 for ten subsequent steps. The batch size and the training epochs are 4 and 10, respectively. It can be seen from Fig. 6 that both training loss and validation loss were quickly reduced at the beginning and gradually converged after 500 iterations. The performance of the well-trained models on unseen images of arm phantom and human forearms are summarized in Table I. The average dice coefficient achieved 0.94 and 0.95 on arm phantom and human forearms, respectively. Besides, the segmentation time for each image only takes around 5 ms.

To demonstrate the performance of the segmentation result, four representative results on arm phantom and two different human arm are shown in Fig. 7. Regarding the arm phantom, Fig. 7(a) and (b) display segmentation results form RGB images when the arm surface is without/with occlusion caused by US probe, respectively. Regarding the experiments on volunteers, a normal blue tape was warped around human elbow. The tape servers as boundary of forearm and its position is easily changed and determined based on target anatomies by human operators. The segmentation results on two different volunteers are shown in Fig. 7(c) and (d). Even when the
occlusion happens, the forearm boundary is still completely extracted. These results demonstrate that well-trained models has the potential to accurately extract the target objects. The dice coefficient of these four representative results are 0.96, 0.95, 0.97, and 0.96, respectively. Accurate and robust segmentation results have been achieved on both phantom and human arms, which demonstrates that the position of the RGB-D camera will not significantly affect the results. In addition, more live segmentation results can be found in the online video.

B. ICP-Based Surface Registration Performance

To demonstrate the performance of the ICP-Based surface registration approach in our setup, some representative results have been shown in Fig. 8. The mse of the distance between corresponding points in the two point clouds \( \mathbf{P}_t \) and \( \mathbf{P}_c \) quickly decreased at the beginning and converged \( (0.2 \text{ mm}) \) after 25 iterations. The initial alignment is shown in Fig. 8(b), which is computed using the multiscale feature resistance approach [24]. Then the result after ten and 40 iterations are depicted in Fig. 8(c) and (d). It can be seen that the point cloud obtained using an RGB-D camera can be well registered to a preoperative template (mse is 0.16 mm).

Considering the potential occlusion in a real scenario, the extracted point cloud from an RGB-D camera \( \mathbf{P}_c \) could be incomplete. To further validate the robustness of the ICP-based Registration, \( \mathbf{P}_c \) is cropped by a plane being orthogonal to the first principal direction computed using PCA. To investigate the influence caused by various levels of occlusion, \( \mathbf{P}_c \) is cropped by the plane at 10%, 20%, and 40% on the first principal direction [see Fig. 9(a)–(c), respectively] and two planes at 10% and 90% on the first principal direction. \( \mathbf{P}_c \) is cropped by two planes at 10% and 90% on the first principal direction. (a) Iteration = 64. (b) Iteration = 68. (c) Iteration = 83. (d) Iteration = 168.

C. Performance of Confidence-Based Trajectory Correction

Due to the error of hand-eye calibration and CT-to-object registration, the computed trajectory may result in sub-optimal contact during the scan. This reduces the quality of US imaging, like introducing a shadow in the image view as shown in Fig. 10(b). Thus, a confidence-based orientation correction is developed to guarantee good contact condition between the probe and the arm surface during the scan. To demonstrate the performance of the confidence-based correction, two sweeps with an exactly same initial trajectory are performed. The representing results are depicted in Fig. 10.

It can be seen from Fig. 10, part of the target artery is out of the imaging view for the case without correction. This is represented by the red part in 3-D and white circle (incomplete) in 2-D [see Fig. 10(a) and (b), respectively. In the case involving correction, once a shadow is detected using the confidence map, an in-plane adjustment angle \( \theta_c \) is computed and the correction is automatically carried out by the robot. After such correction, the entire artery geometry has been successfully visualized in the imaging view [see Fig. 10(d)]. In addition, since the poses of the remaining positions in the trajectory are also updated using (5), the target vessel is completely visualized inside the imaging view [see Fig. 10(d)]. This result demonstrates that the proposed confidence-based orientation can help to improve the visibility of the target object during the sweep.

D. Movement Compensation Results

The compensation is done by two successive procedures: 1) surface registration-based compensation and 2) the fine adjustment process using both robotic tracking information and corresponding B-mode images. To qualitatively and quantitatively validate the compensation algorithm, a complete artery of an arm phantom is used as the target object (total length is around 440 mm).

1) Performance of Surface Registration-Based Compensation: To quantitatively assess the surface registration performance, the phantom arm together with the used table was randomly moved inside of a rectangle \( (110 \text{ mm}) \) and the maximum rotation angle variation was 80°. To assess the performance of the surface registration-based compensation, an ArUco marker is physically placed on the flat table (see Fig. 3). The real-time pose (position and orientation) of the ArUco marker can be computed using an ROS package.\(^5\)

\(^5\)https://github.com/pal-robotics/arucoros
Thus, the motion compensation error $e_{mc}$ is computed by (8).

Since the arm surface could be partly occluded by the used robotic manipulator (see Fig. 7(b) and (d)) during the scanning, the experiments are separately performed when (case 1) the robotic arm is out of the view and (case 2) is in contact with the phantom arm surface. The experiments were repeated 20 times for each case. The final results are shown in Fig. 11. The $e_{mc}$ ($\pm$SD) of case 1 and case 2 are $6.1 \pm 1.6$ and $6.9 \pm 1.6$ mm, respectively. The results obtained when there is an occlusion in the camera view (case 2) is slightly larger than the results of the ideal case (case 1). Besides, based on a t-test (probability $p = 0.14 > 0.05$), it was concluded that no significant difference exists between the two experimental sets.

To perform the t-test, the results obtained from both case 1 and case 2 are proved to be normally distributed using the Lilliefors goodness-of-fit test. Considering our application, since $6$ mm is much smaller than the probe width $37.5$ mm, the surface registration-based compensation algorithm is promising to be used for compensating a large movement that happens during the scanning. Thereby, a complete geometry of the anatomy can be computed as in Fig. 12(b).

2) Performance of Fine-Tuning Procedure: After surface registration-based compensation, two partitions have been stitched together. However, due to the error $e_{mc}$, there is a significant gap and dislocation at the connected part as in Fig. 12(b). To further address this issue, two-step fine adjustment procedures are carried using (9) and (10) based on the both robotic tracking information and B-mode images. The tracking information is used to fully overlap the last frame of the first partition and the first frame of the second partition. Then an image-based in-plane adjustment is performed to overlap the two centroids of the target artery in the overlapped two frames. The 3-D result after fine adjustment is shown in Fig. 12(c), where the stitching has been successfully compensated. This will further enable autonomous diagnosis.

V. DISCUSSION

Precise repositioning of US probe is one of the crucial techniques for guaranteeing accurate and complete geometry of target anatomies when the object’s pose is changed during scanning. This study first proposed a systematic pipeline for RUSS to monitor and further compensate for the potential object motion only on a depth camera. The proposed motion-aware RUSS performs boldly in our setup of experiments (on a human-like arm phantom). However, there are still some limitations that need to be discussed. First, in this work, we have so far considered the rigid motion of object. The articulated motion and deformation of object are existing as well for US scanning. However, the articulated motion could be consider as multiple rigid motion around joints. Regarding the force-induced deformation, one can apply stiffness-based approach to achieve zero-compress volume [16]. In addition, the ICP-based compensation method was designed for visible object motions during US scans. The tremor or vessel motion [33] have not been considered. However, due to the use of the compliant controller, the small and fast physiological motion can be partially adapted by the controller. Thus,
compared with free-hand US acquisition, the effect of such small motions is not aggravated in our setup. At this moment, the presented RUSS can be used for the potential applications on rigid body parts, i.e., forearm, upper arm, thigh, shank, or whole chest. By further considering articulated motion [34] and integrating deformation correction approach [16] and nonrigid registration [35] in the future, the system has the potential to provide reliable US images in clinical practice. It is also noteworthy that the fine-tuning process theoretically cannot improve the accuracy of the 3-D geometry of the vessel of interest. The objective of the fine-tuning process is to improve the visualization performance and further enable the autonomous diagnosis of PAD by estimating the local diameter variations in future work.

VI. CONCLUSION

This work introduces a motion-aware RUSS only based on an RGB-D camera. This system enables achieving accurate and consistent 3-D images of target anatomy when the object is changed during US scans. In this way, the proposed system possesses the advantages of free-hand US (flexibility) and robot US (accuracy and stability) at the same time. To this end, the scanning trajectory on skin is generated by registering the surface point cloud obtained from the camera to the CT template with annotated objects under skin. To monitor object movements, dice coefficients are computed based on the binary masks of object segmented from two different frames using a well-trained UNet-VGG16 model. Once the object is moved, the surface registration-based approach is employed to automatically update the sweep trajectory and resume the scan from the break point to seamlessly accomplish the planned sweep. Since the ICP-based registration may result in suboptimal contact condition between probe and object after repositioning, a confidence-based probe orientation optimization approach is employed. Finally, to accurately display 3-D objects without significant stitching gaps as [18], a stepwise fine adjustment procedure is carried out based on both tracking data and B-mode images [see Fig. 12(c)]. Such improvement we believe can make RUSSs more robust and thus, bring them closer to clinical use. Furthermore, it enables the research topic of automatic examination and diagnosis, which is significantly meaningful in pandemic and undeveloped countries and areas.

ACKNOWLEDGMENT

The authors would like to acknowledge the Editor-in-Chief, Associate Editor and anonymous reviewers for their contributions to the improvement of this article.

REFERENCES

[1] P. R. Hoskins, K. Martin, and A. Thrush, Diagnostic Ultrasound: Physics and Equipment. Boca Raton, FL, USA: CRC Press, 2019.

[2] V. Aboyans et al., “2017 ESC guidelines on the diagnosis and treatment of peripheral arterial diseases, in collaboration with the European Society for vascular surgery (ESVS),” Eur. Heart J., vol. 39, no. 9, pp. 763–816, 2018.

[3] R. Collins et al., “A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease,” Health Technol. Assessment (Winchester, England), vol. 11, no. 20, pp. iii–184, 2007.

[4] A. Fenster, C. Blake, I. Gyacsov, A. Landry, and J. D. Spence, “3D ultrasound analysis of carotid plaque volume and surface morphology,” Ultrasonics, vol. 44, pp. e153–e157, Dec. 2006.

[5] S. Merouche, L. Allard, E. Montagnon, G. Soulez, P. Bigras, and G. Cloutier, “A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of B-mode images,” IEEE Trans. Ultrason., Ferroelectr.,Freq. Control, vol. 63, no. 1, pp. 35–46, Jan. 2016.

[6] D. Ottacher, A. Chan, E. Parent, and E. Lou, “Positional and orientational accuracy of 3-D ultrasound navigation system on vertebral phantom study,” IEEE Trans. Instrum. Meas., vol. 69, no. 9, pp. 6412–6419, Sep. 2020.

[7] J. A. Hossack, T. S. Sumanaweera, S. Npel, and J. S. Ha, “Quantitative 3-D diagnostic ultrasound imaging using a modified transducer array and automated imaging technique,” IEEE Trans. Ultrason., Ferroelectr.,Freq. Control, vol. 49, no. 8, pp. 1029–1038, Aug. 2002.

[8] A. Gee, R. Prager, G. Treece, and L. Berman, “Engineering a free-hand 3D ultrasound system,” Pattern Recognit. Lett., vol. 24, no. 4, pp. 757–777, Feb. 2003.

[9] C. Yang, M. Jiang, M. Chen, M. Fu, J. Li, and Q. Huang, “Automatic 3-D imaging and measurement of human spines with a robotic ultrasound system,” IEEE Trans. Instrum. Meas., vol. 70, pp. 853–859, Dec. 2021.

[10] M. W. Gilbertson and B. W. Anthony, “Force and position control system for freehand ultrasound,” IEEE Trans. Rob., vol. 31, no. 4, pp. 853–849, Aug. 2015.

[11] Q. Huang, B. Wu, J. Lan, and X. Li, “Fully automatic three-dimensional ultrasound imaging based on conventional B-scan,” IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 2, pp. 426–436, Apr. 2018.

[12] Q. Huang, J. Lan, and X. Li, “Robotic arm based automatic ultrasound scanning for three-dimensional imaging,” IEEE Trans. Ind. Informat., vol. 15, no. 2, pp. 1173–1182, Feb. 2019.

[13] Z. Jiang et al., “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, Apr. 2020.

[14] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab, “Automatic force-based probe positioning for precise robotic ultrasound acquisition,” IEEE Trans. Ind. Electron., vol. 68, no. 11, pp. 11200–11211, Nov. 2021.

[15] S.-Y. Sun, B. W. Anthony, and M. W. Gilbertson, “Trajectory-based deformation correction in ultrasound images,” Proc. SPIE, vol. 7629, Mar. 2010, Art. no. 76290A.

[16] Z. Jiang, Y. Zhou, Y. Bi, M. Zhou, T. Wendler, and N. Navab, “Deformation-aware robotic 3D ultrasound,” IEEE Robot. Autom. Lett., vol. 6, no. 4, pp. 7675–7682, Oct. 2021.

[17] S. Virga et al., “Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Oct. 2016, pp. 508–513.

[18] Z. Jiang et al., “Motion-aware robotic 3D ultrasound,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2021, pp. 12494–12500.

[19] A. Karamalis, W. Wei, and N. Klein, and N. Navab, “3D ultrasound confidence maps using random walks,” Med. Image Anal., vol. 16, no. 6, pp. 1101–1112, Aug. 2012.

[20] C. Hennerperger et al., “Towards MRI-based autonomous robotic U.S. acquisitions: A first feasibility study,” IEEE Trans. Med. Imag., vol. 36, no. 2, pp. 538–548, Feb. 2017.

[21] Z. Jiang et al., “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, Jul. 2022.

[22] C. Balakrishna, S. Dadashzadeh, and S. Soltaninejad, “Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 encoder,” 2018, arXiv:1806.07554.

[23] D. Dunbar and G. Humphreys, “A spatial data structure for fast Poisson-disk sample generation,” ACM Trans. Graph., vol. 25, no. 3, pp. 503–508, Jul. 2006.

[24] C. Holz, A. E. Ichinomiya, D. Tombari, R. B. Rusu, and S. Behnke, “Registration with the point cloud library: A modular framework for aligning in 3-D,” IEEE Robot. Autom. Mag., vol. 22, no. 4, pp. 110–124, Dec. 2015.

[25] A. Boezaart and B. Ihnatsenka, “Ultrasound: Basic understanding and learning the language,” Int. J. Shoulder Surgery, vol. 4, no. 3, pp. 55, 2010.

[26] P. Chatelain, A. Krupa, and N. Navab, “Confidence-driven control of an ultrasound probe,” IEEE Trans. Rob., vol. 33, no. 6, pp. 1410–1424, Dec. 2017.
Zhongliang Jiang (Member, IEEE) received the M.Eng. degree in mechanical engineering from the Harbin Institute of Technology, Shenzhen, China, in 2017. He is currently pursuing the Ph.D. degree in computer science with the Technical University of Munich, Munich, Germany. From January 2017 to July 2018, he was a Research Assistant with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Science (CAS), Shenzhen. His research interests include medical robotics, robotic learning, human–robot interaction, robotic ultrasound, and computer vision.

Nehil Danis received the B.Sc. degree in computer science from Hacetette University, Ankara, Turkey, in 2018, and the M.Sc. degree in computer science from the Technical University of Munich, Munich, Germany, in 2021. She is currently a Software Engineer at Brainlab AG, Munich. Her research interests include 3-D geometry processing, visual SLAM, and robotic manipulation.

Yuan Bi (Graduate Student Member, IEEE) received the B.Sc. degree in mechatronics from Tongji University, Shanghai, China, in 2017, and the M.Sc. degree in robotics, cognition, intelligence from the Technical University of Munich, Munich, Germany, in 2021, where he is currently pursuing the Ph.D. degree in computer science. His research interests include robot manipulation, robotic learning, and computer vision.

Mingchuan Zhou (Member, IEEE) received the Ph.D. degree in computer science from the Technical University of Munich, Munich, Germany, in 2020. He was a Visiting Scholar with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA, in 2019. He held a joint post-doctoral position at the Institute of Biological and Medical Imaging (IBMI), Helmholtz Center Munich, Oberschleißheim, Germany, and the Chair for Computer Aided Medical Procedures Augmented Reality (CAMP), Technical University of Munich, from 2019 to 2021. He is currently an Assistant Professor with Zhejiang University, Hangzhou, China, where he is leading multiscale robotic manipulation laboratory for agricultural robotics. His research interests include the autonomous system, agricultural robotics, medical robotics, and image processing.

Markus Kroenke studied physics and medicine at Ludwig Maximilians University, Munich, Germany, and the Technical University of Munich, Munich, Germany. He is board certified in Nuclear Medicine and Attending Physician at the Department of Radiology and Nuclear Medicine, German Heart Center Munich, Technical University of Munich. His main research interests include multimodal imaging.

Thomas Wendler received the B.Sc. degree in electrical engineering from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2004, and the M.Sc. degree in biomedical engineering and the degree in computer science from the Technical University of Munich, Munich, Germany, in 2007 and 2010, respectively. He is currently the Vice-Director of the Chair for Computer-Aided Medical Procedures and Augmented Reality (CAMP), Technical University of Munich (TUM), the CEO of ScintHealth GmbH, Munich, and an interim CTO of SurgicEye GmbH, Munich. His main interests include multimodal image analysis methods for clinical decision support into clinical applications.

Nassir Navab (Fellow, IEEE) received the Ph.D. degree in computer and automation from INRIA, Paris, France, and the University of Paris XI, Paris, in 1993. He is currently a Full Professor and the Director of the Laboratory for Computer-Aided Medical Procedures, Technical University of Munich, Munich, Germany, and an Adjunct Professor at Johns Hopkins University, Baltimore, MD, USA. He has also secondary faculty appointments with the both affiliated Medical Schools. He enjoyed two years of a Post-Doctoral Fellowship with the MIT Media Laboratory, Cambridge, MA, USA, before joining Siemens Corporate Research (SCR), Princeton, NJ, USA, in 1994. He has authored hundreds of peer-reviewed scientific articles, with more than 50 400 citations and enjoy an H-index of 104 as of August 11, 2022. He has authored more than 30 awarded papers, including 11 at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), five at the International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), and three at the IEEE International Symposium on Mixed and Augmented Reality (ISMAR). He is the Inventor of 50 granted U.S. patents and more than 50 International ones.

Dr. Navab is a fellow of the Academy of Europe, MICCAI, and Asia-Pacific Artificial Intelligence Association (AAIA). He was a Distinguished Member and was a recipient of the Siemens Inventor of the Year Award in 2001 at SCR, the SMT Society Technology Award in 2010 for the introduction of Camera Augmented Mobile C-arm and Freehand SPECT technologies, and the “10 Years Lasting Impact Award” of the IEEE ISMAR in 2015.