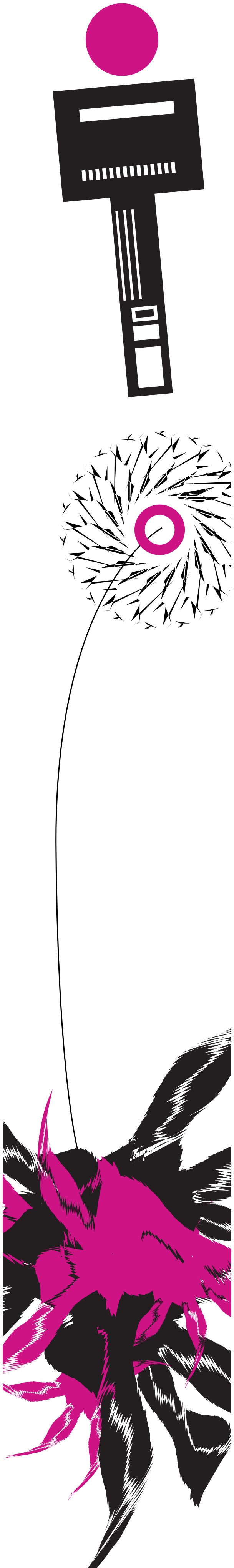

ARUS - AUTOMATIC ROBOTIC ULTRASOUND SCANNING FOR MUSCLE SEGMENTATION AND RECOGNITION

A. (Alessandro) Cappellari

MSC ASSIGNMENT

Committee:
dr. ir. M. Abayazid
dr. ir. K. Niu
prof. dr. ir. N.J.J. Verdonschot

November, 2024

076RaM2024
Robotics and Mechatronics
EEMCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Automatic Robotic Ultrasound Scanning for Muscle Segmentation and Recognition

Alessandro Cappellari Kenan Niu

***Abstract*—This work presents the development and evaluation of an Automatic Robotic Ultrasound Scanning (ARUS) system capable of real-time segmentation and 3D reconstruction of muscles and bones. The ARUS system integrates a robotic arm, ultrasound imaging, and stereo vision to achieve fully automated scanning, eliminating the need for operator expertise. The system employs hybrid position-force control, advanced calibration techniques, and a U-Net-based segmentation model to ensure precision in dynamic environments. Experimental evaluations were conducted on simple and realistic phantoms, as well as in vivo, demonstrating robust reconstruction accuracy with deviations below 1 mm under optimal conditions. Visual servoing was incorporated to enhance reconstruction quality, though its effectiveness was limited by processing delays and challenges in dynamic and non-uniform environments. Force regulation exhibited consistent performance across all experiments, with mean forces maintained within safe limits. Heat maps provided insights into force distribution, highlighting areas for system improvement. This study underscores the ARUS system's potential as a cost-effective alternative to MRI for muscle analysis, offering real-time insights into muscle dynamics during movement. Future work includes developing interactive 3D muscle reconstructions, optimizing computational efficiency, and exploring innovative control and reconstruction strategies to enhance clinical applicability.**

***Index Terms*—Franka, Automatic Ultrasound scanning, Robot control, Muscle segmentation, Real-time reconstruction**

I. Introduction

Ultrasound (US) imaging is a widely used medical diagnostic tool thanks to its non-invasiveness, real-time (RT) imaging capability, and relatively low cost, which makes it particularly effective for soft tissue visualization and examination by medical professionals [1]. The evolution of US technology now allows for RT reconstruction of 3D models [2], significantly enhancing its potential for medical applications and enabling clinicians to visualize and monitor complex anatomical structures instantly. Therefore, 3D reconstruction of dynamic and non-rigid bodies like muscles becomes feasible. Athletes, trainers, doctors, and other clinicians would benefit from having RT access to a continuous and detailed analysis of muscle behavior during movement, offering insights that were previously unattainable. The ability to perform automatic scan-based reconstruction of patient-specific models, driven by the 3D capabilities, could further revolutionize personalized medical care. RT 3D muscle modeling would enable the continuous monitoring of muscle dynamics during physical activity, improving motion analysis and providing additional diagnostic tools for sports conditioning [3].

Accurate muscle assessments and diagnostics are typically performed using Magnetic Resonance Imaging (MRI). MRI is highly effective for identifying muscle injuries, tears, and other soft-tissue abnormalities, making it a gold standard in diagnosing many musculoskeletal conditions [4]. However, this technology is expensive and requires the patient to remain perfectly still. In contrast, US imaging is significantly less expensive than MRI and provides RT imaging [5]. It is dependent on the operator's expertise and is limited by its 2D output.

Automatic US scanning with 3D reconstruction aims to solve the above-mentioned issues. First, using a reliable robot manipulator to perform the scan eliminates the need for an expert operator. Second, this technology can reconstruct the area of interest - muscles in this case- in 3D using advanced segmentation algorithms. The reconstructed area can be further processed and displayed as 3D models, supporting doctors or athletes in analyzing muscle contraction.

Researchers have extensively explored 3D image reconstruction from US imaging, particularly focusing on semi-automated robotic scanning systems. For example, Li et al. [2], along with other studies such as [6] and [7], investigate 3D spinal reconstruction using semi-automated robotic scans. Besides focusing merely on bone reconstruction, these studies implement assisted robot trajectory planning by manually defining the scanning paths. Most semi or fully-automated robotic US scanning systems either follow predefined trajectories without external sensing or awareness [3][7], or they employ external sensors like calibrated optical tracking systems to provide environmental awareness during scanning [8][9][10]. However, researchers in the field have conducted limited studies on enabling visual guidance through camera sensors mounted directly on the robot's end effector (EE) that, compared to external optical tracking systems, are less expensive and require little calibration [11][12].

In previous research, 3D reconstruction in US imaging has predominantly focused on skeletal structures, with less attention on the muscles. While automated systems for trajectory planning in bone imaging exist, adapting these methods to generate 3D muscle modeling poses new challenges due to the deformable nature of muscle tissue. Unlike bones, muscles change shape during movement, making it more difficult to develop automatic scanning systems that can account for these dynamic deformations using visual servoing and real-time sensor feedback.

This work aims to contribute to the US reconstruction field

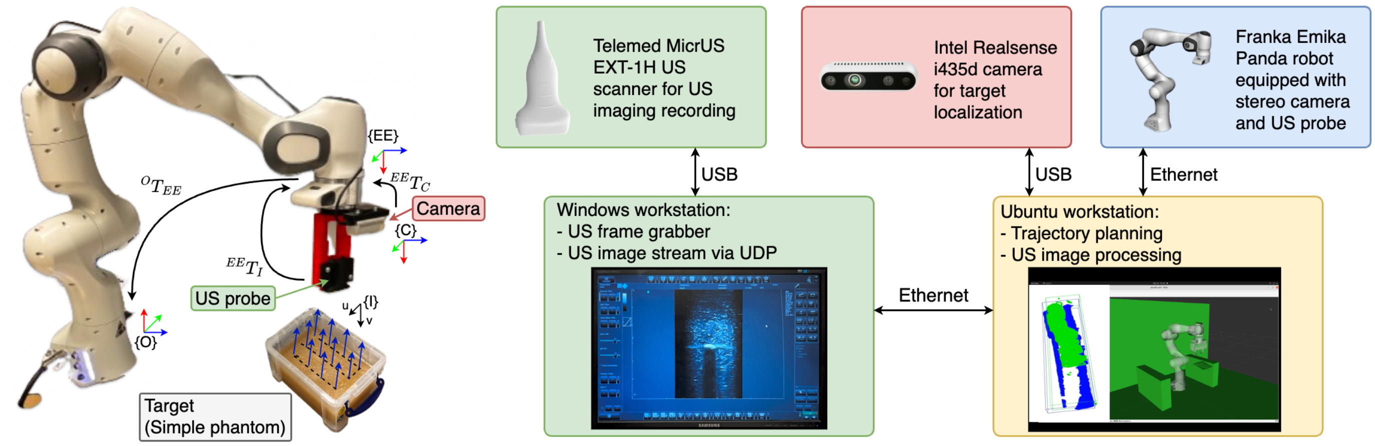

Fig. 1. System overview. On the left, Franka Emika Panda robot equipped with the Telemed MicrUS US scanner and the Intel Realsense stereo camera, with all relevant frames and tranformation matrices highlighted. On the right, the main components and the relative connections.

by introducing the development of an Automatic Robotic Ultrasound Scanning (ARUS) system that performs real-time segmentation and 3D reconstruction of muscles and bones simultaneously. This fully automated system is capable of scanning, segmenting, and reconstructing patient-specific muscle and bone models, and was tested and evaluated through various experiments. These improvements are expected to benefit medical professionals and sports scientists, opening new possibilities for muscle model generation, motion analysis, and interactive diagnostics. The integration of these technologies creates a more comprehensive and data-driven platform for enhancing muscle reconstruction and motion monitoring.

The paper is organized as follows. Section II describes all the components of the proposed system, as well as introducing the experimental evaluation. Section III presents the experimental results alongside their validations. Section IV analyses and discusses the obtained results. Finally, Section V concludes the work and proposes further research.

II. Method

A. Experimental setup

The proposed ARUS system consists of three main building blocks: a robotic arm (Franka Emika Panda, Munich, Germany), a US imaging system (Telemed MicrUs EXT-1H, Vilnius, Lithuania) and a stereo camera (Intel Realsense d435i, Santa Clara, USA).

A custom, 3D printed, holder part was mounted onto the end effector of the robot to hold the US probe and the stereo camera. The US system was set to have a clear image view of the targets both during image calibration and testing. The linear transducer was operated at 12 MHz and a depth of 6 cm utilizing TGC. A Windows 10 PC workstation (Intel Xeon, CPU @4.00GHz, 32GB RAM) was used to stream the US frames via UDP to the main workstation, a Ubuntu 20.04 PC (Intel i7-7700, CPU @3.60GHz, 16GB RAM) used for data processing and robot control. The two PCs were connected using direct Ethernet connection and the stream was at 2Hz during scan. The main components of the system are illustrated in Fig. 1.

B. Stereo camera-to-robot calibration

In stereo camera-to-robot calibration, the objective is to estimate the homogeneous transformation $^{EE}\mathbf{T}_C$, which describes how the camera (C) is attached to the robot's end-effector (EE).

The calibration starts by acquiring n images of a calibration grid (e.g., a chessboard), which are necessary for determining the camera's intrinsic parameters, along with the corresponding EE poses in base frame. This step allows the estimation of $^{C}\mathbf{T}_{obj}$, the homogeneous transformation between the camera frame and the calibration grid. Finally, $^{EE}\mathbf{T}_C$ can be estimated from the pairs $\{^{O}\mathbf{T}_{EE}, ^{C}\mathbf{T}_{obj}\}_i$, where $^{O}\mathbf{T}_{EE}$ expresses the EE position in base frame (O) and i represents the $i-th$ acquired image. The ViSP library was used to calculate the matrix by solving the $AX = XB$ problem [12][13][14].

Once the calibration is complete, it's possible to express any point from the depth camera frame in the base frame as:

$$\mathbf{p}^O = {}^{O}\mathbf{T}_{EE}\,{}^{EE}\mathbf{T}_C\mathbf{p}^C \qquad (1)$$

where $\mathbf{p}^O = [p_x^O, p_y^O, p_z^O, 1]^T$ and $\mathbf{p}^C = [p_x^C, p_y^C, p_z^C, 1]^T$.

C. Automatic robot ultrasound image calibration

To conduct ultrasound image calibration is to determine the transformation from the position of each pixel on the 2D ultrasound image relative to the 3D position of the robot's EE. Other than the homogeneous transformation $^{EE}\mathbf{T}_I$, that describes the conversion from ultrasound Image (I) frame to the EE frame, this calibration also provides the scale matrix $\mathbf{T}_s$, which converts the pixel size in meters. The calibration was performed automatically using a sphere phantom, similar to what was described in prior work [15][8]; this means estimating the two aforementioned matrices by minimizing the difference between the position of the sphere's center

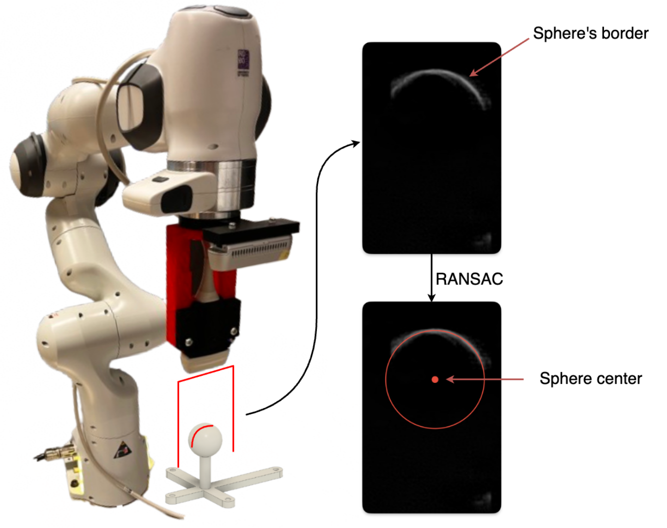

Fig. 2. During automatic robot ultrasound image calibration, hundreds of US images of the sphere phantom are recorded. Each one is processed by a RANSAC algorithm to estimate its center.

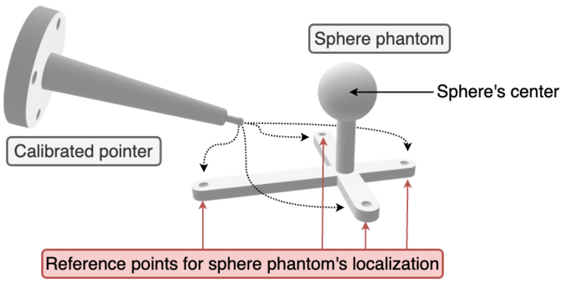

Fig. 3. In order to localize the sphere's center, a pre-calibrated pointer is used to first localize the four reference points of the phantom. These reference points to calculate the sphere's exact position in the base frame via trilateration.

expressed in base frame and its projection from Image frame to the base frame, as follows:

$$f = \min \sum_{i=1}^{n} \left| {}^{O}\mathbf{T}_{EE,i} \, {}^{EE}\mathbf{T}_{I} \mathbf{T}_{s} \mathbf{p}_{i}^{I} - \mathbf{p}^{O} \right| \tag{2}$$

Where ${}^{O}\mathbf{T}_{EE,i}$ is the transformation matrix that describes the pose of the end effector in the base frame in the i-th captured image (determined using a RANSAC fitting algorithm as shown in Fig. 2), ${}^{EE}\mathbf{T}_{I}$ and $\mathbf{T}_{s}$ are the unknown matrices, $\mathbf{p}_{i}^{I}$ is the sphere's center position in the i-th ultrasound image frame and $\mathbf{p}^{O}$ is the corresponding point in the base frame. The position of this latter was determined using trilateration by accurately measuring the positions of four known reference points on the phantom. In this paper, the proposed method of recording the positions is by touching them with a virtualized pointer previously attached to the Franka's EE and calibrated via pivot calibration (see Fig. 3) [16].

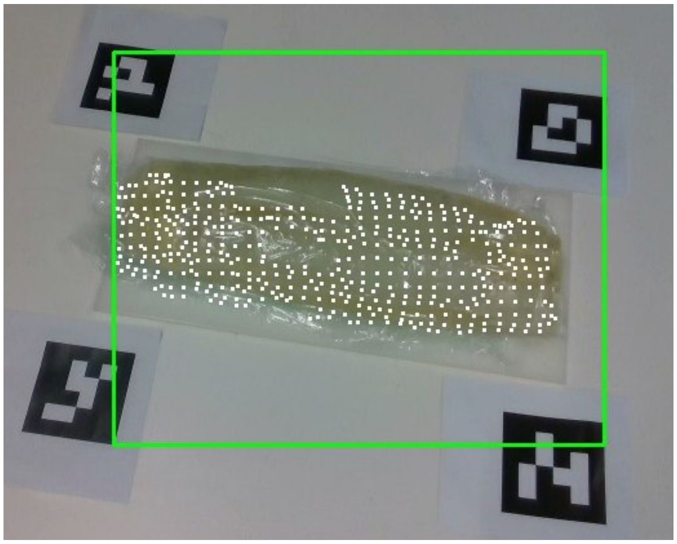

Fig. 4. Four fiducial markers define a Region of Interest (ROI) to help localize the target, in this case a forearm phantom.

After calibration, each pixel $\mathbf{p}^{I} = [u, v, 0, 1]^{T}$ can be converted to point cloud $\mathbf{p}^{O} = [p_{x}^{O}, p_{y}^{O}, p_{z}^{O}, 1]^{T}$ in robot base frame using:

$$\mathbf{p}^{O} = {}^{O}\mathbf{T}_{EE} \, {}^{EE}\mathbf{T}_{I} \mathbf{T}_{s} \mathbf{p}^{I} \tag{3}$$

Note: Ultrasound (US) and Image (I) frames aren't equal. The homogeneous transformation ${}^{EE}\mathbf{T}_{I}$ describes how each US image pixel is positioned in the EE frame, and is defined in (4), where ${}^{EE}\mathbf{R}_{I}$ is the rotation matrix and ${}^{EE}\mathbf{t}_{I} = [t_x, t_y, t_z]^{T}$ is the translation vector.

$${}^{EE}\mathbf{T}_{I} = \begin{bmatrix} {}^{EE}\mathbf{R}_{I} & {}^{EE}\mathbf{t}_{I} \\ 0 & 1 \end{bmatrix} \tag{4}$$

On the other hand, ${}^{EE}\mathbf{T}_{US}$ simply shifts the EE by ${}^{EE}\mathbf{t}_{I,z} = [0, 0, t_z]^{T}$, while keeping the two frames aligned, as follows:

$${}^{EE}\mathbf{T}_{US} = \begin{bmatrix} \mathbf{I}_{3x3} & {}^{EE}\mathbf{t}_{I,z} \\ 0 & 1 \end{bmatrix} \tag{5}$$

D. Robot scanning trajectory

The robot trajectory precisely defines the positions and orientations that the US probe must follow and is highly dependent on the initial target scan with the stereo camera. This initial step involves moving the robot toward the target so that the stereo camera can scan it using both visual and depth information. Specifically, depth information is used to create a point cloud of the camera's field of view. Then, assuming the target is placed on a flat surface, a RANSAC-based algorithm fits a plane on it and filters out any point that isn't at least 1 cm above the surface. Visual information is used to further remove outliers, reflections and noise by defining a Region of Interest (ROI) using four fiducial markers (see Fig. 4). To improve image frame rate, the point cloud is down sampled with voxel size = $0.008\mathrm{m}$. The point cloud representing the target is then converted to the robot's base

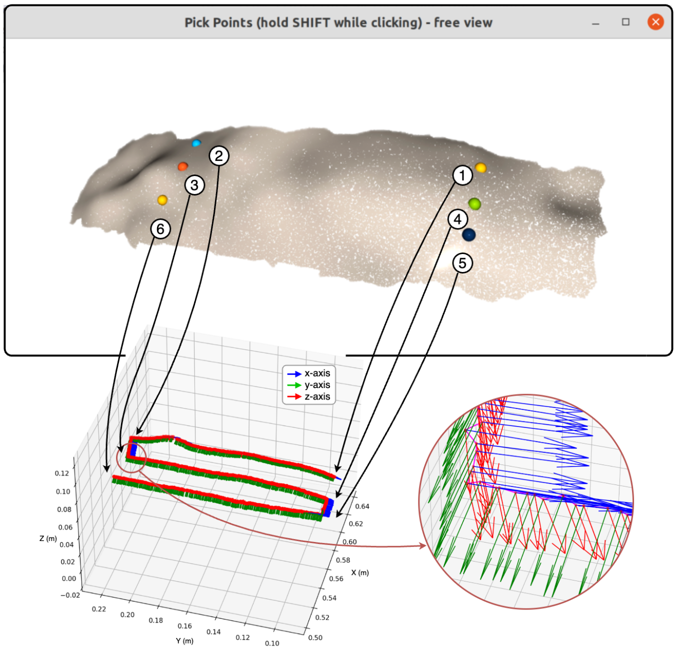

Fig. 5. Mesh reconstruction from the point cloud. A Graphical User Interface allows to select specific points from it (6 points in this example) and generate a trajectory. Each trajectory point is associated with an orientation: the blue arrows represent the x-axis, the green arrows the y-axis and red arrows represent the z-axis.

frame and a statistical outlier remover is applied to identify and remove points that deviate significantly from their neighbors based on statistical measures, in order to eliminate any outliers.

A mesh reconstruction is created from the point cloud, and a custom Graphical User Interface (GUI) allows the operator to interact with it and pinpoint markers, as visible in the top part of Fig. 5. These markers are used to generate the desired US probe's trajectory. The trajectory is a $N \times 7$ matrix, where N is the number of poses Each pose $\mathbf{p}_i^{US} = [p_x^{US}, p_y^{US}, p_z^{US}, q_x^{US}, q_y^{US}, q_z^{US}, q_w^{US}]^T$ is defined such that they are positioned 0.5mm apart and oriented perpendicular to the surface, while maintaining the yaw angle constant relative to the EE. An example is shown in the bottom-right part of Fig. 5, illustrating that each trajectory point is associated with an orientation where the z-axis (red arrow) is directed towards the target, while the x-axis (blue arrow) maintains a consistent direction throughout the trajectory.

During scanning, the trajectory poses are continuously updated and fine tuned using an offset parallel to the EE z-axis to account for the deformable nature of the skin and muscles.

E. Robot control

A precise and reliable control strategy is essential for performing high-quality US scan. To achieve this, it was implemented a hybrid position-based control, as illustrated in Fig. 6. This strategy allows for the deactivation of the force controller whenever the US probe is not directly in contact with a surface and thus enhancing the system stability.

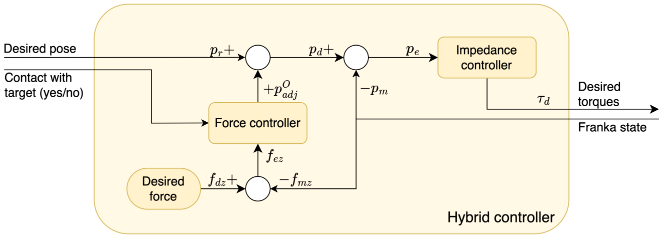

Fig. 6. Hybrid controller overview. An advantage of this controller is being able to activate or deactivate the Force controller depending on whether the US probe is in contact with the target or not.

The force controller is responsible for maintaining a constant force of $f_{dz} = 3\text{N}$ on the target for optimal scanning quality. Although the robot can only measure the forces applied to the EE, we assume the forces applied on it and the US probe are equal, as their frames are aligned. The external wrench is calculated through torque sensors positioned at joint level, using the Jacobian to map joint torques to forces and moments at the EE. Given the measured force $\mathbf{f}_{mz}$, the force error is computed as $\mathbf{f}_{ez} = \mathbf{f}_{dz} - \mathbf{f}_{mz}$.

A PI controller is used to compute the desired position adjustment $\mathbf{p}_{adj}^{EE}$ in US frame:

$$\mathbf{p}_{adj}^{EE} = \begin{bmatrix} 0 \\ 0 \\ k_p \mathbf{f}_{ez} + k_i \int \mathbf{f}_{ez} dt \end{bmatrix} \tag{6}$$

The found position adjustment is finally converted in base frame:

$$\mathbf{p}_{adj}^{O} = {}^{O}\mathbf{T}_{EE} \mathbf{p}_{adj}^{EE} \tag{7}$$

This allows to calculate the new desired pose as its sum with the reference pose $\mathbf{p}_r = [\tilde{\mathbf{p}}_r, \tilde{\mathbf{q}}_r] \in \Re^7$:

$$\mathbf{p}_d = [\tilde{\mathbf{p}}_r + \mathbf{p}_{adj}^{O}, \tilde{\mathbf{q}}_r] \tag{8}$$

The Cartesian impedance controller is responsible for calculating the necessary joint torques based on the current pose error. The total desired joint torque is computed using a PD controller as:

$$\boldsymbol{\tau}_d = \boldsymbol{\tau}_{\text{task}} + \boldsymbol{\tau}_{\text{nullspace}} + \boldsymbol{\tau}_{\text{coriolis}}$$

The task-space torque is calculated as:

$$\boldsymbol{\tau}_{\text{task}} = \mathbf{J}^\top \left(-\mathbf{K}_{\text{cart}} \mathbf{p}_e - \mathbf{D}_{\text{cart}} \mathbf{J} \dot{\mathbf{q}}\right)$$

Here, $\mathbf{J}$ is the Jacobian matrix, $\mathbf{K}_{\text{cart}}$ is the Cartesian stiffness matrix, $\mathbf{D}_{\text{cart}}$ is the Cartesian damping matrix, and $\mathbf{p}_e = [\tilde{\mathbf{p}}_e, \tilde{\theta}_e] \in \Re^6$ is the pose error, where the $\tilde{\mathbf{p}}_e$ is the difference between desired position $\tilde{\mathbf{p}}_d$ and measured one $\tilde{\mathbf{p}}_m$, while $\tilde{\theta}_e$ is composed by the $[x, y, z]$ components of the quaternion operation $\tilde{\mathbf{q}}_d^{-1} \tilde{\mathbf{q}}_m$. In this equation the mass term is not present as it's managed by Franka's low-level controller. Moreover, it's also assumed that the damping ratio $\zeta = 1$, hence we assume that cartesian stiffness and cartesian damping are proportional.

The nullspace torque is used to control joint motion without affecting task space and is calculated as:

$$\boldsymbol{\tau}_{\text{nullspace}} = \left(\mathbf{I} - \mathbf{J}^\top \mathbf{J}^+\right) \left(\mathbf{K}_{\text{null}} (\mathbf{q}_d^{\text{null}} - \mathbf{q}) - 2\sqrt{\mathbf{K}_{\text{null}}} \dot{\mathbf{q}}\right)$$

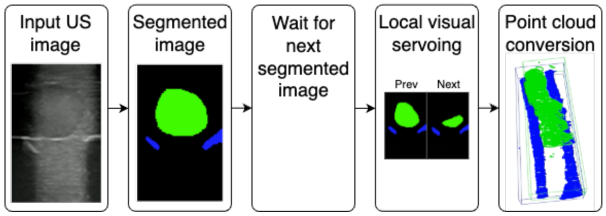

Fig. 7. Image segmentation and reconstruction pipeline.

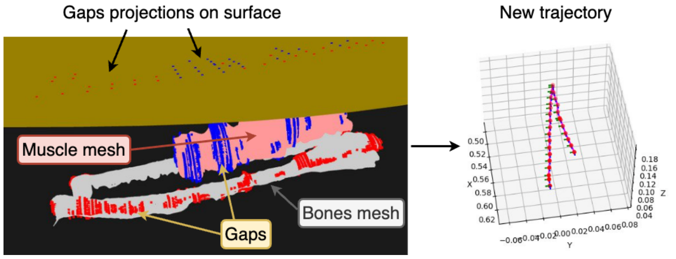

Fig. 8. (Left) Depiction of the global Visual Servoing (VS) algorithm: the mesh derived from the bone class reconstruction is shown in gray, with the corresponding gaps highlighted in red. The mesh from the muscle layer reconstruction is shown in pink, with the corresponding gap points highlighted in blue. The phantom surface is depicted in sand color, with red and blue dots representing the projections of the gap points. (Right) The output trajectory generated by the algorithm.

Here, $\mathbf{q}_d^{\text{null}}$ is the desired nullspace configuration, $\mathbf{J}^+$ is the pseudoinverse of the Jacobian, and $\mathbf{K}_{\text{null}}$ is the nullspace stiffness matrix.

$\boldsymbol{\tau}_{\text{coriolis}}$, the Jacobian $\mathbf{J}$, joint positions $\mathbf{q}$, desired nullspace configuration $\mathbf{q}_d^{\text{null}}$, and joint velocities $\dot{\mathbf{q}}$ are obtained from the Franka robot's state during each control loop iteration.

The controller parameters were fine tuned to achieve low steady-state error while maintaining the robot relatively compliant. Safety mechanisms are implemented: if a force $f_{mz} >$ 6N is measured, the robot stops momentarily and reduces the force before continuing, whereas if a force $f_{mz} >$ 15N is measured, the scan is aborted and the robot goes back to the initial pose.

F. Image segmentation and 3D reconstruction

In order to achieve accurate 3D reconstructions of the muscles, it's critical to firstly perform accurate image segmentation. The U-Net [17] is widely used for biomedical image segmentation due its ability to efficiently extract both high-level and detailed spatial features with superior performance when compared to traditional image processing, making it particularly effective for identifying soft tissue structures in ultrasound data. Two U-Net models were trained for the phantom and in-vivo experiments. Both have a depth of 4 and with a number of 64 starting filters. The input images have a resolution of 128x128 pixels and manually-labeled augmented images were used for training.

After segmenting the image frame, all the detected pixels $\mathbf{p}^I = [u, v, 0, 1]^T$ are converted to point clouds $\mathbf{p}^O$ in robot base frame using (3).

Finally, surface meshes are rendered and visualized in a custom GUI that constantly updates with the most recent data and allows to toggle the visibility of each segmented class.

G. Visual servoing feedback

Two visual servoing (VS) based algorithms were also developed to enhance the quality of US scans, one based on local performance, and the other based on the global one.

The first one takes a segmented image I_k as input and determines their validity based on two criteria. First, it waits until another frame is received, and determines the pixel-wise union between the images I_{k-1} and I_{k+1}:

$$I_{k,union} = I_{k-1} \cup I_{k+1} \tag{9}$$

Then, it performs the average surface distance between $I_{k,union}$ and I_k. This latter operation is computed for each segmented class, and each one is either labelled as acceptable or not based on how low the distance is, as shown in Fig. 7. Second, it checks whether the force applied to the target is within a tolerance range of $f_{dz} \pm 2\text{N} = [1, 5]$ N. If either one of these criteria is not satisfied, the corresponding US pose is fed back into the trajectory planning script and will re-scanned once the original trajectory is completed.

The second algorithm waits until the initial trajectory is completed, then evaluates the overall quality of the 3D reconstruction by inspecting the presence of any gaps or missing areas in the scan. This is done by converting the output point cloud into a mesh, which, by definition, represents a continuous surface. The two are then compared. Areas with low point densities are detected as holes, and the algorithm generates a new trajectory for the robot to follow, specifically targeting the unscanned or poorly scanned regions. The robot then re-scans those parts, aiming to enhance the completeness and quality of the overall 3D reconstruction.

H. Ultrasound image transmission

The US images, captured from the Windows workstation, are streamed via User Datagram Protocol (UDP), a communication protocol that sends data without establishing a persistent connection between devices, prioritizing transmission speed. Each frame is converted into pixel arrays and split into chunks with an 8-byte header for identification, allowing proper reconstruction on the receiver side.

The main workstation and the Franka robot use ROS to communicate internally. The main nodes are highlighted in yellow Fig. 9. The $Pose\ publisher$ node is responsible for publishing the next desired pose; each pose is coupled with a boolean topic that describes if the robot must enable or not the force sensor, and is read by the controller to precisely reach the target poses with the desired force. Concurrently, the $Image\ publisher$ node, that listens for UDP packages, reconstructs the US images and publishes them paired with the current probe's pose and force applied to the target. Then, $Image\ segmentation$ applies the pre-trained UNet model to segment the image and publishes the segmented image together with the same pose and force. $Local\ visual\ servoing$

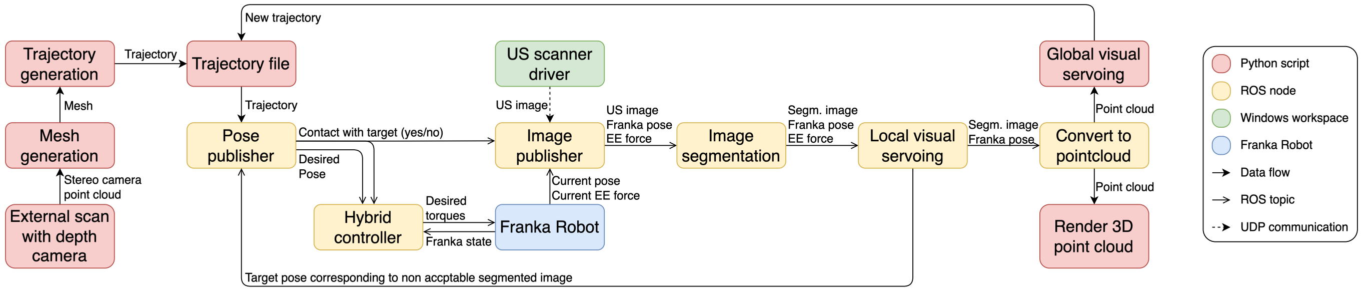

Fig. 9. Flow chart of the system's code blocks. Yellow blocks indicate ROS nodes. Red blocks indicate python scripts. The green block distinguishes the secondary workstation and the blue one indicates the Franka robot.

then verifies the quality of the single segmented images and every acceptable frame is finally converted to point cloud via $Convert\ to\ pointcloud$ node. To visualize the RT reconstruction, the $3D\ reconstruction$ script creates a render that shows the current generated point cloud and allows to toggle the visibility of the classes.

I. Evaluation metrics

The objective of evaluating the ARUS system is to assess its performance in accurately reconstructing the bones and muscles of the target, while being safe for the patient. Each selected evaluation metric directly contributes to enhancing the system's ability to achieve this goal, ensuring that every component functions collaboratively to produce precise and reliable reconstructions. The evaluation process is the following:

1) 3D Reconstruction Quality: Good reconstruction quality is the main goal of the entire system and is directly related to the accuracy of the following metrics:
   1.1 3D Reconstruction Accuracy: Clearly, reconstruction accuracy is the factor that most of all impacts the reconstruction quality. To validate it, the US reconstruction was compared to the corresponding MRI, used as ground truth.
   1.2 Calibration: Calibration accuracy was measured for both stereo camera-to-robot and ultrasound image, which are essential for precise 3D reconstructions.
   1.3 Segmentation: The U-Net model's segmentation performance for muscle and bone regions was evaluated using the Dice coefficient and Intersection over Union (IoU) metrics.
   1.4 Control: The hybrid controller's robustness in following the specified trajectory and maintaining a constant normal force during scanning was evaluated, with accuracy measured by Root Mean Square Error (RMSE), average and standard deviation.
   1.5 Visual Servoing: Visual servoing feedback was tested by scanning the target in four different configurations: (1) without VS feedback, (2) with local VS image feedback, (3) with post-scan feedback, and (4) using both feedback types simultaneously.
2) Safety: Ensuring operational safety during scanning is critical for preventing damage to both the subject and the equipment. Monitoring and controlling the applied force and contact areas were essential aspects of the system's safety protocol, aimed at guaranteeing consistent performance within safe limits.
   2.1 Force: The applied force during scans was monitored to ensure consistent levels within safe limits for each target.
   2.2 Force spatial distribution - Force Heat Map: The heat map combines force and accuracy data to identify potential correlations between them. Each point is represented as a horizontal line matching the width of the ultrasound probe, making it easier to detect any areas that were not scanned.
3) Experimental Evaluation: The evaluation process involved testing on different targets, visible in Fig. 10, each providing insight into the system's capabilities and areas for potential improvement.
   3.1 Simple Phantom: An initial feasibility analysis was conducted using a simple, rectangular gelatine phantom that incorporated PVA components to simulate a muscle layer (Fig. 10, left) and PLA components to represent bones.
   3.2 Realistic Phantom: Further testing was performed on a more anatomically realistic gelatine phantom designed to mimic a human forearm, allowing for validation of system performance in a more lifelike scenario. This phantom included four distinct PVA muscle layers (three of which are visible in Fig. 10, center) as well as PLA parts representing the forearm bones.
   3.3 In-vivo: Final evaluations were carried out in vivo on a human subject, offering insights into the system's effectiveness in real-world conditions (Fig. 10, right).

Detailed results from these experiments are presented in the next sections, organized as follows: results from the simple phantom are discussed first, followed by those from the complex phantom, and finally the in-vivo experiments.

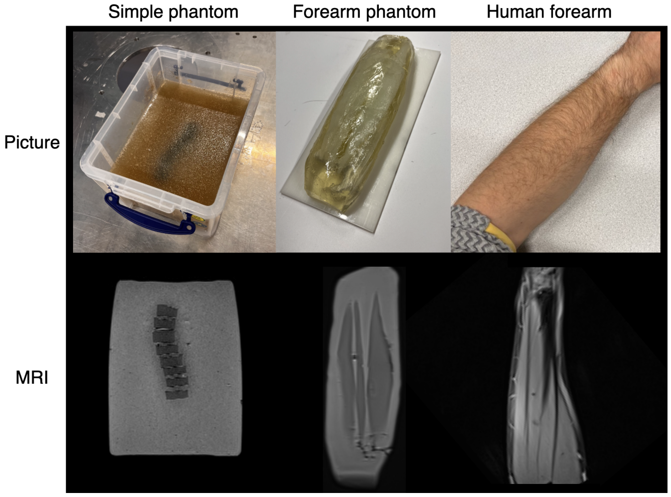

Fig. 10. MRI image comparison between targets. From left to right: simple phantom, complex phantom, and real human forearm. Both phantoms consist of a gelatine base with embedded PVA components to simulate muscle layers (visible in the image) and 3D-printed PLA pieces representing the ulna and radius bones.

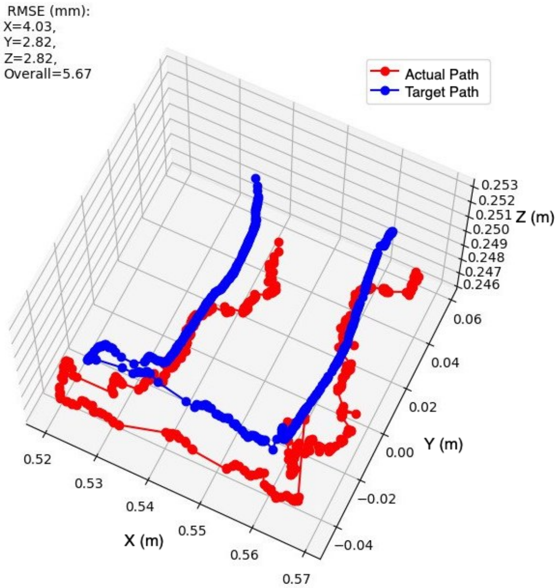

Fig. 11. Controller position accuracy performance. In blue, the target path to follow, while in red the actual one. The RMSE is 5.67 mm.

III. Results

A. Simple phantom results

1.1) 3D Reconstruction Accuracy: The system's 3D reconstruction output was quantitatively compared to the MRI ground truth (see Fig. 13), achieving a RMSE of 1.22 mm, a mean error of 0.94 mm, and a standard deviation of 0.77 mm.

1.2) Calibration: Stereo camera-to-robot calibration was validated by developing a script to detect a visual marker using the stereo camera and automatically move the robot's EE to that position. The calibration's accuracy was assessed by visually inspecting the distance between the two. Automatic US calibration, on the other hand, was validated by estimating the sphere phantom's position using (3) for each US image. This led to a RMSE of 2.75 mm and standard deviation of 1.04 mm.

1.3) Segmentation: The U-Net model was trained for 200 epochs on 200 augmented images with one muscle class and one bone class (visible depicted in green and blue in Fig. 7). The model performed well, returning a Dice coefficient of 0.85 and an IoU score of 0.84.

1.4) Control: Several tests were performed on the phantom. Depicted in Fig. 11 it's possible to see an example of tracking performance. In this case, the RMSE was 5.67 mm.

1.5) Visual Servoing: VS experiments are presented in Table I. The impact of each VS feedback configuration on position accuracy, force regulation, reconstruction quality, and execution time is shown. Position RMSE was $3.35\,\mathrm{mm}$ without feedback, $3.10\,\mathrm{mm}$ with local VS feedback, $3.28\,\mathrm{mm}$ with global VS feedback, and $3.25\,\mathrm{mm}$ with combined feedback. Standard deviations remained low. Average force applied was similar across configurations, with the highest at $2.654\,\mathrm{N}$ and the lowest at $2.543\,\mathrm{N}$. Force RMSE was $0.7183\,\mathrm{N}$ to $0.7264\,\mathrm{N}$ and the standard deviation was $0.7121\,\mathrm{N}$ to $0.7174\,\mathrm{N}$. Reconstruction RMSE was $1.22\,\mathrm{mm}$ without feedback, $1.20\,\mathrm{mm}$ with local and global feedback, and $1.21\,\mathrm{mm}$ with combined feedback. Standard deviations were consistent. Execution time varied significantly, with the no-feedback configuration taking the shortest at 295 seconds, followed by local VS feedback at 394 seconds, global VS feedback at 425 seconds, and combined feedback at 549 seconds.

2.1) Force: In the experiment shown in Fig. 12, the applied force demonstrated consistent performance with an RMSE of 0.2485 N, an MRE of 0.0894, a mean force of 2.9582 N, and a standard deviation of 0.2450 N.

2.2) Heat Map: The heat map in Fig. 12 illustrates the distribution of applied force, highlighting areas of consistent scanning and potential correlations between force stability and scan coverage.

B. Realistic Phantom Results

1.1) 3D Reconstruction Accuracy: The system's 3D reconstruction output for the realistic phantom was quantitatively assessed, showing an RMSE of 3.71 mm, a mean error of 2.57 mm, and a standard deviation of 2.68 mm (see Fig. 14)

1.2) Calibration: Stereo camera-to-robot calibration and automatic US calibration followed the same procedures as described for the simple phantom.

1.3) Segmentation: The U-Net model, trained on 50 augmented images with 5 classes, performed segmentation with an overall Dice coefficient of 0.85 and an IoU score of 0.92 (more detailed results are visible in Table II).

1.4) Control: Position tracking tests on the realistic phantom are presented in Fig. 15. The tracking accuracy, measured by RMSE, was 6.20 mm.

TABLE I
EVALUATION OF EVERY VISUAL SERVOING FEEDBACK CONFIGURATION USING THE SIMPLE PHANTOM TARGET.

Configuration	Position (mm)		Force (N)			Reconstruction (mm)		Execution time (s)
	RMSE	Std	Avg	RMSE	Std	RMSE	Std	
No Feedback	3.35	1.23	2.654	0.7219	0.7174	1.22	0.77	295
Local VS Feedback	3.10	2.34	2.543	0.7219	0.7174	1.22	0.77	394
Global VS Feedback	3.28	3.15	2.610	0.7264	0.7121	1.20	0.78	425
Both Feedbacks	3.25	2.95	2.635	0.7183	0.7159	1.21	0.76	549

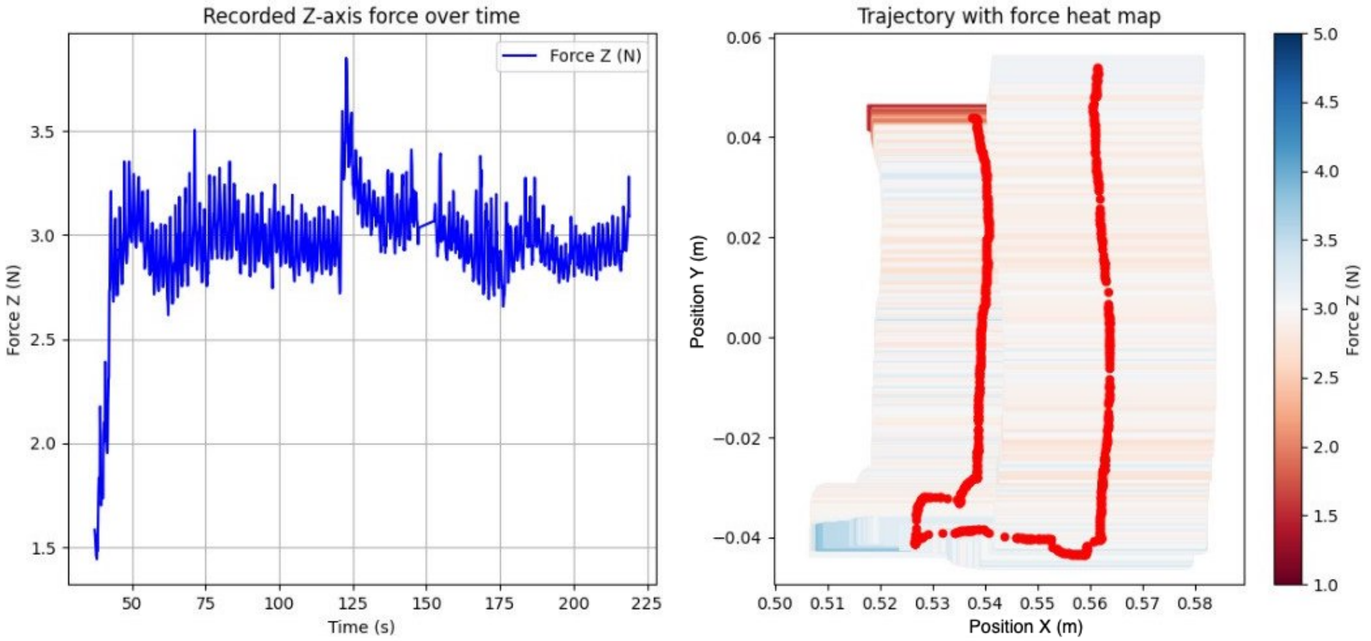

Fig. 12. Simple phantom force results. (Left) Force plot showing consistent performance with an RMSE of 0.3561 N, an MRE of 0.0894, a mean force of 3.0120 N, and a standard deviation of 0.3559 N. (Right) Heat map illustrating the applied force distribution, with each point represented as a horizontal line matching the ultrasound probe's width to highlight scanned and unscanned areas.

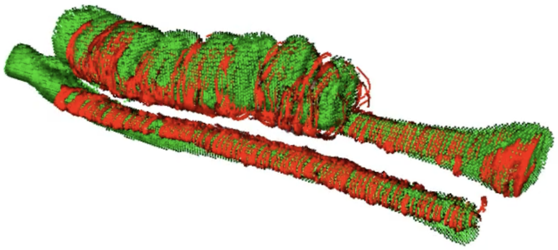

Fig. 13. The comparison between MRI (green) and US reconstruction (red) demonstrates good alignment and overlap between the two visual representations. Quantitatively, the RMSE is 1.22 mm, with a mean error of 0.94 mm and a standard deviation of 0.77 mm.

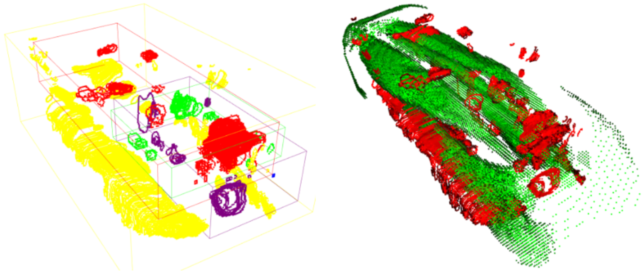

Fig. 14. (Left) 3D reconstruction of realistic phantom where each color is a different class. (Right) Comparison between MRI, in green, and US, in red. The RMSE is 3.71 mm, the mean error is 2.57 mm and the standard deviation is 2.68 mm.

TABLE II
U-NET MODEL TRAINING RESULTS FOR REALISTIC GELATINE PHANTOM EXPERIMENTATION.

Class	Dice Score	IoU
Ulna and radius (class 0)	0.71	0.84
Extensor carpi ulnaris (class 1)	0.92	0.96
Extensor digit minimi (class 2)	0.89	0.95
Extensor digitorum (class 3)	0.84	0.92
Extensor carpi radialis brevis (class 4)	0.88	0.93
Overall	0.85	0.92

2.1) Force: Force control results are illustrated in Fig. 16. The RMSE of the applied force was 0.5611 N, with a mean force of 3.1101 N and a standard deviation of 0.5502 N.

2.2) Heat Map: The heat map in Fig. 16 visualizes the force distribution across the realistic phantom. It reveals areas of consistent scanning and highlights challenges in maintaining uniform force on more detailed surfaces. This map provides valuable insights for further improving the force controller.

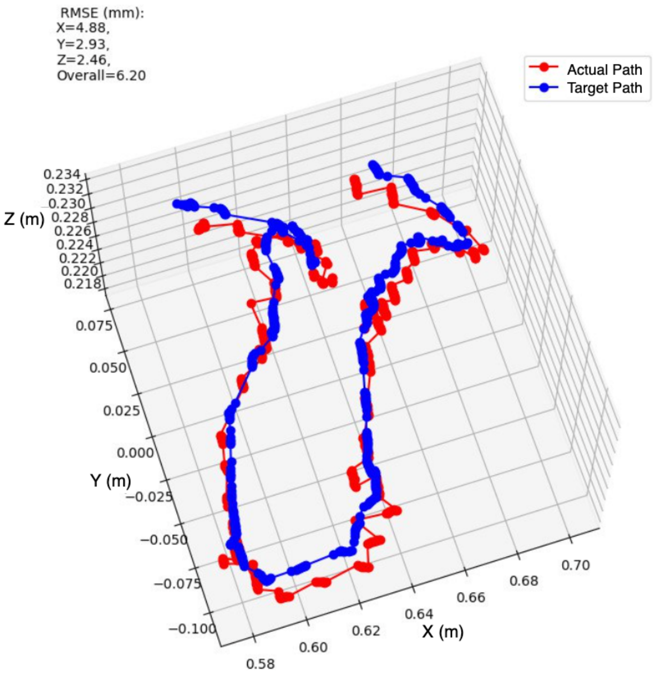

Fig. 15. Controller position accuracy performance on the realistic phantom. The target path (blue) and actual path (red) are shown. RMSE was 6.20 mm.

TABLE III
U-NET MODEL TRAINING RESULTS FOR IN-VIVO SEGMENTATION.

Class	Dice Score	IoU
Class 0 (bones)	0.79	0.82
Class 1 (muscle layer 1)	0.71	0.88
Class 2 (muscle layer 2)	0.83	0.89
Class 3 (muscle layer 3)	0.86	0.90
Overall	0.80	0.87

C. In-Vivo Results

1.1) 3D Reconstruction Accuracy: Due to challenges in accurately localizing features in the MRI caused by its quality, the comparison between the ultrasound reconstruction and the MRI is primarily qualitative rather than quantitative. The ultrasound reconstruction measures 189.7 mm in length and 38.3 mm in width, while the MRI dimensions are 190.7 mm in length and 40.9 mm in width.

1.2) Calibration: The calibration process for in-vivo testing was identical to that of the simple and realistic phantoms.

1.3) Segmentation: The U-Net model's segmentation performance was tested on real muscle and bone structures. As summarized in Table III, the model achieved an overall Dice score of 0.80 and an IoU of 0.87.

1.4) Control: In-vivo position tracking tests demonstrated an RMSE of 7.03 mm. Fig. 18 illustrates the tracking performance.

2.1) Force: Force control results during in-vivo testing are presented in Fig. 19. The RMSE of the applied force was 0.475 N, with a mean force of 2.937 N and a standard deviation of 0.471 N.

2.2) Heat Map: The force heat map in Fig. 19 illustrates the distribution of applied force across the in-vivo target.

IV. Discussion

The results of this study demonstrate that the proposed ARUS system successfully automates the ultrasound scanning process while generating real-time 3D reconstructions of both muscle and bone tissues. In comparison to existing semi-automated systems for spinal reconstruction [2], our fully automated approach offers several improvements, particularly in its ability to adapt to dynamic muscle deformations. Unlike bone structures, muscles present unique challenges due to their non-rigid nature, requiring precise real-time adjustments during scanning. While previous methods rely on predefined trajectories [6], our system integrates visual servoing and real-time feedback, enabling dynamic adaptation to muscle movement, thus improving the accuracy and consistency of the generated models.

A. 3D reconstruction quality

The experimental evaluation demonstrates the ARUS system's robustness and accuracy in muscle reconstruction. Specifically, the accuracy of the reconstructed models closely aligned with ground truth MRI data, with average deviations below $1\mathrm{mm}$ when performing in optimal conditions, thus confirming the system's potential as a reliable alternative to MRI for muscle assessments.

While both local and global VS were integrated into the ARUS system to enhance the adaptability and precision of the scanning process, their contributions proved to be less effective than initially anticipated. Local VS was expected to facilitate precise adjustments when tracking fine muscle contours; however, its responsiveness was limited by the system's processing frequency, leading to minor but noticeable lags in high-speed muscle motions. Similarly, global VS, which aimed to provide robust spatial awareness of the entire scanning area, faced challenges in effectively targeting low-density point cloud areas. This limitation reduced the effectiveness of global adjustments, and as a result, the anticipated improvements in scan accuracy were not fully realized. The results in Table I highlight these performance gaps. Position RMSE values showed marginal improvements with local VS feedback compared to no feedback, but global and combined feedback configurations resulted in only slight differences. Similarly, force regulation performance was consistent across configurations. Reconstruction accuracy remained largely unaffected by the VS configurations, with RMSE values staying between 1.20 mm and 1.22 mm. One notable drawback of incorporating VS was the increase in execution time. While the no-feedback configuration completed scans in 295 seconds, the local, global, and combined feedback configurations required significantly longer durations. In real-life applications, maintaining a good trade-off between scanning time and accuracy is crucial, as prolonged durations increase the likelihood of patient movement, which can compromise the accuracy and reliability of the reconstruction.

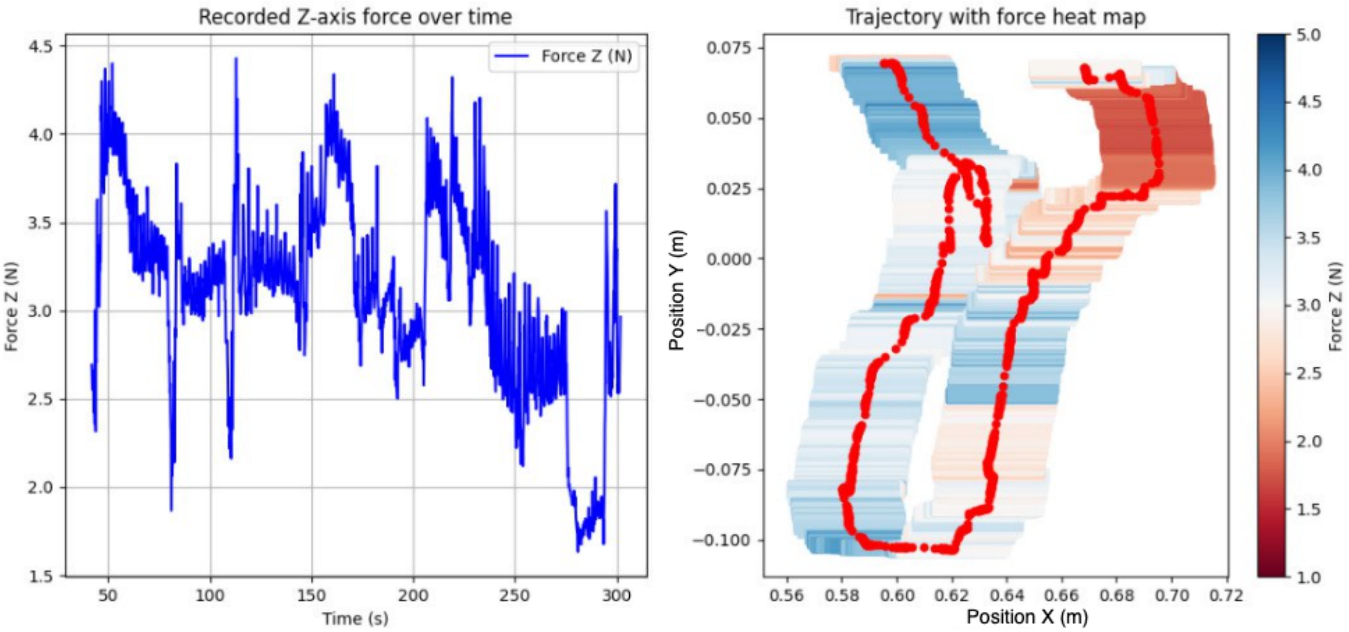

Fig. 16. Complex phantom force results. (Left) Force plot showing consistent performance with an RMSE of 0.5611 N, a mean force of 3.1101 N, and a standard deviation of 0.5502 N. (Right) Heat map illustrating the applied force distribution across the realistic phantom, highlighting scanned and unscanned areas.

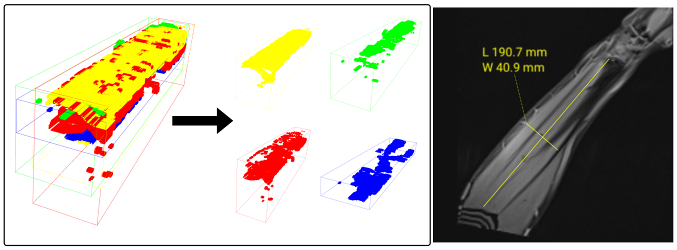

Fig. 17. (Left) In-vivo 3D ultrasound reconstruction of the subject's left forearm. For clarity, each class is displayed individually. The reconstructed point cloud measures 189.7 mm in length and 38.3 mm in width. (Right) MRI of the same forearm, with dimensions of 190.7 mm in length and 40.9 mm in width.

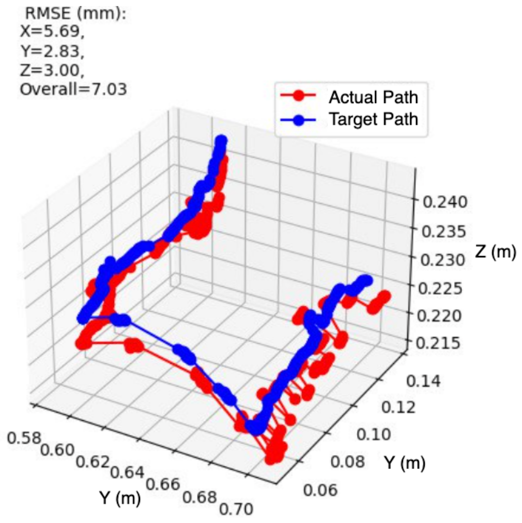

Fig. 18. Position accuracy during in-vivo tests. The blue line indicates the target path, while the red line shows the actual path followed by the end-effector. RMSE was 7.03 mm.

These findings suggest that while visual servoing remains a valuable asset in controlled settings, further refinement is needed to fully harness its potential in dynamic real-time scanning applications, particularly for improving speed and robustness without compromising accuracy.

B. Safety

Across all tests, the force regulation exhibited precise and consistent performance. For the simple phantom, the RMSE of 0.2485 N and a mean force of 2.9582 N suggest that the system maintains stable contact with the phantom surface, minimizing risks of excessive force that could lead to tissue deformation or damage in real-world scenarios. Similarly, the realistic phantom and in-vivo tests showed slightly increased force variability (RMSE 0.5611 N and 0.471 N respectively), reflecting the added complexity of maintaining stability on anatomically detailed surfaces. Nonetheless, the mean force remained within a safe range (3.1101 N and 2.937 N respectively), indicating the robustness of the force controller under varying conditions. This is confirmed by force peaks never exceeding 5 N across all experiments.

The heat maps provided a visual assessment of force distribution, revealing areas of consistent and inconsistent

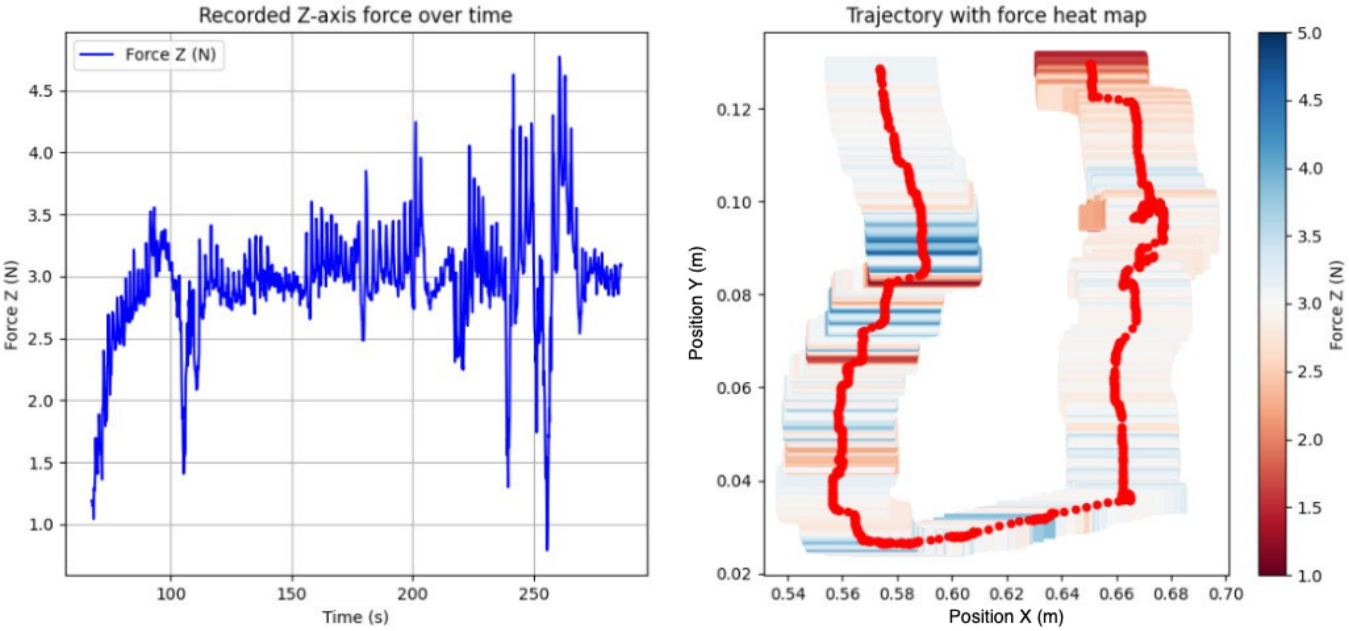

Fig. 19. In-vivo force results. (Left) Force plot showing consistent regulation with an RMSE of 0.475 N, a mean force of 2.937 N, and a standard deviation of 0.471 N. (Right) Heat map of applied force during in-vivo testing, showing regions of consistent force application and areas requiring improvement.

scanning. For the simple phantom (Fig. 12), the heat map confirmed uniform coverage, ensuring safety by avoiding areas of excessive pressure or skipped regions. However, the realistic phantom map (Fig. 16) and especially the in-vivo one (Fig. 19) highlighted challenges in maintaining even force across complex geometries, with unscanned areas indicating potential risks of incomplete coverage. These insights are critical for refining the system to ensure both safety and effectiveness in clinical applications.

Overall, the system demonstrated a strong capacity to regulate force safely and consistently, even under complex conditions, while the heat maps provided valuable feedback for improving scanning quality and trajectory planning.

Despite the promising results, there were also challenges associated with measuring the external forces during the scanning process. In this setup, force estimation was based on joint torque sensors rather than a dedicated external F/T sensor. This approach requires precise payload calibration, which proved difficult to achieve accurately, potentially affecting the force measurement's fidelity. Moreover, calculating external forces from joint torques involves inverting the robot's Jacobian matrix:

$$\mathbf{F}_{\text{ext}} = \mathbf{J}^{-T}(\mathbf{q})\ \tau_{\text{measured}} \tag{10}$$

where $\mathbf{J}$ is the Jacobian matrix, $\mathbf{q}$ represents joint positions, and τ_{measured} denotes the measured joint torques. This inversion process is computationally expensive, with a maximum achievable frequency of 30 Hz, which limits the responsiveness of force feedback. Additionally, force estimation becomes unreliable in particularly stretched positions. In such configurations, retrieving external force data is either unfeasible or an offset in force measurements is observed.

V. Conclusions

The integration of 3D reconstruction and segmentation in a fully automated system represents a significant advancement in ultrasound imaging technology. In the context of clinical applications, the ARUS system provides medical professionals with a reliable tool for assessing muscle conditions in real time, potentially reducing reliance on more expensive and time-consuming imaging modalities, such as MRI. Its ability to monitor muscle behavior continuously during movement could lead to innovative diagnostic tools for conditions like muscle tears, strains, or performance issues in athletes, offering an unprecedented level of insight into muscle dynamics.

Despite its success in achieving real-time segmentation and reconstruction, the ARUS system has areas for improvement. While visual servoing was incorporated to enhance adaptive scanning control, it proved less effective than anticipated in both local and global modes. These limitations highlight the need for more robust visual servoing approaches in challenging, real-world clinical settings.

A. Future Work

Future work on the ARUS system could focus on several exciting directions. One promising avenue is the development of interactive 3D muscle reconstructions based on real-time scans. Such reconstructions could also incorporate scans of muscles in different configurations, such as contracted and relaxed states, offering a dynamic view of muscle behavior and potentially enhancing diagnostic capabilities.

Another critical area for improvement is the system's computational efficiency. A lighter U-Net model, for instance, with a reduced depth or fewer initial filters (e.g., depth of 3 instead of 4 or 32 initial filters instead of 64), could be trained to improve the frame rate and, consequently, the reaction time of the local visual servoing. This enhancement could enable faster and more precise adjustments during dynamic scanning scenarios.

In addition, novel control strategies could be explored to optimize the force applied during scanning. For instance, adaptive force modulation based on image quality or the depth of the target muscle could improve both the safety and the accuracy of the scanning process.

Finally, further innovation in reconstruction techniques is essential to enhance the overall quality of the generated models. New strategies leveraging advanced image processing algorithms, machine learning, or physics-based modeling could address existing challenges, such as low-density areas in the point clouds or inconsistencies in the reconstructed surfaces.

By addressing these areas, the ARUS system could become an even more powerful tool for clinical and research applications, pushing the boundaries of what is possible in real-time ultrasound imaging and analysis.

References

[1] Douglas L Miller et al. “Overview of therapeutic ultrasound applications and safety considerations”. In: *Journal of Ultrasound in Medicine* 31.4 (2012), pp. 623–634. DOI: 10.7863/jum.2012.31.4.623.

[2] Ruixuan Li et al. “Automatic Robotic Scanning for real-time 3D Ultrasound Reconstruction in Spine Surgery”. In: *11th Conference on New Technologies for Computer and Robot Assisted Surgery*. Naples, Italy, Apr. 25, 2022. URL: https://lirias.kuleuven.be/retrieve/674549.

[3] Ayoob Davoodi et al. “A Comparative Study for Control of Semi-Automatic Robotic-assisted Ultrasound System in Spine Surgery”. In: *21st International Conference on Advanced Robotics (ICAR)* (2023).

[4] M.C. Erlandson et al. “Muscle analysis using pQCT, DXA and MRI”. In: *European Journal of Radiology* 85.8 (2016), pp. 1505–1511. ISSN: 0720-048X. DOI: https://doi.org/10.1016/j.ejrad.2016.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0720048X16300742.

[5] Jon A. Jacobson. “Musculoskeletal Ultrasound: Focused Impact on MRI”. In: *American Journal of Roentgenology* 193.3 (2009), pp. 619–627. DOI: 10.2214/AJR.09.2841. eprint: https://doi.org/10.2214/AJR.09.2841. URL: https://doi.org/10.2214/AJR.09.2841.

[6] Maria Victorova, David Navarro-Alarcon, and Yong-Ping Zheng. “3D Ultrasound Imaging of Scoliosis with Force-Sensitive Robotic Scanning”. In: *2019 Third IEEE International Conference on Robotic Computing (IRC)*. 2019, pp. 262–265. DOI: 10.1109/IRC.2019.00049.

[7] Ruixuan Li et al. “Robot-assisted ultrasound reconstruction for spine surgery: from bench-top to pre-clinical study”. In: *International Journal of Computer Assisted Radiology and Surgery* 18.9 (Sept. 2023), pp. 1613–1623. DOI: 10.1007/s11548-023-02932-z.

[8] Ruixuan Li, Kenan Niu, and Emmanuel Vander Poorten. “A Framework for Fast Automatic Robot Ultrasound Calibration”. In: *2021 International Symposium on Medical Robotics (ISMR)*. 2021, pp. 1–7. DOI: 10.1109/ISMR48346.2021.9661495.

[9] Ruixuan Li et al. “Comparative Quantitative Analysis of Robotic Ultrasound Image Calibration Methods”. In: *2021 20th International Conference on Advanced Robotics (ICAR)*. 2021, pp. 511–516. DOI: 10.1109/ICAR53236.2021.9659341.

[10] Yuyu Cai et al. “Development of Robot-assisted Ultrasound System for Fetoscopic Tracking in Twin to Twin Transfusion Syndrome Surgery”. In: *2023 International Symposium on Medical Robotics (ISMR)*. 2023, pp. 1–7. DOI: 10.1109/ISMR57123.2023.10130208.

[11] Dianye Huang et al. “Robot-Assisted Deep Venous Thrombosis Ultrasound Examination Using Virtual Fixture”. In: *IEEE Transactions on Automation Science and Engineering* (2024), pp. 1–12. DOI: 10.1109/TASE.2024.3351076.

[12] Ikenna Enebuse et al. “A Comparative Review of Hand-Eye Calibration Techniques for Vision Guided Robots”. In: *IEEE Access* 9 (2021), pp. 113143–113155. DOI: 10.1109/ACCESS.2021.3104514.

[13] E. Marchand, F. Spindler, and F. Chaumette. “ViSP for visual servoing: a generic software platform with a wide class of robot control skills”. In: *IEEE Robotics and Automation Magazine* 12.4 (Dec. 2005), pp. 40–52.

[14] Rainbow research. *Tutorial: Camera eye-to-hand extrinsic calibration*. URL: https://visp-doc.inria.fr/doxygen/visp-daily/tutorial-calibration-extrinsic.html.

[15] Maarten Schoovaerts et al. “Quantitative Assessment of Calibration Motion Profiles in Robotic-assisted Ultrasound System”. In: *2022 International Symposium on Medical Robotics (ISMR)*. 2022, pp. 1–7. DOI: 10.1109/ISMR48347.2022.9807524.

[16] Ziv Yaniv. “Which pivot calibration?” In: *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Robert J. Webster III and Ziv R. Yaniv. Vol. 9415. International Society for Optics and Photonics. SPIE, 2015, p. 941527. DOI: 10.1117/12.2081348. URL: https://doi.org/10.1117/12.2081348.

[17] P. Fischer O. Ronneberger and T. Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *MICCAI*. Vol. 9351. 2015.