



UNIVERSITÀ  
DI TRENTO

Department of Information Engineering  
and Computer Science

Master's Degree in  
Artificial Intelligence Systems

FINAL DISSERTATION

A TELEHEALTH SHARED CONTROL  
FRAMEWORK FOR ASSISTED REMOTE  
LUNG ULTRASOUND EXAMINATION

Supervisor  
Luigi Palopoli

Co-supervisor  
Edoardo Lamon

Student  
Davide Nardi

Academic year 2023/2024

# Acknowledgements

*...thanks to...*

# Contents

<b>Abstract</b>	<b>2</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Lung Ultrasound Protocols</b>	<b>5</b>
2.1 The BLUE Protocol . . . . .	5
2.2 The new 12-Points Protocol . . . . .	6
<b>3 Framework Description</b>	<b>8</b>
3.1 Leader . . . . .	8
3.2 Follower . . . . .	10
<b>4 Connection requirements</b>	<b>14</b>
<b>5 Human Body Estimation</b>	<b>15</b>
5.1 SMPL models family . . . . .	15
5.2 Human body pose estimation . . . . .	16
5.3 Human body shape estimation . . . . .	17
5.4 Human Instance Segmentation with YOLOv8 . . . . .	17
5.5 Point Cloud Filtering using YOLOv8Seg Mask . . . . .	18
5.6 SKEL, a biomechanical model . . . . .	19
5.7 SKEL Fitting from SMPL . . . . .	20
5.8 Modelling the rib cage projection on the skin . . . . .	21
<b>6 Virtual Fixtures</b>	<b>23</b>
6.1 Classification of VFs . . . . .	23
6.2 VFs Generation . . . . .	24

<b>Bibliography</b>	<b>24</b>
---------------------	-----------

<b>A Title first appendix</b>	<b>30</b>
-------------------------------	-----------

# Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

## 1 Introduction

Telehealth is commonly defined as the provision of healthcare services through remote means, obviating the necessity for in-person visits. This paradigm is notably efficacious in reaching geographically isolated areas and has experienced increased adoption during the Covid-19 pandemic, primarily due to its inherent capacity to mitigate the risk of pathogen transmission between patients and healthcare providers. Remote ultrasound examination enhances patient comfort and reduces costs for both patients and the healthcare service [12, 10]. Telerobotics, a promising advancement, involves remote-controlled robotic systems interacting with patients, integrating tools like virtual reality, haptic interfaces, and precise control. Among various applications, medical telerobotics, particularly in ultrasonography, shows significant promise with recent successful implementations [4, 17]. Ultrasound imaging, utilising high-frequency sound waves for noninvasive evaluation of internal structures, provides a safe diagnostic method for detecting and monitoring diseases. However, it necessitates skilled sonographers and risks work-related musculoskeletal disorders due to exerted force [9]. The quality of the diagnosis very much depends on the skills of the sonographer, and the number of experienced operators is not sufficient to deliver the

service in remote areas [20]. Robotised solutions, which involve lightweight manipulators and accurate perception systems, have the potential to eliminate such issues. A crucial aspect is the design of a safe and intuitive framework that allows the physician to obtain faithful and reactive control of the ultrasound probe. For this task, there are two major research fields: robot planning and control theory and the design of the feedback for the physician. Regarding robot planning and control, the literature in this area offers both autonomous [2, 20, 28] and teleoperated [37, 17] solutions. While the two approaches differ mainly on position reference computation; both require the regulation of the contact forces with the patient’s body and an accurate patient location estimation. The latest relevant work with respect to fully autonomous frameworks is [2] in which the authors proposed a complete pipeline for the ultrasound scanning of the thyroid. A rough planning algorithm that guide the robot to the interest region is computed exploiting the point cloud of the patient measured from a RGB-D camera **TODO** continuaires Indeed, applying excessive force during this contact can potentially distort the target anatomical structure and cause harm to the patient. Conversely, insufficient force will not guarantee effective transmission of acoustic waves, leading to poor image quality. Therefore, the goal of most existing robot control methods for ultrasonography is applying a constant force on the probe in the normal direction of the patient’s surface [22, 15, 35]. However, in the scientific community, it’s widely recognised that force controllers can behave unsafely for humans in cases of contact loss. Tsumura et al. [35] suggest using passive springs as an alternative to robot actuators to ensure that maximum force remains within safe limits, prioritising patient safety. However, a significant drawback is their lack of patient specificity. Unlike manual acquisitions, where the sonographer’s experience guides the force, these methods often rely on clinical survey data at best. To address this, Virga *et al.* [36] propose encoding a patient-specific optimal contact force based on ultrasound confidence maps, providing prior information on image quality. Other approaches tailor force to the patient based on estimates of their tissue biomechanical properties. While promising for tissues with similar properties, certain exams, like lung or heart ultrasounds, which involve passing the probe through bones and soft tissue, necessitate adaptable compliance. Additionally, none of these methods can handle exceptional situations at the control level, such as when a patient moves or there is a loss of contact between the probe tip and the patient’s body (potentially due to the presence of gel reducing surface friction). Therefore, closely monitoring and restricting the energy and power transmission the controller can introduce into

the manipulator is crucial for achieving a safe human-robot interaction. By guaranteeing a passive behaviour and decreasing the controller's action in energy-demanding tasks, the potential safety risks are minimised [13, 24, 32, 19, 16]. With respect to the problem of bestowing adequate feedback to the physicians, some interesting designs were proposed in the literature. In *Duan et al.* [11], the physician is able to teleoperate using an appropriate controller panel capable of measuring the force exerted, and the visual feedback is broadcast on the physician's monitors. Other works exploit the capability of mixed reality [7] to improve the visual feedback by virtualising the patient state. Furthermore, augmented reality is used in *Fu et al.* [14] to face the delay problem due to the stream of high-quality data, and a haptic interface is used for the force feedback.

The optimisation problem is formulated as a quadratic programme (QP) with physical constraints obtained by estimating the viscoelastic parameters in advance and safety constraints imposed by the addition of an energy tank. To initialise the proposed control strategy, an offline phase is required, which consists of a discrete biomechanics characterisation and a smoothing operation to obtain a continuous body description. In the online phase, these parameters are used to compute the impedance parameters to ensure accurate force tracking and limited penetration into the body, eliminating the need for precise control in cases of perception inaccuracy or failure. Additionally, the impedance paradigm allows the robot to be controlled in the direction of force exertion, which is not possible with pure force controllers. Furthermore, an energy tank is used to limit both the energy introduced into the system and its power flow, preventing the instantaneous injection of large amounts of energy. The framework was evaluated with a comparison of the main feature that our framework implements against what the state-of-the-art frameworks in the literature offers.

## 2 Lung Ultrasound Protocols

Lung ultrasound (LUS) is used in different healthcare settings, including low- to middle-income countries, where access to advanced imaging technologies like computed tomography (CT) may be limited. Therefore, researchers developed protocols to use of pulmonary ultrasound in clinical settings in a easier way.

### 2.1 The BLUE Protocol

Among the most commonly adopted is the BLUE (Bedside Lung Ultrasound in Emergency) protocol [3], established by Lichtenstein and his team, which is widely used by healthcare professionals to help diagnose the underlying causes of acute respiratory failure.

BLUE were initially intended to be straightforward and fast to diagnose the causes of acute respiratory failure in clinical settings. It was assumed that when acute respiratory failure occurred, the lungs, being such a huge organ, would reveal broad and readily observable diseases. Conditions like pulmonary edema, pulmonary embolism, pneumonia, chronic obstructive pulmonary disease (COPD), asthma, and pneumothorax can be identified using specific profiles outlined in the BLUE protocol. During COVID-19 this protocol became obsolete, since performing a 6-points lung ultrasound exam may miss lung pathology related to COVID-19.

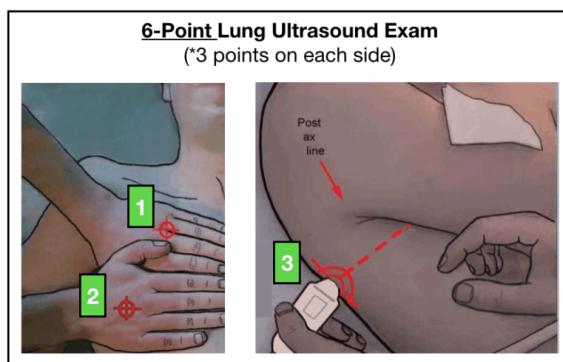


Figure 2.1: 6-points protocol (3 for each side) of BLUE

## 2.2 The new 12-Points Protocol

Soldati et al. [33] presents a comprehensive framework aimed at optimizing the use of LUS in the diagnosis and management of COVID-19 employing a 12-points protocol. The authors highlight the role of LUS as a diagnostic tool, especially in the context of the COVID-19 pandemic as it is capable of detecting various lung conditions. Some of these are lung disease and acute respiratory distress syndrome, which are critical in managing COVID-19 pneumonia. The authors pointed out the need for a unified approach, hence the team has developed a standardized protocol for its use in COVID-19 management. This protocol includes specific recommendations regarding the equipment used, such as the adoption of wireless transducers and tablets. These choices are made to minimize the risk of contamination during examinations, which is particularly important in the context of infectious diseases. By establishing a standardized method, the authors aim to ensure consistency in LUS practices across different healthcare settings. A significant contribution of the article is the introduction of a scoring system designed to classify the severity of lung involvement in COVID-19 patients based on LUS findings. This scoring system was developed through a collaborative effort involving the review of a substantial number of anonymized cases. By categorizing the severity of lung conditions, clinicians can make more informed decisions regarding patient management and treatment strategies, ultimately leading to better patient care. The authors advocate for the establishment of an international database dedicated to sharing LUS images and videos to facilitate the collaboration among healthcare professionals and researchers worldwide. By pooling resources and knowledge, the medical community can enhance the development of pattern recognition algorithms and improve telemedicine capabilities. Such collaboration is crucial in a global health crisis, as it allows for the rapid dissemination of best practices and findings. The article provides practical guidelines for conducting LUS examinations effectively. Key recommendations include avoiding the use of cosmetic filters and specific imaging modalities that may obscure important findings. The authors stress the importance of achieving high frame rates during imaging to capture subtle lung changes accurately. Additionally, they recommend saving data in the Digital Imaging and Communications in Medicine (DICOM) format or as video files to ensure that visual findings can be reviewed comprehensively. With respect to the acquisition, 14 areas should be scanned per patient for 10 seconds. The areas are divided in this way:

- 3 lateral
- 2 lateral
- 2 anterior

. The paper image to display the area of interest is illustrated in Figure 2.2 To capture the largest surface area with a single scan, intercostal scans are required, thus an efficient and accurate way of detecting the ribs is crucial in the design of an autonomous ultrasound system. This thesis aims to face this problem using perception and human body modelling.

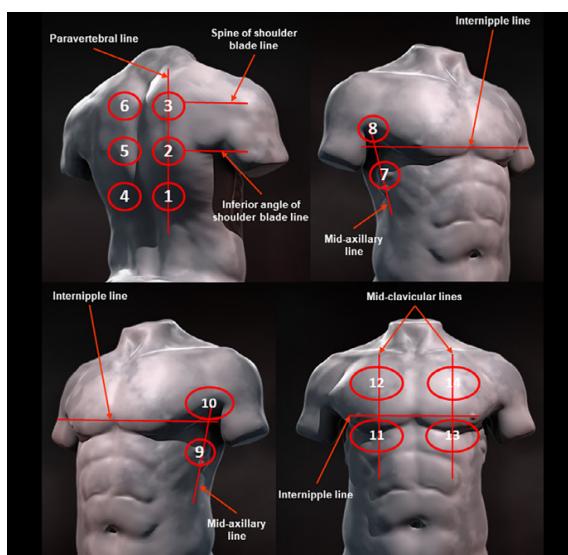


Figure 2.2: The 14-points protocol proposed by Soldati et al.

# 3 Framework Description

The remote robotised US framework consists of two main parts: the leader's side, which corresponds to the environment for the physician, and the follower's side, which corresponds to the environment for the patient (see Figure 3.1). The leader side profits from a real-time updated virtual reality scene streamed over a VR headset with the incoming feedback from the follower, such as the point cloud rendering of the patient overlapped to its SMPL model, the force and wrenches readings of the force and torque sensor mapped into the haptic interface, and the ultrasound probe stream, while the follower side capitalises on a 6Dof robot arm with a force and torque sensor that is controlled with a passive variable Cartesian impedance controller, an ultrasound probe and an RGB-D camera (see Figure 3.2).

## 3.1 Leader

The haptic interface is the input device that the physician uses to control the relative position and orientation of the robot in real-time. It is controlled with a standard impedance control. We mapped the positions as follows:

$$\mathbf{P}_{des}^E = \mathbf{P}_{start}^E + \lambda(\mathbf{P}_{cur}^H - \mathbf{P}_{start}^H) \quad (3.1)$$

where  $\mathbf{P}_{des}^E$  and  $\mathbf{P}_{start}^E \in \mathbb{R}^m$  are respectively the desired and starting position of the end effector,  $\mathbf{P}_{cur}^H$  and  $\mathbf{P}_{start}^H \in \mathbb{R}^m$  are the current and starting position of the haptic interface, and  $\lambda \in \mathbb{R}^{m \times m}$  is a diagonal matrix scaling factor. For the orientation representation, we used unit quaternions, which are well known to be computationally efficient and singularity-free. We implemented orientation mapping with unit quaternions in the following way:

$$\mathcal{Q}_{diff}^H = \overline{\mathcal{Q}}_{cur}^H * \mathcal{Q}_{start}^H \quad (3.2)$$

$$\mathcal{Q}_{des}^E = \overline{\mathcal{Q}}_{diff}^H * \mathcal{Q}_{start}^E \quad (3.3)$$

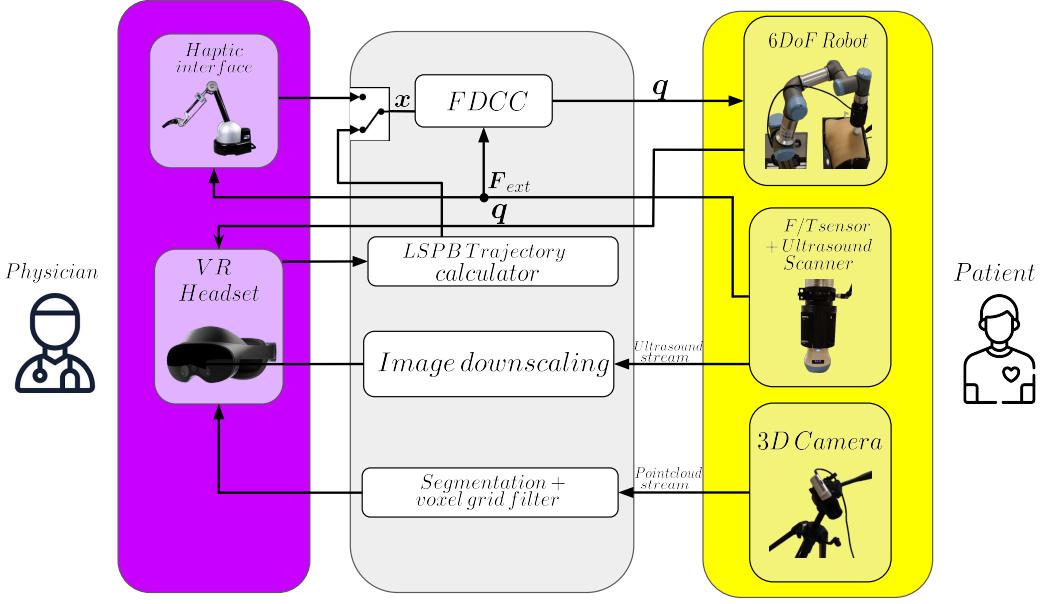


Figure 3.1: The proposed framework: the physician can teleoperate with the haptic interface or with the waypoint placing method. The visual signals coming from the sensors in the follower’s setup are rendered into the VR headset while the force and wrenches feedback are used as input of haptic interface.

where  $\mathcal{Q}_{diff}^H \in \mathbb{R}^4$  is the relative orientation of the haptic interface with respect to its starting position,  $\mathcal{Q}_{start}^H \in \mathbb{R}^4$  and  $\overline{\mathcal{Q}}_{cur}^H \in \mathbb{R}^4$  are the starting and the conjugate of the current orientation of the haptic interface,  $\mathcal{Q}_{des}^E \in \mathbb{R}^4$  is the desired orientation of the end effector,  $\overline{\mathcal{Q}}_{diff}^H \in \mathbb{R}^4$  is the conjugate of  $\mathcal{Q}_{diff}^H$  and  $\mathcal{Q}_{start}^E$  is the starting orientation of the end effector. The absence of the scaling factor in (3.3) is motivated by the fact that we experienced a counterintuitive orientation control if we scale  $\overline{\mathcal{Q}}_{diff}^H$ . At this point, a doctor would already be able to control the robot manipulator and receive force feedback from the patient, but he would miss the visual context of the operation. For this purpose, we recreated the room for ultrasonography in virtual reality (see Figure 3.3), that can be explored through a VR headset. We displayed the robot state by applying the joint values to a virtual model, which is visible to the physician. The point cloud computed from the RGBD sensor in the follower framework is rendered over the SMPL model, which is used just to monitor the patient pose. In our virtual scene, we also added a screen for broadcasting the ultrasound image. This feature can exploit VR capabilities to address the problem of not having the patient and the output of the ultrasonography in the same field of view since we can move the screen near the patient or make it follow the headset orientation in a dynamic way. This means that the physician will always be able to

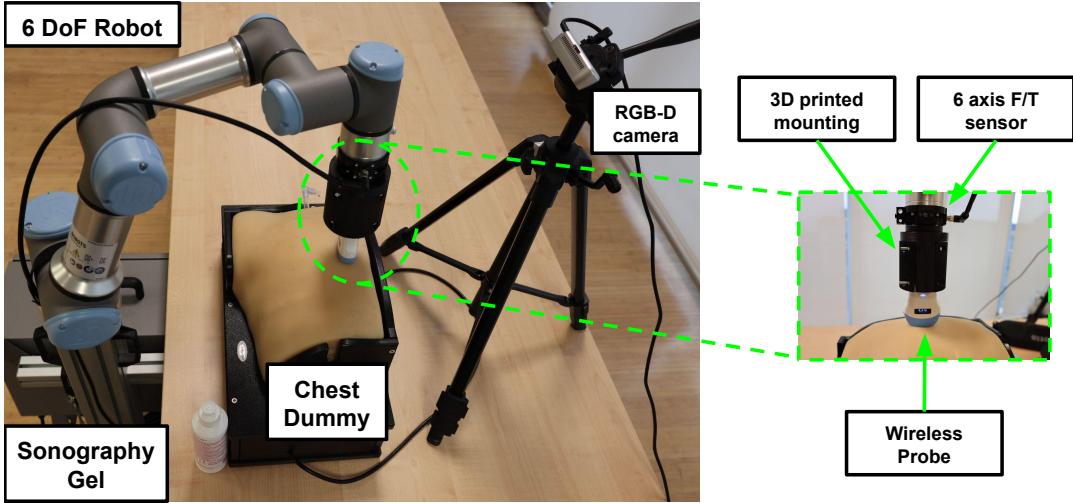


Figure 3.2: In the follower’s setup, a chest dummy is used for the examination done by a robot manipulator equipped with a force/torque sensor and an ultrasound probe. A RGBD camera captures the 3d scene, creating a pointcloud.

monitor both the patient and the ultrasound output, enhancing the control and safety of the teleoperation. Another consequence related to the use of a virtual environment is the possibility for the doctor to move without limitations around the virtual patient, avoiding constraints on motion due to objects or the patient itself.

### 3.2 Follower

For the robot’s controller, we used the FDCC controller[30], where the closed-loop behaviour of a Cartesian compliant controller relates the external wrenches applied to the end-effector of the robot to a mass-spring-damper model as follows:

$$\mathbf{F}^{ext} = \mathbf{\Lambda}^d \ddot{\tilde{x}} + \mathbf{D}^d \dot{\tilde{x}} + \mathbf{K}^d \tilde{x}, \quad (3.4)$$

where  $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_d \in \mathbb{R}^m$  is the Cartesian error computed with respect to the desired Cartesian end-effector pose  $\mathbf{x}_d$ ,  $\mathbf{F}^{ext} \in \mathbb{R}^m$  is the external wrench applied on the end-effector, and  $\mathbf{\Lambda}_d, \mathbf{D}_d, \mathbf{K}_d \in \mathbb{R}^{m \times m}$  are the desired Cartesian inertia, damping, and stiffness, respectively.

As system stability might be violated in the presence of a variable impedance controller, we enforced system passivity through the concept of passivity of the power port  $\dot{\mathbf{x}}^T \mathbf{F}^{ext}$ . To do so, we introduce the formalism of port-Hamiltonian systems to describe the interaction model of the variable Cartesian impedance augmented with an energy tank [13] with

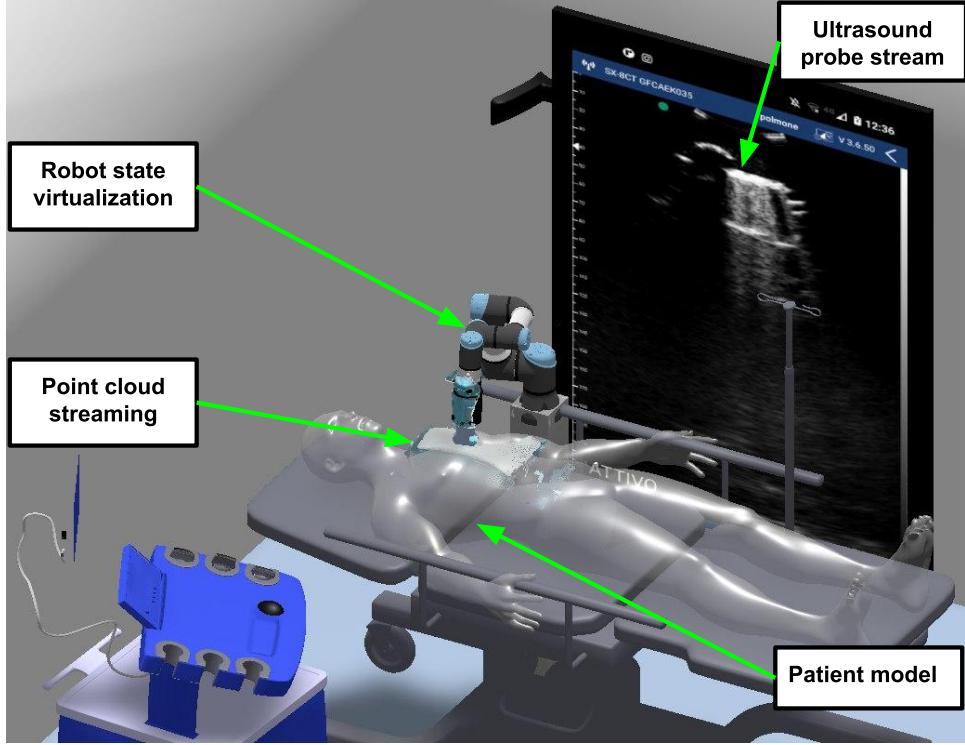


Figure 3.3: The virtual scene seen by the physician: the pointcloud of the chest dummy is mapped in the same relative pose of the patient’s chest, and the stream of the ultrasound probe is put right in front of the physician’s view as long as the robot is a digital twin.

dynamics:

$$\dot{\mathbf{x}}_t = \frac{\sigma}{\mathbf{x}_t} \dot{\tilde{\mathbf{x}}}^T \mathbf{D}^d \dot{\tilde{\mathbf{x}}} - \frac{\mathbf{w}^T}{\mathbf{x}_t} \dot{\tilde{\mathbf{x}}}, \quad (3.5)$$

where  $\mathbf{x}_t \in \mathbb{R}$  is the state of the tank,  $\sigma \in \{0, 1\}$  modulates the energy storage, and  $\mathbf{w}$  an the extra input of the port-Hamiltonian system defined as:

$$\mathbf{w}(t) = \begin{cases} -\mathbf{K}^v(t) \tilde{\mathbf{x}} & \text{if } T(\mathbf{x}_t) > T_{min} \\ 0 & \text{otherwise,} \end{cases} \quad (3.6)$$

where  $\mathbf{K}^v(t)$  is time-varying component of the stiffness ( $\mathbf{K}^d(t) = \mathbf{K}^{min} + \mathbf{K}^v(t)$ ). At each time instant, the tank’s energy is defined by  $T(\mathbf{x}_t) = \frac{1}{2} \mathbf{x}_t^2$  and  $T_{min} \in \mathbb{R}^+$  is the minimum energy that the tank is allowed to store. Thanks to (3.6), we can infer the condition  $T(\mathbf{x}_t) > T_{min}$  when the stiffness is allowed to raise without violating the passivity constraint. However, this bound does not prevent the energy of the tank to be drained instantaneously, situation which leads to the complete loss of performance. For this reason, it is reasonable to further constrain the power flow of the tank when the energy is extracted

from the tank ( $\dot{T}(\mathbf{x}_t) > -\eta$ ), where  $\eta \in \mathbb{R}^+$  is the maximum allowed power. The goal is to track a constant force  $\mathbf{F}^d$  changing the stiffness matrix, which can be computed solving the QP in (3.7). The optimisation problem involves a trade-off between the precise tracking of a desired wrench and the necessity to uphold a limited level of stiffness. Inspired by [38], we formulated the QP as follows:

$$\begin{aligned} \min_{\mathbf{K}^d \in \mathbb{R}^{m \times m}} \quad & \frac{1}{2} \left( \|\mathbf{F}^{ext} - \mathbf{F}^d\|_{\mathbf{Q}}^2 + \|\mathbf{K}^d - \mathbf{K}^{min}\|_{\mathbf{R}}^2 \right) \\ \text{s.t. } \quad & \mathbf{K}^{min} \leq \mathbf{K}^d \leq \mathbf{K}^{max} \\ & \mathbf{F}^{min} \leq \mathbf{F}^{ext} \leq \mathbf{F}^{max} \\ & -\mathbf{K}^d \tilde{\mathbf{x}} \leq \sigma \dot{\mathbf{x}}^T \mathbf{D}^d \dot{\mathbf{x}} - \tilde{\mathbf{x}}^T \mathbf{K}_{min} \dot{\mathbf{x}} + \frac{T_{t-1} - T_{min}}{\Delta t} \\ & -\mathbf{K}^d \tilde{\mathbf{x}} \leq \sigma \dot{\mathbf{x}}^T \mathbf{D}^d \dot{\mathbf{x}} - \tilde{\mathbf{x}}^T \mathbf{K}_{min} \dot{\mathbf{x}} - \eta \end{aligned} \tag{3.7}$$

where  $N$  is the length of the time window,  $\mathbf{Q}$  and  $\mathbf{R} \in \mathbb{R}^{m \times m}$  are diagonal positive definite weighting matrices,  $\mathbf{K}^d \in \mathbb{R}^{m \times m}$  is the desired stiffness of the Cartesian impedance controller,  $\mathbf{K}^{min}$  and  $\mathbf{K}^{max} \in \mathbb{R}^{m \times m}$  are respectively the minimum and maximum allowed stiffness,  $\mathbf{F}^{ext} \in \mathbb{R}^m$  is the wrench of the impedance interaction model, which can be modelled with (3.4),  $\mathbf{F}^d \in \mathbb{R}^m$  is the desired interaction wrench and  $\mathbf{F}^{max}/\mathbf{F}^{min} \in \mathbb{R}^m$  is the maximum/minimum wrench that the robot can exert. The last two constraints limit the maximum energy  $T_{min}$  which can be injected in the system and the rate  $\eta$  at which the energy is injected and are obtained from  $T \leq T_{min}$  and  $\dot{T} \leq -\eta$ .  $1/\Delta t$  is the controller frequency. When  $T \leq T_{min}$  is not satisfied, the stiffness decreases to its minimum ( $\mathbf{K}^d = \mathbf{K}^{min}$ ).

More details on the QP problem formulation and energy tank constraint definition are available in [5] referring to the VF-CF control law, but assuming the target force to be fixed instead of pre-computed offline. The measured force and torque are used for control but also as feedback for the physician through the haptic interface. A straight one-to-one mapping of these quantities from the sensor to the device produces some vibrations on the haptic interface. This is related to the noise in the F/T sensor readings. To address this problem, we added an exponential moving average filter (EWMA) [25], which represent a good trade-off between smoothness and introduced delay:

$$\mathbf{F}_H^{ext} = \alpha \mathbf{F}_{cur}^{ext} + (\mathbf{I} - \alpha) \mathbf{F}_{old}^{ext} \tag{3.8}$$

where  $\mathbf{F}_H^{ext} \in \mathbb{R}^m$  is a vector of forces and torques that are the target output of the haptic device,  $\mathbf{F}_{cur}^{ext}$  and  $\mathbf{F}_{old}^{ext} \in \mathbb{R}^m$  are the vectors of forces and torques of the current and precedent control loop,  $\mathbf{I} \in \mathbb{R}^{m \times m}$  is the identity matrix, and  $\boldsymbol{\alpha} \in \mathbb{R}^{m \times m}$  is a diagonal matrix of weights.

## 4 Connection requirements

The problem of transmitting large data is one of the open problems in the information technology research area. A teleoperated system like the one presented in this work need a stable, low-latency connection with high bandwidth requirements. In the proposed framework I needed to transmit different types of data with different requirements. We defined requirements as *critical* when a connection drop could harm the patient or cause damage to the equipment.

	Critical	High Bandwidth(>5MB/s)	Low Latency (<5ms)
Ultrasound image stream	✗	✓	✓
Force/Torque measurement	✓	✗	✓
Haptic device pose measurement	✓	✗	✓
Point cloud of the patient	✗	✓	✗
SMPL body parameters	✗	✗	✓
Robot state	✗	✗	✓

Table 4.1: Connection requirements for the teleoperation framework

# 5 Human Body Estimation

In this chapter, I will explain how I managed to estimate the parametric model of the patient exploiting visual information.

## 5.1 SMPL models family

For the body shape and pose, there is a well known model called SMPL (Skinned Multi-Person Linear Model) [21] which is a realistic 3D model of the human body that is based on skinning and blend shapes. SMPL is learned from thousands of 3D body scans. SMPL model pose is described by 23 body joints together with body position and orientation called  $\vec{\theta}$ , while the body shape is described by 10 parameters called  $\vec{\beta}$ . Each  $\theta_i, i \in [1, 21]$  represents the 3D orientation of a human joint with respect to its kinematic parent. The kinematic tree is expressed as a connectivity map that connect each joint to its child, as shown in Figure 5.1. Orientations are represented by rotation vectors and can be converted

```
[ 0, 1 ],
[ 0, 2 ],
[ 0, 3 ],
[ 1, 4 ],
[ 2, 5 ],
[ 3, 6 ],
[ 4, 7 ],
[ 5, 8 ],
[ 6, 9 ],
[ 7, 10 ],
[ 8, 11 ],
[ 9, 12 ],
[ 9, 13 ],
[ 9, 14 ],
[ 12, 15 ],
[ 13, 16 ],
[ 14, 17 ],
[ 16, 18 ],
[ 17, 19 ],
[ 18, 20 ],
[ 19, 21 ],
[ 20, 22 ],
[ 21, 23 ],
```

Figure 5.1: The SMPL kinematic tree

to rotation matrices using the Rodriguez formula. Since the model is linear, the inference time is fast. There exist multiple variants of this model:

- SMPL-H[26] is the extended version of SMPL that accounts also for hands position. It leverages on 21 body joints from the SMPL model, replacing the two additional joints related to the hand position with 15 additional joints for each hand.

- SMPLX[23] (SMPL eXpressive) is the extended version of SMPL-H which accounts also for the face expressions. The total number of joints is 54 which includes joints for the neck, jaw, eyeballs, and fingers.

The difference between SMPL and SMPLX is show in Figure 5.2. These more recent

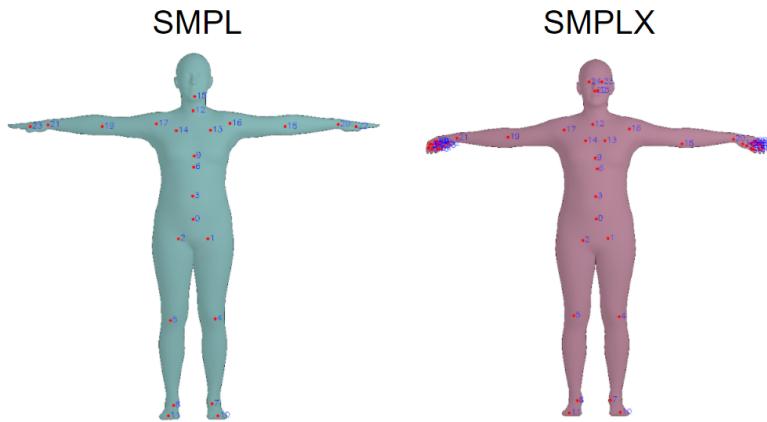


Figure 5.2: The comparison of the number and location of joints for SMPL and SMPLX

extensions of SMPL can be used in tasks where knowing the exact hand configuration is crucial or when the facial expression information can be exploited in some way. In my opinion, SMPL-H and SMPLX will be soon introduced in different human-robot collaboration scenarios. Since for RLU hands and facial expressions were not included in the logic of my thesis I took the SMPL for modelling the patient state.

## 5.2 Human body pose estimation

As already mentioned in the previous section, the SMPL model is composed by  $\vec{\theta}$  and  $\vec{\beta}$ . It is possible to animate a SMPL model in real-time if  $\vec{\beta}$  does not need to be optimized, updating just the  $\vec{\theta}$  parameters of the model. For this purpose I used a ZED2 stereo camera which integrates an AI body tracking module[1] based on the depth and RGB image measured by the sensor itself. The body tracking module provides different skeleton models. I chose the one more similar to SMPL, which is also the most informative, since it is the one with the highest number of keypoints (38). A figure showing the pose of the keypoints is reported in Figure 5.3.

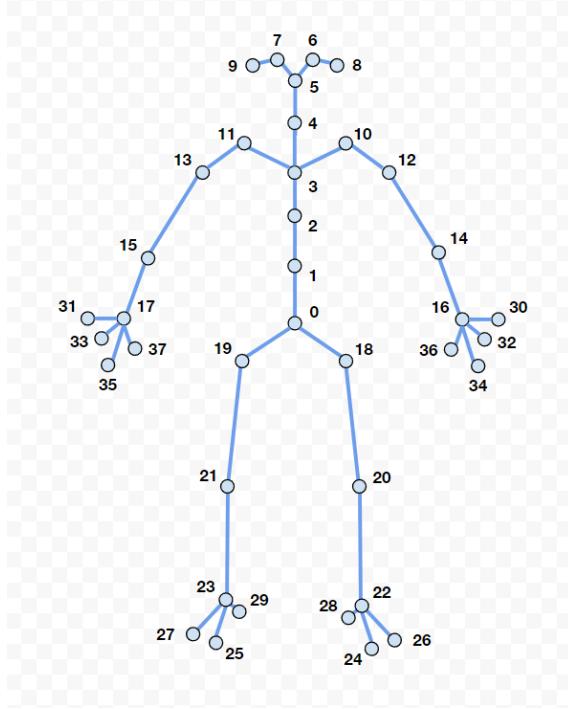


Figure 5.3: BODY38 model provided by the ZED SDK

Comparing Figure 5.2 and Figure 5.3 it is easy to notice that the joint location is similar between the two models, so it was straight forward to compute the right input for the SMPL model for the real time animation. Since the ZED AI module already provides the local orientation of each joints with respect to its parent in the kinematic tree (which is the same of the SMPL), it was sufficient to convert from the quaternion notation for the orientation to the rotation vector notation.

### 5.3 Human body shape estimation

Given that this master’s thesis aims to address the problem of the physical variability between patients in RLU,  $\vec{\beta} \in \mathbb{R}^{10}$  is a crucial parameter to be considered since it affects the rib cage dimension and position. An external source of information is required in order to evaluate a metric for the shape parameters optimization. I decided to use the same exteroceptive sensor used for the pose estimation procedure: the ZED2. Therefore, I retrieved the raw point cloud of the scene.

### 5.4 Human Instance Segmentation with YOLOv8

Since I needed an intelligent way of ignoring points not belonging to the human, I exploited YOLOv8 segmentation neural network from Ultralytics in order to detect which image

pixel belongs to the patient. YOLOv8 segmentation is an advanced computer vision model that combines object detection with instance segmentation. Ultralytics provides different models for the instance segmentation of 80 different classes. When applied to human body segmentation, it not only identifies humans in an image but also generates a precise mask outlining the shape and boundaries of the body. This is achieved through pixel-level predictions, allowing the model to differentiate the human body from the background and other objects. YOLOv8's segmentation capabilities are highly efficient, making it suitable for real-time applications like pose estimation, augmented reality, and medical imaging.

## 5.5 Point Cloud Filtering using YOLOv8Seg Mask

The mask computed by YOLO is represented by the vertices of a polygon. I converted it to a bit mask matrix through the `fillPoly` of opencv having 1 or 0 for each pixel of the original image. The resulting image is then used to decide which points of the point cloud to discard. Moreover, I exploited the bounding box provided by YOLO in addition to the mask to limit the amount of computation needed to filter the point cloud. To be more explicit, I restricted the double nested loop over the image pixels to the bounding box instead of the full image. The resulting point cloud is illustrated in Figure 5.4

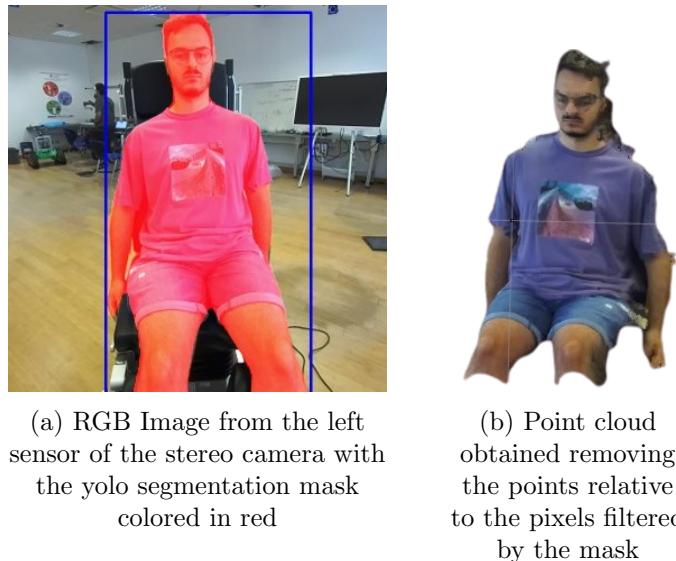


Figure 5.4

## 5.6 SKEL, a biomechanical model

SKEL[18] is a biomechanical skeleton model that maps the SMPL skinned model to its anatomical skeleton. The new model differs from the existing SMPL model in several key ways:

- Anatomical accuracy: In comparison to SMPL, SKEL offers more precise joint placements and anatomically correct bone orientations. It includes biomechanical factors, like better knee alignment and arm supination, that enable a more accurate depiction of human anatomy.
- Posture parameters: SKEL is more effective in regressing precise anatomical joints from video data because it has fewer posture parameters (46), as opposed to SMPL’s (72). Applications in computer vision and biomechanics benefit from this reduction in parameters.
- Biomechanical features: SKEL incorporates a biomechanical shoulder blade that enables for more realistic movement in place of the SMPL’s approximation shoulder structure. Additionally, it simulates how the ulna and radius bones move in the forearm pronation and supination, which constitute a major advancement above the SMPL’s basic rotation around the elbow.
- Integration of skin and skeleton: SKEL enables direct animation of both components by concurrently rigging the skin and skeleton meshes with the same pose parameters. For this reason, SKEL can inherit the pose and shape directly from SMPL.
- Data utilization: To improve the accuracy of the skeletal structure and joint placements, SKEL makes use of the BioAMASS dataset, which integrates motion capture data with anatomical information. In comparison, SMPL does not include such comprehensive biomechanical data.

Compared to the SMPL framework, SKEL represents a substantial leap in producing 3D human models that are more lifelike and anatomically precise. An illustrative image of SKEL can be observed in Figure 5.5



Figure 5.5: An image taken from the SKEL paper, as we can see in the images on the left, both the skin and the skeleton are present in the output of the SKEL model

## 5.7 SKEL Fitting from SMPL

Since I computed the SMPL from the patient state in section 5.3, the next step was to compute the skeleton inside the SMPL mesh. Since SKEL has the same surface mesh topology and shape parameters  $\vec{\beta}$  as SMPL, it can be directly fit to existing SMPL meshes by optimizing its pose parameters to minimize the vertex-to-vertex error between the SMPL skin and the SKEL skin. The alignment is avoid local minima by aligning the model in this order:

1. Align pelvis
2. Align forearm and thighs
3. Align elbows and knees
4. Align hands, feet, and neck

The order is the same of the kinematic tree since each joint position depends on its predecessor. In order to locally optimize some particular joints, it is sufficient to mask the not relevant contributions in the loss computation. The loss used for the optimization were:

- Pose loss: The loss between the joint position of SMPL and the one of SKEL
- Vertex loss: Since the skin mesh of SMPL is the same as SKEL, it is possible to compute vertex-wise displacement
- Scapula loss: loss added to regularize the scapula fitting avoiding large values, implemented with masking on non-relevant joints
- Spine loss: loss added to regularize the spine fitting avoiding large values, implemented with masking on non-relevant joints

**TODO** add formula

## 5.8 Modelling the rib cage projection on the skin

Once the SKEL model is computed from the SMPL retrieved from the point cloud, we need to understand where it is located during the robot interaction with the skin. Different geometrical intuition and approximation can be used to estimate how the rib cage is projected on the skin. My first idea was to radially project the skeleton mesh vertices to the skin. The result was poor, especially for ribs far from the projection centre. I then switched to cylindrical projection, considering the cylinder axis parallel to the backbone. The results were qualitatively more realistic than the one produced by the spherical projection. The projection was implemented using ray casting algorithm provided by `Open3d` that finds the 3D intersection point of a ray to a mesh. A visual comparison between radial and cylindrical projection is shown in Figure 5.6.



Figure 5.6: Comparison between the two projection methods tested: on the left the radial projection, on the right the cylindrical projection

Since each vertex is projected on the skin, the result is a point cloud. We can reconstruct a mesh from the projection connecting each vertex to his adjacent vertices in the original rib cage mesh. The result of such operation is shown in Figure 5.7. As we can notice, the ribs are reconstructed correctly.

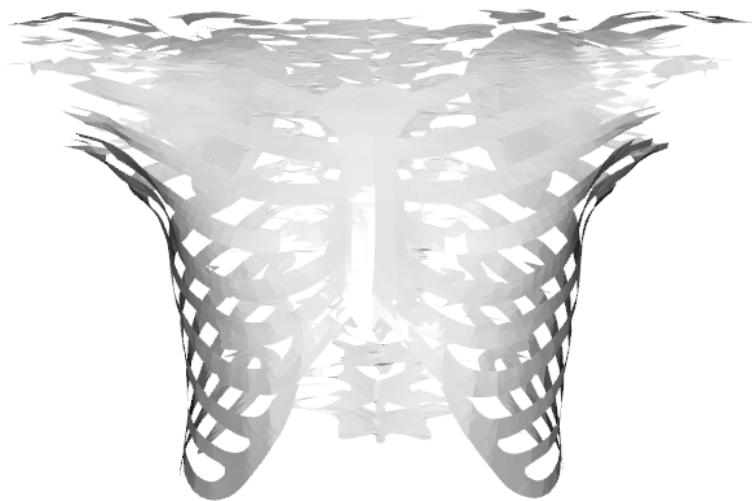


Figure 5.7: The resultant mesh obtained from the point cloud retrieved by the cylindrical projection of the rib cage original mesh.

The idea is to create a virtual fixture that guide the physician towards the intercostal area, exploiting the semantic knowledge expressed by the SKEL model to compensate the lacks of the teleoperation.

# 6 Virtual Fixtures

Virtual fixtures (VF), also called active constraints, are a family of algorithms used for limiting the workspace of the robot for safety and accuracy reasons, thus they are widely used in medical robotics. They were introduced in 1992 [27] by Rosemberg in an augmented reality teleoperation setup to simulate physical barriers, fields, and guides, designed to assist in the user during the teleoperation. VFs can be modelled with different shapes as shown in [6] where the author proposed a tube-like boundary with an admittance control and haptic feedback. Later on, different works show the crucial importance of VFs in medical robotics scenarios where the workspace of the robot must be limited to the area of interest to avoid hurting patients. In 2012 Rydén et al. [29] proposed an algorithm for the identification of a forbidden region virtual fixture from point clouds analysis that was used for operating on a beating heart. The virtual fixture region is computed, considering each point of the point cloud as a 3D sphere with a particular radius. A known problem of VFs is the non passivity of the system when crossing or getting closer to the boundaries.

## 6.1 Classification of VFs

A good classification of these algorithms is given in [8]. One of the most significant classifications is whether the fixture belongs to the *forbidden region VF* or *guidance VF*. The former bounds a manipulator’s tool to certain regions within its task or joint space, the latter encourages the user to move the tool along a specific pathway or toward a specific target. Usually, guidance constraints are more intrusive upon the user than forbidden region constraints in normal operation. The second meaningful classification for this work highlighted by the author is whether the constraint is *unilateral* or *bilateral*. A unilateral virtual fixture is a constraint active only by one side, while the bilateral acts on both sides. Regional VFs (forbidden region) often are implemented with unilateral constraints, whereas guidance VFs leverage on bilateral limits. The last distinction is between *static* and *dynamic* constraints that depend on if the geometry exploited by the algorithm change during the operation.

## 6.2 VFs Generation

In the literature, different methods for the generation of virtual fixture were proposed. The survey [8] mentions the most common types of constraint representations used in the literature.

1. (a) Point: A single 3d point can be used to generate an attractive or repulsive potential field
2. (b) Line: can be used to generate a guidance
3. (c) Parametric curve: generalization of the line for task that requires some dexterity
4. (d) Plane: Usually used as a forbidden region constraint to limit the robot workspace
5. (e) Parametric surface: generalization of the plane. A well-known way for generating a parametric surface from a set of points is the Non Uniform Rational Basis-Splines (NURBS)
6. (f) Polygonal Mesh: Non-parametric but realistic way of representing a real object.  
Can be used when the fidelity of the constraint is crucial for the task.
7. (g) Point Cloud: A virtual fixture can be generated by a point cloud as explained in [29]
8. (h) Volumetric primitives: Boolean algebra can be applied to 3d mesh to obtain complex shapes maintaining computation efficiency in order to model situations in which the efficiency is critical
9. (i) Explicitly described areas: work well in small scale scenarios but suffer from the curse of dimensionality.

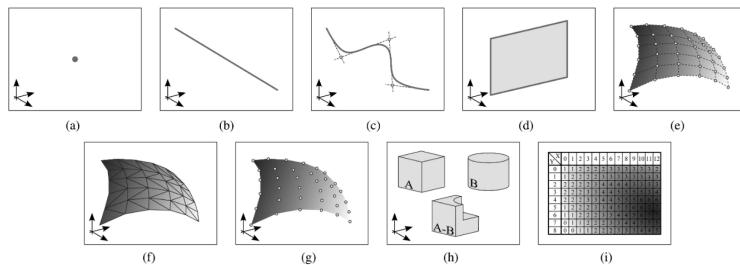


Figure 6.1: The most common types of virtual fixture geometries reported by Bowyer et al.

**TODO** storia delle VF con reference [31] [34]

# Bibliography

- [1] Body Tracking Overview - Stereolabs.
- [2] A fully autonomous robotic ultrasound system for thyroid scanning | Nature Communications.
- [3] The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments.
- [4] Sotiris Avgousti, Eftychios G. Christoforou, Andreas S. Panayides, Sotos Voskarides, Cyril Novales, Laurence Nouaille, Constantinos S. Pattichis, and Pierre Vieyres. Medical telerobotic systems: current status and future trends. *BioMedical Engineering OnLine* 2016 15:1, 15(1):1–44, 8 2016.
- [5] Luca Beber, Edoardo Lamon, Davide Nardi, Daniele Fontanelli, Matteo Saveriano, and Luigi Palopoli. A passive variable impedance control strategy with viscoelastic parameters estimation of soft tissues for safe ultrasonography. In *arXiv*, 9 2023.
- [6] A. Bettini, P. Marayong, S. Lang, A.M. Okamura, and G.D. Hager. Vision-assisted control for manipulation using virtual fixtures. *IEEE Transactions on Robotics*, 20(6):953–966, December 2004. Conference Name: IEEE Transactions on Robotics.
- [7] David Black and Septimiu Salcudean. Mixed reality human teleoperation. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 375–383. IEEE, 3 2023.
- [8] Stuart A. Bowyer, Brian L. Davies, and Ferdinando Rodriguez y Baena. Active Constraints/Virtual Fixtures: A Survey. *IEEE Transactions on Robotics*, 30(1):138–157, February 2014. Conference Name: IEEE Transactions on Robotics.

- [9] Carolyn T. Coffin. Work-related musculoskeletal disorders in sonographers: A review of causes and types of injury and best practices for reducing injury risk. *Reports in Medical Imaging*, 7(1):15–26, 2014.
- [10] Digital Health and Innovation. Consolidated telemedicine implementation guide. Technical report, World Health Organization, 2022.
- [11] Bin Duan, Linfei Xiong, Xiaolong Guan, Yongqing Fu, and Yongwei Zhang. Tele-operated robotic ultrasound system for medical diagnosis. *Biomedical Signal Processing and Control*, 70:102900, 9 2021.
- [12] European Reference Networks and Digital Health. Market study on telemedicine. Technical report, European Commission, 2018.
- [13] Federica Ferraguti, Nicola Preda, Auralius Manurung, Marcello Bonfe, Olivier Lamberty, Roger Gassert, Riccardo Muradore, Paolo Fiorini, and Cristian Secchi. An Energy Tank-Based Interactive Control Architecture for Autonomous and Teleoperated Robotic Surgery. *IEEE Transactions on Robotics*, 31(5):1073–1088, 10 2015.
- [14] Yongqing Fu, Weiyang Lin, Xinghu Yu, Juan J. Rodriguez-Andina, and Huijun Gao. Robot-assisted teleoperation ultrasound system based on fusion of augmented reality and predictive force. *IEEE Transactions on Industrial Electronics*, 70:7449–7456, 7 2023.
- [15] Christoph Hennersperger, Bernhard Fuerst, Salvatore Virga, Oliver Zettinig, Benjamin Frisch, Thomas Neff, and Nassir Navab. Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study. *IEEE Transactions on Medical Imaging*, 36(2):538–548, 2 2017.
- [16] Sebastian Hjorth, Edoardo Lamon, Dimitrios Chrysostomou, and Arash Ajoudani. Design of an Energy-Aware Cartesian Impedance Controller for Collaborative Disassembly. *Proceedings - IEEE International Conference on Robotics and Automation*, 2 2023.
- [17] Zhongliang Jiang, Septimiu E. Salcudean, and Nassir Navab. Robotic ultrasound imaging: State-of-the-art and future perspectives. *Medical Image Analysis*, 89:102878, 10 2023.

- [18] Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C. Karen Liu, and Michael J. Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. In *ACM ToG, Proc. SIGGRAPH Asia*, volume 42, December 2023.
- [19] Johannes Lachner, Felix Allmendinger, Eddo Hobert, Neville Hogan, and Stefano Stramigioli. Energy budgets for coordinate invariant robot control in physical human–robot interaction. *International Journal of Robotics Research*, 40(8-9):968–985, 8 2021.
- [20] Keyu Li, Yangxin Xu, and Max Q.H. Meng. An Overview of Systems and Techniques for Autonomous Robotic Ultrasound Acquisitions. *IEEE Transactions on Medical Robotics and Bionics*, 3(2):510–524, 5 2021.
- [21] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.
- [22] Samir Merouche, Louise Allard, Emmanuel Montagnon, Gilles Soulez, Pascal Bigras, and Guy Cloutier. A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(1):35–46, 1 2016.
- [23] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 10975–10985, 2019.
- [24] Gennaro Raiola, Carlos Alberto Cardenas, Tadele Shiferaw Tadele, Theo De Vries, and Stefano Stramigioli. Development of a Safety- and Energy-Aware Impedance Controller for Collaborative Robots. *IEEE Robotics and Automation Letters*, 3(2):1237–1244, 4 2018.
- [25] S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1:239–250, 8 1959.

- [26] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), November 2017.
- [27] Louis B. Rosenberg. The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments.
- [28] Mariadas Capsran Roshan, Adrian Pranata, and Mats Isaksson. Robotic Ultrasonography for Autonomous Non-Invasive Diagnosis - A Systematic Literature Review. *IEEE Transactions on Medical Robotics and Bionics*, 4(4):863–874, 11 2022.
- [29] Fredrik Rydén and Howard Jay Chizeck. Forbidden-region virtual fixtures from streaming point clouds: Remotely touching and protecting a beating heart. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3308–3313, October 2012. ISSN: 2153-0866.
- [30] Stefan Scherzinger, Arne Roennau, and Rudiger Dillmann. Forward Dynamics Compliance Control (FDCC): A new approach to cartesian compliance for robotic manipulators. *IEEE International Conference on Intelligent Robots and Systems*, 2017-September:4568–4575, 12 2017.
- [31] Mario Selvaggio, Giuseppe Andrea Fontanelli, Fanny Ficuciello, Luigi Villani, and Bruno Siciliano. Passive Virtual Fixtures Adaptation in Minimally Invasive Robotic Surgery. *IEEE Robotics and Automation Letters*, 3(4):3129–3136, October 2018. Conference Name: IEEE Robotics and Automation Letters.
- [32] Erfan Shahriari, Lars Johannsmeier, Elisabeth Jensen, and Sami Haddadin. Power Flow Regulation, Adaptation, and Learning for Intrinsically Robust Virtual Energy Tanks. *IEEE Robotics and Automation Letters*, 5(1):211–218, 1 2020.
- [33] Gino Soldati, Andrea Smargiassi, Riccardo Inchegolo, Danilo Buonsenso, Tiziano Perrone, Domenica Federica Briganti, Stefano Perlini, Elena Torri, Alberto Mariani, Elisa Eleonora Mossolani, Francesco Tursi, Federico Mento, and Libertario Demi. Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19. *Journal of Ultrasound in Medicine*, 39(7):1413–1419, 7 2020.
- [34] Gino Soldati, Andrea Smargiassi, Riccardo Inchegolo, Danilo Buonsenso, Tiziano Perrone, Domenica Federica Briganti, Stefano Perlini, Elena Torri, Alberto Mariani,

Elisa Eleonora Mossolani, Francesco Tursi, Federico Mento, and Libertario Demi. Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19. *Journal of Ultrasound in Medicine*, 39(7):1413–1419, July 2020. Publisher: John Wiley & Sons, Ltd.

- [35] Ryosuke Tsumura and Hiroyasu Iwata. Robotic fetal ultrasonography platform with a passive scan mechanism. *International Journal of Computer Assisted Radiology and Surgery*, 15(8):1323–1333, 8 2020.
- [36] Salvatore Virga, Oliver Zettinig, Marco Esposito, Karin Pfister, Benjamin Frisch, Thomas Neff, Nassir Navab, and Christoph Hennersperger. Automatic force-compliant robotic Ultrasound screening of abdominal aortic aneurysms. *IEEE International Conference on Intelligent Robots and Systems*, 2016-November:508–513, 11 2016.
- [37] Felix von Haxthausen, Sven Böttger, Daniel Wulff, Jannis Hagenah, Verónica García-Vázquez, and Svenja Ipsen. Medical Robotics for Ultrasound Imaging: Current Systems and Future Trends. *Current Robotics Reports* 2021 2:1, 2(1):55–71, 2 2021.
- [38] J. Zhao, A. Giammarino, E. Lamon, J.M. Gandarias, E.D. Momi, and A. Ajoudani. A Hybrid Learning and Optimization Framework to Achieve Physically Interactive Tasks With Mobile Manipulators. *IEEE Robotics and Automation Letters*, 7(3), 2022.

## Appendix A Title first appendix