



**UNIVERSITÀ
DI TRENTO**

**Department of
Industrial Engineering**

XXXIV cycle

**Doctoral School in Materials, Mechatronics
and System Engineering**

**Measurement of Neuro-Mechanical
Parameters for the Biomechanical Analysis
and a Mixed Reality Data Visualization to
Augment the Clinical Eye**

Nicola Covre

June 2022

Measurement of Neuro-Mechanical Parameters for the Biomechanical Analysis and a Mixed Reality Data Visualization to Augment the Clinical Eye

Nicola Covre

Email: nicola.covre@unitn.it

Approved by:

Prof. Mariolino De Cecco, Advisor

Dept. of Industrial Engineering

University of Trento, Italy

Ph.D. Commission:

Prof. Francesco Crenna

Dept. of Mechanical Engineering

University of Genova

Prof. Paolo Bosetti

Dept. of Industrial Engineering

University of Trento

Prof. Matteo Lancini

Dept. of Industrial Engineering

University of Brescia

University of Trento
Department of Industrial Engineering

June 2022

I would like to dedicate this thesis to my parents Anna and Stefano, to my brothers Giovanni, Luca and Francesco, and to Elena who loved and supported me all these years...

¹I would like to dedicate a special thanks to Professor De Cecco for believing and supporting me along with my academic research work. I would like to thank my colleagues and friends Paolo Tomasin, Alessandro Luchetti, Matteo Lancini, Fabrizio Nunnari, Irene Sorianini, Federico Gala, Clara Sanz, Luca Maule, Matteo Zanetti, Prof. Davide Brunelli, Matteo Nardello, Alessandro Torrisi, Davide Massella, Damiano Fruet e Alberto Fornaser for all the great projects and collaborations we experienced together. Finally, I would like to thank my friends Alessandro Cazzolli, Francesco Galvagnini, Anna Saiani, Filippo Perri, Fancesco Sica, Martin Zatelli, Nicola Cestaro, Andrea Vigorito and Valeria Pia Laurito for their support and friendship over these years.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Nicola Covre
June 2022

Acknowledgements

And I would like to acknowledge that this research project was developed in collaboration with the Mirolab research team of the Department of Industrial Engineering of the University of Trento, the MTTLab Team of the University of Brescia and the Consejo Superior de Investigaciones Científicas of Madrid.

In addition, I hereby state that in the first chapter, the development, integration and testing of the camera-based markerless system for the evaluation of the human pose has been accomplished in collaboration with Clara Sanz, Alessandro Luchetti and Davide Massella. In the second Chapter, the acquisition system for the estimation of mass and inertia of the anatomical segments has been designed, developed and coded together with Alessandro Luchetti, Luca Maule, Paolo Tomasin, Matteo Lancini and Damiano Fruet. In the third chapter, the conditioning system has been designed and developed in collaboration with the company Selmo ing. Antonio Franco, T4I and that the whole project has been pursued in collaboration with Irene Sorianini, Federico Gala, Alessandro Torrisi, Matteo Nardello and prof. Davide Brunelli. In the last chapter, the development of the augmented reality environment and the testing of the final application were conducted and evaluated by Davide Massella, Alberto Vacilotto, Clara Sanz and David Pinto.

Finally, I declare that part of my research activities has received funding from the European Union's Horizon 2020 research and innovation programme, via an Open Call issued and executed under Project EUROBENCH (grant agreement N°779963). The activities were approved by the ethics committee "EUROBENCH: European Robotic framework for bipedal locomotion benchmarking - Validation experiments" with reference to the H2020 call, under the internal code 091/2021.

Abstract

This thesis research proposes a set of technical systems which support the biomechanical analysis and it is structured as four connected sub-projects. First, it proposes a markerless motion capture system which estimates the human pose via Time of Flight Cameras. Secondly, it is illustrated the design, development and validation of a cutting-edge RGB-D camera-based system which autonomously estimates the volume, weight and the Inertia of the human anatomical segments. Subsequently, it reports the development of custom-made electromyographic sensors which perform real-time monitoring and data processing of the surface muscle activation. Finally, it proposes a mixed reality application for returning real-time visual feedback both for assisting the test subjects and for augmenting the physician's eye.

In particular, the first part of this work proposes a comparison between two motion capture systems. The study compares a customised breakthrough markerless multi-camera system, Azure Kinect based, and a gold standard Marker-Based optical tracking System, Vicon Nexus. On one hand, a Kalman Filter sensor fusion collects each joint information provided by each Azure Kinect and merges the measures considering the uncertainty related to each camera. On the other hand, the Vicon motion capture uses a plug-in gait model with a standard subject marketization to reconstruct the position of each joint. A further innovative aspect is the peculiar focus on subjects wearing exoskeletons for lower limbs, allowing to highlight the performances and limitations of both systems in presence of noisy environments and occlusions.

The computation of volume, weight, and inertia of each anatomical segment has been conducted within the Bullet project as a part of the Eurobench European framework. The goal of this project aims to establish the reaction forces generated within the articulations when a paraplegic subject walks with a lower limb's exoskeleton and uses the crutches for balancing and support. The anatomical parameters are therefore required to perform biomechanical analysis and run simulations. With respect to the tools and methods reported by the state of the art, here is proposed a novel approach which autonomously estimates the volumes, weight, and inertia from the RGB and depth images of the subjects laying on a hospital bed.

The monitoring of the muscle activation aims to develop and optimize a custom device for offline and real-time EMG applications, such as the visual feedback of muscular activation in mixed reality or the remote actuation of a robotic prosthesis. For this reason, the analysis has been pursued both by using commercial devices, such as Delsys Trigno, and also designing, developing, and validating a custom device for electromyography.

Finally, this work reports the development of mixed reality environments through HoloLens 2 devices for the visualization and rendering of the measurements. This information is supposed to help both the therapist and the patient address the rehabilitative exercises more intuitively and objectively. In fact, on one hand, this allows the therapist to have a fast real-time evaluation of the patient's parameters without the need to consult paper documents or nearby monitors. On the other hand, it helps the patient by exploiting ghost avatars or geometrical guideline indicators that compare the subject's movements with the correct ones.

Table of contents

List of figures	xiii
List of tables	xvii
Introduction	xix
1 The Video-Based Markerless Evaluation Of Human Pose	1
1.1 The Evaluation Of Human Pose and Motion	1
1.2 Vicon	3
1.3 Azure Kinect	3
1.4 Set up and structure the Tests	5
1.5 Calibration	8
1.6 Data collection and processing	10
1.6.1 The Kalman Filter sensor fusion	13
1.6.2 The model	14
1.7 Results	17
1.8 Conclusions	20
2 Estimation of the mass and inertia of the anatomical segments	25
2.1 Traditional approaches for the measurement of anthropometric parameters	25
2.2 Technologies for the 3D virtual reconstruction	26
2.3 The Bullet project	28
2.4 Materials and Set up	29
2.5 Data Collection and Processing	32
2.6 Methods for the volume estimation from a surface point cloud of 3D objects	36
2.7 The Monte Carlo-based volume estimation by exploding local cubes faces of a 3D surface point cloud	38
2.8 Results of the volume estimation algorithm	41
2.9 Computation of Mass and Inertia	43

2.10	Conclusions	45
3	sEMG Sensor for monitoring the muscle activation and for Action Recognition	49
3.1	The Electromyography	49
3.2	Action recognition in the clinical Environment	52
3.3	Action recognition	53
3.4	Hardware architecture and set-up	55
3.4.1	Conditioning system	57
3.4.2	ADC and Processing Unit	59
3.5	Deep Neural Network Architecture	59
3.6	Generation of Dataset and Training	62
3.7	Results	64
3.8	Conclusions	66
4	Visual feedback through augmented virtuality and augmented reality	71
4.1	Mixed reality in the rehabilitative environment	71
4.2	Mixed reality for exoskeleton walking	74
4.3	Set up and material	74
4.4	Virtual Environment Set Up	76
4.4.1	Avatars Animation and motion control	76
4.4.2	The subject avatar and the ghost avatar	78
4.5	Acquisition and processing at 30 fps data display	78
4.6	Test and trials	80
4.7	Postprocessing and skeleton comparison	81
4.7.1	The Spatial-temporal gait parameters	81
4.7.2	The Normality Checks	82
4.7.3	The two-sample T-tests	83
4.8	Results	84
4.9	Conclusions	86
	Conclusions	i
	References	v
	Appendix	xvii
	Appendix	xxi

List of figures

1	Table of Contents - The red box refers to the content of chapter one, the blue box refers to the content of chapters two and three, and the yellow box refers to the content of chapter four.	xxii
1.1	Vicon 3D multicamera view of the reflective Markers	4
1.3	Frontal visualization of the Azure Kinect skeleton with joint labelling	7
1.4	multi-perspective view of unassisted walking on a treadmill returned by azure Kinect multicamera system: a) view camera 1, b)view camera 2, c) view camera 3, d) virtual reconstruction of the point cloud with Vicon skeleton (red) and AzureKinect (black) in overlay	9
1.5	multi-perspective view of walking with exoskeleton and walker without the treadmill returned by azure Kinect multicamera system: a) view camera 1, b)view camera 2, c) view camera 3, d) virtual reconstruction of the 3D point cloud with Azure Kinect skeleton (in black) in overlay	10
1.6	Dynamic Calibration of Vicon Cameras	11
1.7	Points selection (blue stars on the yellow cross frame) in the RGB image for the first Azure Kinect Calibration	12
1.8	Pointcloud representation of the 3D remapping in the depth image of the 2D points (in red) selected in the RGB image	12
1.9	3D Pointcloud representation of the first Azure Kinect field of view with respect to the new reference frame	13
1.10	Display of 3D overall point cloud and the Skeletons reconstructed from the three Azure Kinects	14
1.11	Joint to Joint Euclidean Distance (reported in millimetres and coloured in red) between the Vicon marker-based skeleton reconstruction (in black) and the markerless Azure Kinect system skeleton reconstruction (in blue)	17
1.12	Graph of all the Joint to Joint Euclidean Distance between the marker-based Vicon skeleton reconstruction and a single Azure Kinect Skeleton reconstruction	19

1.13	Joint to Joint Euclidean Distance between the Vicon skeleton reconstruction and a single Azure Kinect Skeleton Reconstruction	20
1.14	Graph of all the Joint to Joint Euclidean Distance between the marker-based Vicon skeleton reconstruction and the markerless Azure Kinect Skeleton reconstruction with the Kalman Filter sensor fusion	21
1.15	Euclidean joint-to-joint distance of a subject walking without an exoskeleton, crutches and orthopaedic walker support between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction	21
1.16	Euclidean joint-to-joint distance of a subject walking with a lower limb exoskeleton between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction	22
1.17	Euclidean joint-to-joint distance of a subject walking with an orthopaedic walker between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction	23
1.18	Euclidean joint-to-joint distance of a subject walking with the support of crutches between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction	24
2.1	Measurement with the tailor's tape for the comparison of the methods . . .	26
2.2	Test of the Bullet project during exoskeleton-assisted walking with sensorised crutches and pressure matrices	29
2.3	Volumetric Scanner system Set up composed of the hospital bed, the horizontal wood frame with low deformation mattress, the vertical aluminium frame and the ToF Camera Azure Kinect	30
2.4	Subject on the Hospital bed ready for the acquisition	32
2.5	Cloud of dots of the subject lying on the hospital bed	33
2.6	Plane of the Hospital bed found by the Ransac algorithm	34
2.7	Subject isolated from the surrounding environment	34
2.9	Open Pose Skeleton merged to the 3D Point Cloud	35
2.13	Extrusion of one of the faces of the cube and interception of two clusters of points.	41
2.14	Comparison between the surface point cloud of the 3D object(left), inner point cloud (centre) outer point cloud (right)	41
2.15	Box plot of the Volume computation depending on the amount of generated points	42
2.16	Box plot of the computational time depending on the amount of generated points	43

2.17	Cloud of points on a subject's chest surface reoriented with principal component analysis and expressed with respect to a reference system with axes parallel to the three orthogonal directions of maximum variance and with origin in the centre of gravity.	44
3.1	Generic sEMG signal acquisition pipeline	51
3.2	Testing of different geometries and electrode arrangements to maximise the signal-to-noise ratio	56
3.3	Prototype of the 8-channel EMG acquisition system	57
3.4	Acquisition tests wearing the prototype of the developed system	58
3.5	Conditioning system architecture, designed and developed by Ing. Antonio Selmo	59
3.6	Comparison on the results obtained by training different types of neural networks with the same sub-portion of the EMG open-source Nina Pro dataset of 8 forearm muscles	62
3.7	LSTM Network Architecture	63
3.8	Confusion matrix obtained by training and testing the LSTM neural network	65
3.9	Ensemble of specialized LSTM neural networks	66
3.10	Comparison of results obtained with specialised LSTM neural network ensemble and dataset augmentation	67
3.11	Acquisition and classification of EMG signals in real-time for the control of a robotic limb	69
4.1	3D Augmented Virtuality reproduction with HTC-VIVE of Ausilia domotic apartment with point clouds of the patient and the therapists. In the background on the left side are the graphs of the toilet pressure sensors and on the right side are the graphs of the patient's heartbeat and breathing.	72
4.2	Set up consists of six Vicon cameras, three Azure Kinect cameras, a treadmill and an AsusTUF Gaming Fx504 series laptop	75
4.3	Introduction Scene	76
4.4	Test scene with the two avatars. On the left side, there is the subject avatar, that reproduces the user's movements in real-time. On the other side, there is the ghost avatar, that reproduces the recorded movements of an expert pilot of robotic exoskeletons	77
4.5	Generation of the avatar's musculature in the blender environment	79
4.6	User's view during walking trials wearing Microsoft Hololens 2	80

4.7	Set up acquisition with the test subject wearing Hololens 2 to emulate the movement of the expert user in exoskeleton walking. The colours highlight the height of the subject and some spatiotemporal parameters used for gait assessment.	82
4.8	In red, the cadence of the expert user, and in black the cadence of the test subject using Hololens	83
4.9	Q-Q plot of the distribution obtained by calculating the maximum knee angle reached during each gait cycle	84
4.10	Maximum Knee Angle - Histogram versus Gaussian distribution comparison	85
4.11	Maximum Knee Angles Statistical Comparison using the two-sample t-tests - On the right, the distribution of the spatiotemporal parameter of an experienced user is shown in green. On the left, the distribution of the subject without AR support is reported in yellow.	86
4.12	Maximum Knee Angles graph Comparison between the expert user (in black), the subject supported with (in blue) and without (in red) AR tools	87
4.13	Comparison of the cadence distributions of the expert user (in black) with the cadence distribution of the subject with Hololens (in blue) and the subject without Hololens (in red)	87

List of tables

- 1.1 Azure Kinect Joint Hierarchy 6
- 2.1 Results Comparison with Virtual Objects 46
- 2.2 Results Comparison with real Objects 47

Introduction

The opportunity to take advantage of effective procedures, reliable measurements and intuitive tools brings several advantages to the biomechanical analysis. It also allows a more conscious assessment of the phenomenon under study having the opportunity to work with a greater number of tests and analyze a larger amount of data.

So far, generic analysis of human motion is structured as a sequence of standard procedures consisting in the preparation of the subject, measurement of anthropometric parameters, data collection and processing, and the display of the final results [118, 75].

Taking into account the work reported in the state of the art, it can be seen that subject preparation and data collection typically take place through widely consolidated procedures. However, it is also possible to observe that these procedures may result slow, inefficient and time-consuming, as they require great manual work accomplished by specialized staff acknowledged in biomechanics[49]. In fact, the measurement of anthropometric parameters, such as the subject's weight, height, anatomical circumferences and the distances between retere points, requires a good knowledge of the human body, as it is often pursued manually with the support of anthropometric tables[34]. In addition, the use of anthropometric references and the geometric reconstruction to estimate volumes, masses, and inertias of the anatomical segments requires a solid knowledge of technical aspects related the biomechanical field. This approach not only makes data acquisition extremely slow, but also makes it impossible to determine the uncertainty associated with the measurement. Even if a highly educated staff is involved, the uncertainty will depend on the attentiveness, the competence, as well as the tiredness and stress condition of the operator who is carrying out the measurements. Not only it can jeopardise the entire acquisition chain, but there is also no easy way to detect it. On the contrary, the possibility to automate the measurement of anthropometric parameters by leveraging reverse engineering and 3D scanning technologies with sensors of known precision and accuracy allows to reduce working time and obtain consistent measurements with a more objective estimate of uncertainty. Moreover, in order to pursue high standard tests and properly reconstruct the human motion over time, the ability to deal with marker-based systems is required [49]. Marker-based systems record the trajectory

of the reflective markers placed on the subject's body. Again, the placement of the reflective markers requires a qualified personnel able to deal with challenging complications and contingencies. In fact, marker-based systems can be quite unrobust in certain applications, such as in presence of robotic exoskeletons and crutches. In addition, most of the time the related issues can not be detected until the offline post-processing takes place, along with the subject's pose is reconstructed[82]. In this case, an in-depth knowledge of the measurement system is required to prevent acquisition failures or compensate for data leaks. For example, the pose estimation of a subject walking with a robotic exoskeleton often requires to deal with different redundant sensors, since the reflective markers may be hidden, hit, or even detached during the movement of the subject[82]. Therefore, in addition to the large amount of time required for test preparation and marker placement, it is also necessary to have specialized and competent personnel to keep checking the goodness of the data, making sure that they are not corrupted or even unusable. Instead, the possibility of exploiting the new markerless technologies would considerably streamline the preparation process. This allows to carry out a greater number of tests in the same amount of time and to pay more attention the data analysis. Moreover, the presence of specialized personnel is no longer necessary both for the positioning of the markers and the permanent control of their visibility condition.

Considering the processing and data consultation, the gold standard protocols still rely on offline data processing and the consultation is mainly presented in the form of descriptive reports, tables and graphs. Similarly, in the case of a clinical evaluation for rehabilitation purposes, these results provide qualitative feedback on the subject's conditions and may lack quantitative information to make appropriate comparisons. Moreover, the offline data consultation introduce a great delay in the clinical evaluation and may results to be untimely to guide the therapist while checking and selecting the most suitable therapy for the patient.

Especially in the field of rehabilitation, having the possibility to consult the data on-site, instead, enables a direct comparison between the clinician's perception (related to his/her expertise) and the information provided, leading to notice relevant details that otherwise would be missed with an offline standard analysis [105],[37]. On this front, the emerging Mixed Reality technologies enable the development of useful, versatile and intuitive tools to improve an efficient and more objective data consultation [30, 37, 105]. Mixed reality is a general term describing an environmental merge of digital and physical worlds. By imagining a continuous transition from the real to the virtual world we can define the Augmented Virtuality (VR) and Augmented Reality (AR). AR presents the vision of the real environment with a superimposition of augmented virtual objects. In contrast, VR shows a virtual environment enriched by the information collected from the real world.

On one hand, such data visualisation support the therapist while assessing the subject's movement during the rehabilitative exercises[30]. For example, AR tools, if correctly supported by a real-time EMG system for estimating the muscle activation and a real-time motion capture device for estimating the subject pose, can provide the therapist an intuitive and dynamic visualization of the locomotor apparatus. This would allow a careful clinical eye to analyze each movement with greater awareness, evaluating in real-time symmetries/asymmetries of forces, the subject effort, and the immediate check of the involved muscles. Moreover, by exploiting innovative Artificial intelligence tools, e.g. neural networks, and computationally efficient algorithms applied on EMG signals, it is possible to pursue real-time data processing to return more complex information than raw data, such as automatic classification of actions, subdivision of movements into phases[127] and quantitative assessments.

Such a data can also be saved and consulted offline. As developed at the Pergine rehabilitation hospital Villa Rosa within the Ausilia project[27, 85]. By exploiting VR technologies, it is possible to render saved information and consult patient data offline by visualizing them on-demand within a reconstructed virtual environment. In this case, using wearable devices such as HTC-Vive, the clinician can consult the registered patient's data in a synchronized and parallel viewing mode. Visualizing the patient through a 3D virtual point cloud, the therapists can observe and compare the monitored parameters as if he/she were in presence. The therapist is also able to stop and proceed the animation as winding and rewinding a 3D video, by pausing to evaluate a particular moment as much as needed. The environment, the patient's point cloud and the data reported on the graphs will pause and proceed accordingly.

Considering the subject's point of view, AR technologies can act as a guide and support during experimental tests and the execution of exercises. In addition, an intuitive and simple visualization of the information, such as with a color scale, interactive graphs, or dynamic bar charts, can also assist the subject along the movement he/she is trying to pursue[85]. In this way, the subject can modify the movement according to the AR visual feedback, leading to an easier and better execution of the exercise. For example, in the case of the biomechanical analysis of subjects walking with exoskeletons, usually it is necessary to train experienced users to emulate the correct movements. The introduction of AR technologies may help the user during the learning phase and give support to achieve a smoother simulation of the actual walk. Moreover, this approach allows to reproduce the walk with an exoskeleton even without wearing it. This significantly streamlines the preparation procedure and reduces the risks and complexities associated with the early trials of exoskeleton. AR technologies may result to be more effective than standard protocols due to the ability to turn the rehabilitation

exercises into a playful activity, shifting the subject's attention to more engaging aspects and keeping the patient actively focused. This aspect has already been introduced through traditional video games with on-screen feedback. In the case of augmented reality, it is increasingly emphasized thanks to new immersive technologies that allow a more realistic and intuitive experience of the virtual world.

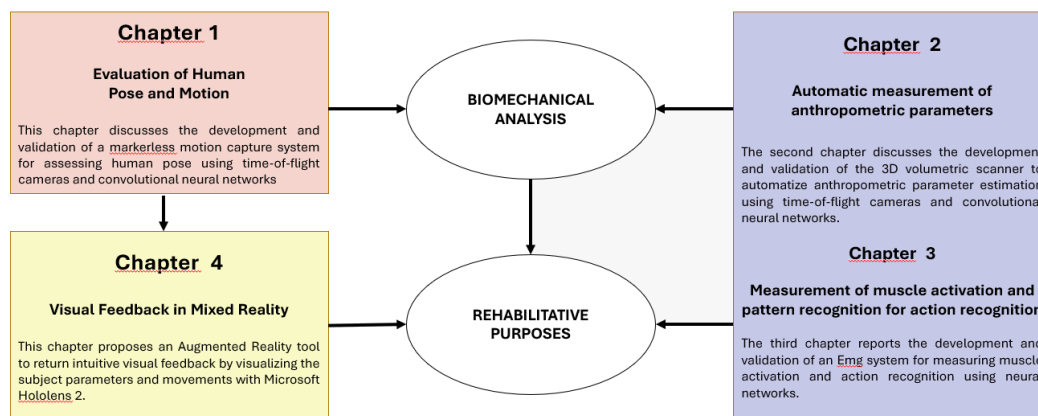


Fig. 1 Table of Contents - The red box refers to the content of chapter one, the blue box refers to the content of chapters two and three, and the yellow box refers to the content of chapter four.

For these reasons, as reported in figure 1 this research work focuses on measurement and visualization tools in the field of biomechanical analysis. The work considers the state-of-the-art technologies, sensors and methods to meet the main scientific requirements and proposes novel approaches and systems to deal with the most common issues. The first chapter addresses the issues related to the gold standard motion capture systems and proposes an innovative marker-less systems to streamline and lean the human pose acquisition process. The second chapter focuses on the measurement of anthropometric parameters illustrating the state of the art procedures and proposing an autonomous scanning system to evaluate volumes, masses and inertias of the anatomical segments. Following, the third and the fourth chapters focus on the real-time data acquisition, elaboration and visualization for the online data consultation. In particular the third chapter describes the development of a custom EMG sensor and machine learning algorithms for the real-time collection and recognition of muscular patterns. Finally, the fourth chapter groups together the systems described in the first and the third chapters to illustrate the development of an AR environment for the real-time data visualization of the human motion and muscular activity.

Chapter 1

The Video-Based Markerless Evaluation Of Human Pose

1.1 The Evaluation Of Human Pose and Motion

The importance of human posture monitoring and action recognition is a key factor in the field of biomechanics, rehabilitation, and support for people with disabilities[59]. In particular, considering subjects with disabilities who require exoskeletons for lower limb locomotion, the motion capture technology allows several applications in clinical environments, such as estimating internal forces on joints during movement, studying load distribution while walking with an exoskeleton, and analyzing performance during training[82]. In the case of elderly individuals who live alone instead, this technology allows support to the person during the most common daily activities, recognizing the pose and acting accordingly for the good of the subject [27], [25], [119].

Several technologies are available to estimate the subject pose to perform kinematic and biomechanical analysis. For example, in the clinical environment, for the study of the gait cycle, the Marker-based optical tracking systems (MBS) are now considered the gold standard to monitor the movement and acquire the trajectories of the joints. These measurement systems, such as Optitrack, Vicon, Phase Space, or Smart-DX track the movement of markers positioned on the subject's body and reconstruct his pose through algorithms based on biomechanical models, such as the plug-in gait of Vicon. On the one hand, these devices allow accurate and reliable measurement of human pose, but on the other hand, they are expensive, time-consuming, and quite difficult to use. Being mainly commercial devices, they present a user-friendly interface, excellent processing capabilities, and a very low difficulty of post-processing by the user. However, they are particularly expensive depending on the

number of cameras and the number of functional packages required, preventing smaller research entities from being able to take advantage of them. In addition, they require special conditions and specialized personnel to place the markers. For example, in the case of Vicon, the subject must be partially undressed or wear special clothing to allow placement of the markers (reflective beads on an adhesive base) and the operator must wear latex gloves to handle the silver beads and preserve their reflective capability. Furthermore, if not positioned correctly, the markers are subject to detachment and occlusions. In fact, in the specific case of subjects wearing robotic exoskeletons, they are covered or hit by the robotic frame causing them to fall or reduce visibility by the cameras.

Inertial measurement units (IMU) have been widely investigated for biomechanical applications, such as gait analysis[60] [94] body orientation and pose estimate [28] [47], sport performance assessment [53] [70] [132], fall detection of elderly subjects[96], [97], gaming [6] and control of robotic prostheses for rehabilitation purposes [116] [21]. While these sensors are particularly suitable for motion capture due to their small size, affordable cost, high accuracy, ease of use and wear, on the other hand, they are affected by drift and require integration with other sensors to mitigate the growing error [130] [92]. Moreover, the introduction of further sensors implies an increase in the complexity of the system and its management. An alternative technology proposes the use of markerless motion capture systems (MLS). This recent approach uses artificial intelligence algorithms to extrapolate the position of the joint from RGB or depth images. In particular, some algorithms such as Open Pose, Open Vino or Mediapipe, exploit the convolution neural networks (CNN) to extract from the RGB images the 2D joint coordinates of the anatomical segments. Similarly, still using CNN, some time-of-flight (Tof) cameras such as Kinect 2 or Azure Kinect, can determine the position of the anatomical joints in the 3D space. Because these technologies are still under development, their accuracy is limited by the AI algorithm that processes the information and by the goodness of the sensor that acquires the image[122]. They also require a remarkable computational power per frame implying further limitations on real-time applications due to the low frame rate and the low capability to capture rapid movements. However, while it is particularly well suited for off-line biomechanical analysis due to its affordable cost, absence of marker placement, and automatic calculation of anatomical joints [23], how it works in the presence of exoskeletons and occlusions is not yet well defined. The CNN has been trained on healthy subjects dressed in everyday clothes, and although its performance is known to depend strongly on several chromatic and geometric factors, no in-depth study has been conducted for gait cycle analysis using robotic exoskeletons.

For these reasons, the first chapter of this work proposes the acquisition with a multicamera markerless system of a subject walking with an exoskeleton and compares the

measurement with the gold standard MBS system, Vicon. The multi-camera markerless system is composed of 3 Azure Kinects that estimate the position of the joints via convolutional neural networks (CNN) and a Kalman Filter sensor fusion algorithm [82] to merge the measurements and reduce the uncertainty.

1.2 Vicon

Vicon is an MBS for human pose tracking. Nowadays considered one of the gold standards for the analysis of human kinematics and movement, it allows following the trajectory of reflective markers with high accuracy and reliability. Through software programs, such as Nexus 2.12¹, based on biomechanical models it is possible to reconstruct the trajectory of the human skeletal joints. Due to its high reputation and reliability, Vicon, like many other MBS, is often used as a reference for the validation of alternative markerless systems, such as Theia [50], [48], [49] or Kinect [4], [106], [82].

However, it does have some downsides. Like all MBS technologies, it requires the placement of markers following standardized mappings, such as the Plug-in gait mapping to allow the skeleton reconstruction software to recognize the position and apply the conversion model. The Plug-in gait is a biomechanical model for calculating joint kinematics. It is commercially available and is based on the Newington-Helen Hayes marker set [22]. There are different types of markers, but all have common recognition characteristics to be seen and tracked by MBS cameras. The markers used with Vicon, for example, consist of a silver reflective sphere mounted on a black non-reflective plastic base. These markers must be handled with latex gloves to avoid depositing skin grease on their surface and thereby compromising their reflective characteristics. They must also be set up by an experienced and knowledgeable staff member who is familiar with the mapping protocols, Figure 1.2b, and be able to proceed with the placement of the marker upon the subject's body, as reported in Figure 1.2a.

1.3 Azure Kinect

Azure Kinect is a time-of-flight (ToF) camera that enables RGB and depth image capture. It is a commercial product from Microsoft developed based on the previous Kinect 1 and Kinect 2 versions. Unlike its predecessors initially developed for gaming applications, Azure Kinect is produced with a greater focus on biomechanics studies and is also designed as a research

¹<https://www.vicon.com/software/nexus/>

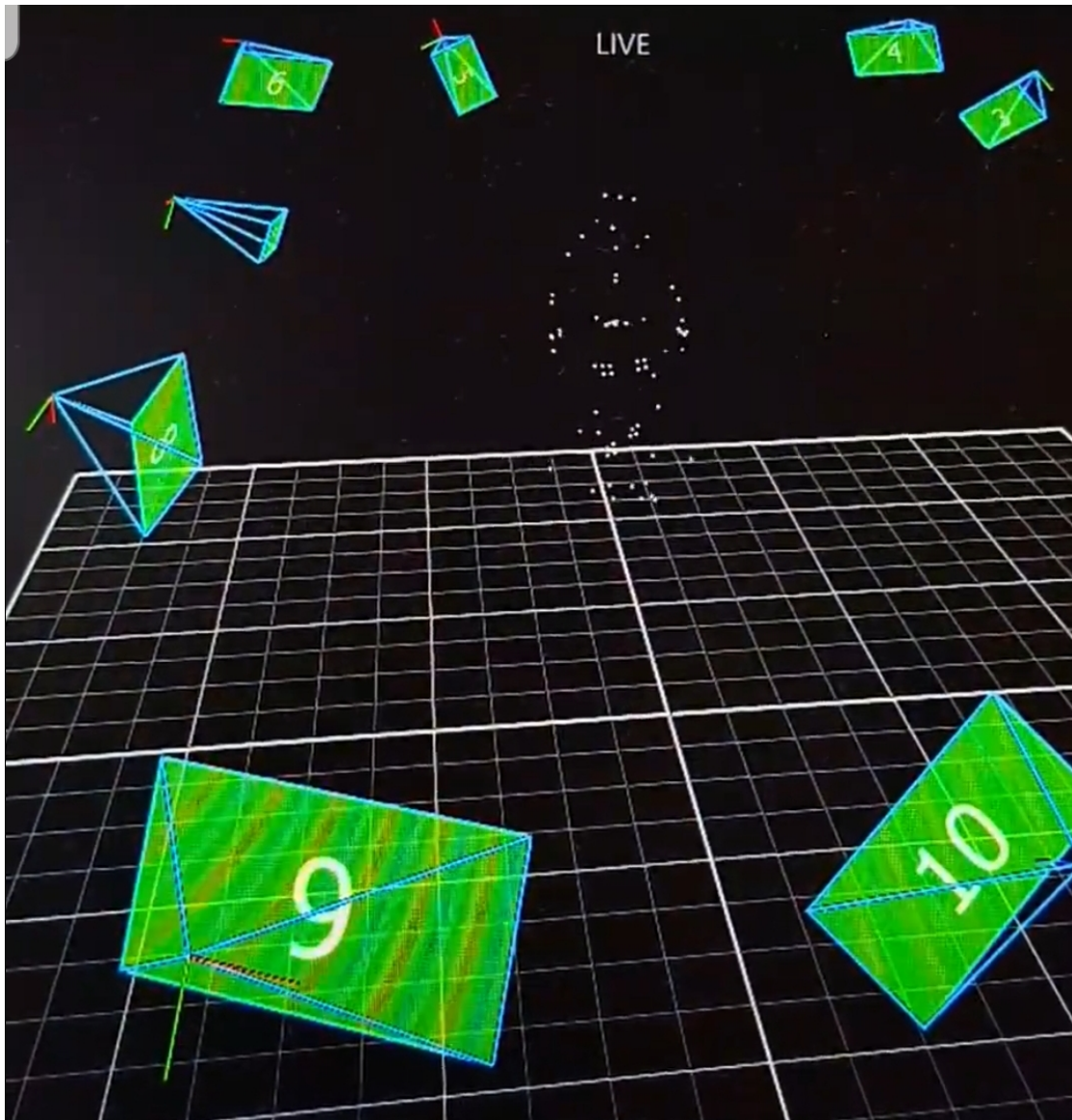
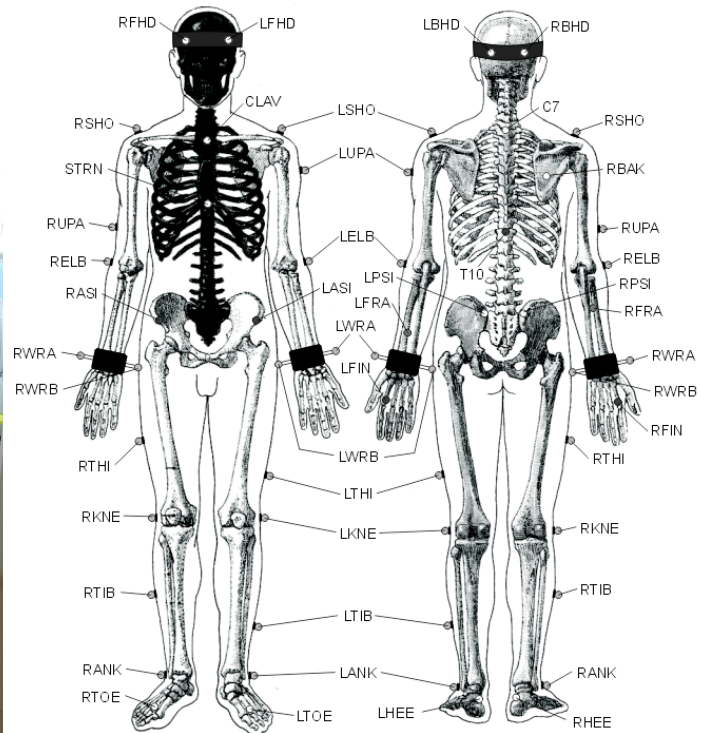


Fig. 1.1 Vicon 3D multicamera view of the reflective Markers

and development tool. For these reasons, it comes with a software development kit (SDK) that allows users to develop their applications based on requirements and tasks. In addition, Microsoft also provides a trained convolutional neural network, Azure Kinect Body Tracker, which takes in input the images captured by the camera and returns the segmentation of the monitored subject and the 3D coordinates of each of its joints. Azure Kinect allows capturing RGB and depth information in video format and post-processing the images by extracting the joints of the body with its CNN and reconstructing the environment as a 3D point cloud. Also, with these tools available, you can set the resolution of the images as well as the acquisition frame rate, up to a maximum of 30 fps. If you use a computer equipped with a



(a) Plug-in Gait marker placement of the reflective markers on a human subject for the Vicon marker-based acquisition system



(b) Plug in Gait marker placement

powerful graphics card, you can then perform real-time data processing, allowing you to use the measured human pose for gaming, serious gaming or augmented reality applications.

The Azure Kinect Skeleton² presents 32 joints which follow a humanoid representation branching from the centre of the body to the distal ends. As shown in the Figure 1.3 and reported in the table 1.1, The skeleton collects all the segments (bones) that connect the proximal joints (parent) to the distal joints (child).

1.4 Set up and structure the Tests

For the comparison between the two motion capture systems, we have chosen to use Vicon and an Azure Kinect multi-camera system. Vicon uses standard V225510 cameras with an OptiTrack marker bases system and processed in post-processing by the Nexus program.

²url: <https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints> 2022

Table 1.1 Azure Kinect Joint Hierarchy

Index	Joint name	Parent joint
0	PELVIS	-
1	SPINE NAVAL	PELVIS
2	SPINE CHEST	SPINE NAVAL
3	NECK	SPINE CHEST
4	CLAVICLE LEFT	SPINE CHEST
5	SHOULDER LEFT	CLAVICLE LEFT
6	ELBOW LEFT	SHOULDER LEFT
7	WRIST LEFT	ELBOW LEFT
8	HAND LEFT	WRIST LEFT
9	HANDTIP LEFT	HAND LEFT
10	THUMB LEFT	WRIST LEFT
11	CLAVICLE RIGHT	SPINE CHEST
12	SHOULDER RIGHT	CLAVICLE RIGHT
13	ELBOW RIGHT	SHOULDER RIGHT
14	WRIST RIGHT	ELBOW RIGHT
15	HAND RIGHT	WRIST RIGHT
16	HANDTIP RIGHT	HAND RIGHT
17	THUMB RIGHT	WRIST RIGHT
18	HIP LEFT	PELVIS
19	KNEE LEFT	HIP LEFT
20	ANKLE LEFT	KNEE LEFT
21	FOOT LEFT	ANKLE LEFT
22	HIP RIGHT	PELVIS
23	KNEE RIGHT	HIP RIGHT
24	ANKLE RIGHT	KNEE RIGHT
25	FOOT RIGHT	ANKLE RIGHT
26	HEAD	NECK
27	NOSE	HEAD
28	EYE LEFT	HEAD
29	EAR LEFT	HEAD
30	EYE RIGHT	HEAD
31	EAR RIGHT	HEAD

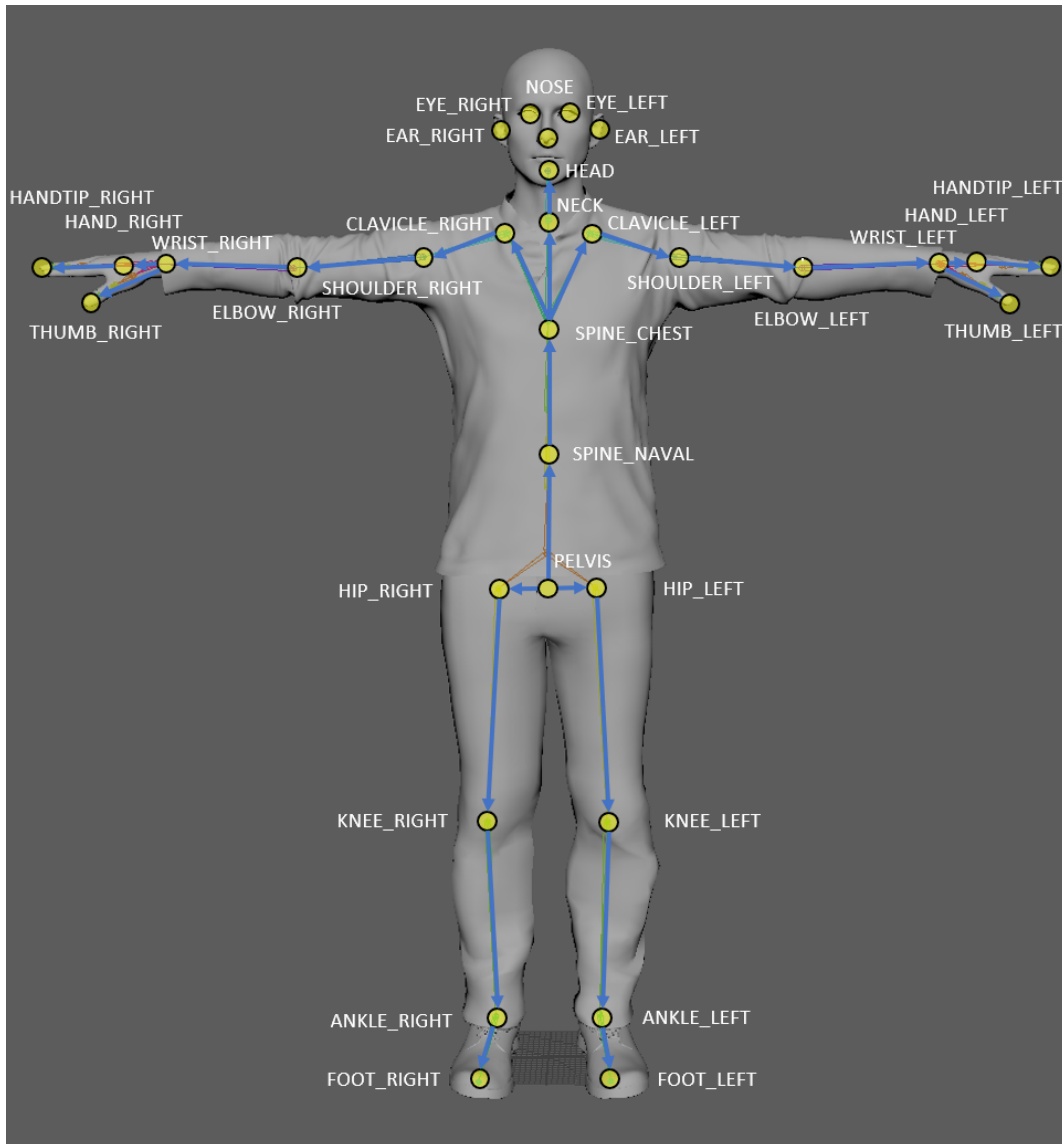


Fig. 1.3 Frontal visualization of the Azure Kinect skeleton with joint labelling

Both the Azure Kinect and the Vicon cameras are arranged around the room with the view directed towards the centre where the subject will perform the walk. The cameras have been positioned to reduce as much as possible interference between the two systems, avoiding that the infrared emissions of Kinect disturb the Vicon measurement and vice versa. The measurement volume is represented by an area five meters long, three meters wide, and two meters high. Each Kinect is mounted on a tripod at a height of 1.8 meters to maximize the field of view. Following the positioning of the cameras, the calibration of the two systems takes place. The calibration operation, partially carried out with the manufacturer's algorithms and partially performed with custom algorithms, returns the rototranslation matrices with respect

to a common reference system allowing to represent of the information of each camera with respect to the same reference system.

The subject will walk in the centre of the room at first without any assistance, later combining the supports of the exoskeleton, crutches, and walker, as partially illustrated in Figure 1.4 and 1.5. The tests, a total of six tests performed with and without a treadmill, are as follows:

1. unassisted walking
2. walking with the exoskeleton
3. walking with crutches
4. walking with a walker
5. walking with exoskeleton and crutches
6. walking with exoskeleton and walker

Acquisition occurs simultaneously for both systems and data are stored on disk for off-line processing. This allows us to take full advantage of the markerless system's acquisition capability without the need for devices with advanced graphics processing units. Each Azure Kinect is connected to an Intel NUC which is not equipped with a dedicated graphics unit. Each NUC device is previously synchronized with all the others and during the acquisition, it records in video format the information acquired by Kinect saving the timestamp since epoch of the beginning and end of each acquisition. Then, in post-processing, each video is elaborated separately allowing to extract the reconstructed 3D point cloud through the depth image and the 3D skeleton of the monitored subject.

1.5 Calibration

Vicon Calibration The calibration of the Vicon system is fully supported by the commercial software provided by the manufacturer. The calibration operation is performed using a black metal cross with red LEDs and is divided into two phases, static and dynamic. Dynamic calibration is performed by swinging the cross frame smoothly and evenly in front of the Vicon cameras as shown in Figure 1.6. The software will thus be able to see from different cameras the same reference object calculating the position of each camera with respect to the others. Static calibration is performed by placing the cross frame approximately in the centre of the viewing space of the cameras. In this way, the Vicon system positions the Global reference system in the intersection of the axes of the cross and orients it parallel to it.

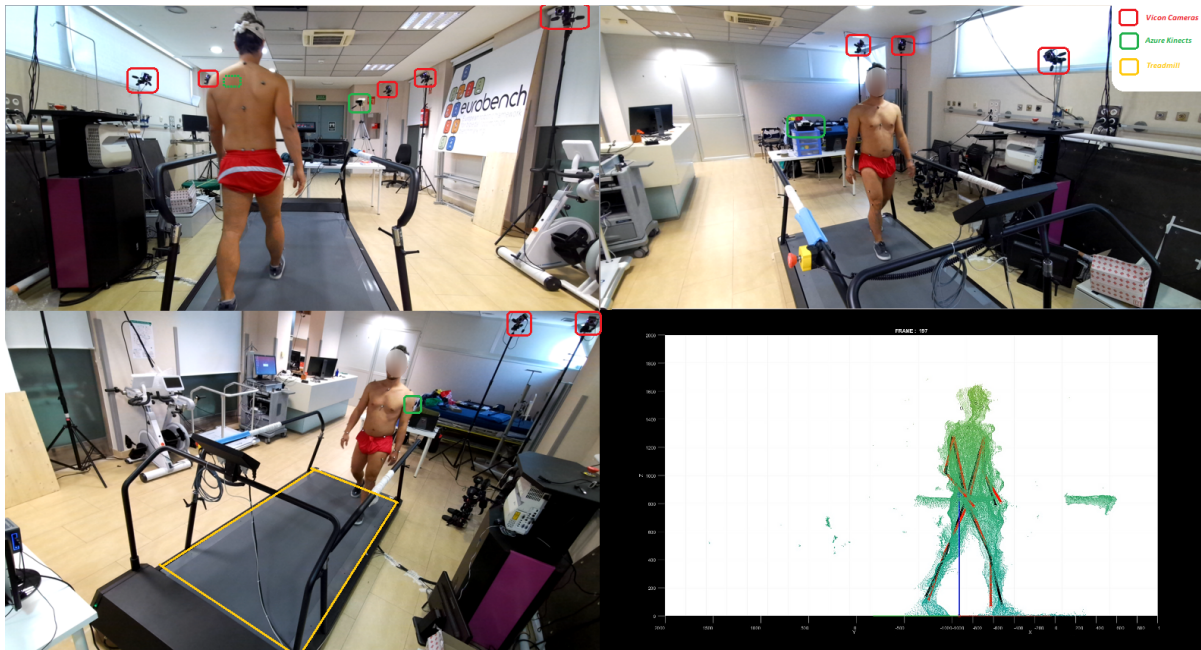


Fig. 1.4 multi-perspective view of unassisted walking on a treadmill returned by azure Kinect multicamera system: a) view camera 1, b) view camera 2, c) view camera 3, d) virtual reconstruction of the point cloud with Vicon skeleton (red) and AzureKinect (black) in overlay

Azure Kinect Calibration For Azure Kinect, however, we implemented a partially automated calibration system that returns a positioned and aligned reference system like that of the Vicon. In this case, only static calibration is performed and is accomplished through the detection of coloured markers placed at the LEDs of the Vicon cross. The same frame cross, in addition to the Vicon calibration, allows the placement of white markers that calibrate the multi Kinect system. A single acquisition by each of the Azure Kinect cameras results in one RGB and one depth image. The algorithm allows for locating the white markers by selecting the corresponding pixels from the RGB image as shown in Figure 1.7. By re-mapping the selected pixels in the depth image it is possible to get the cartesian coordinates x, y , and z of the markers, as reported in Figure 1.8. The markers have been purposely chosen of a reflecting colour to avoid the absorption of the infrared emissions emitted by the Kinect camera and to guarantee the presence of information in the depth image. If the markers had been chosen in black colour the infrared emission would not have bounced back to the camera, consequently producing a lack of information represented as a hole in the depth image.

Once known the spatial coordinates of at least three white markers, through simple vector products it is possible to calculate an orthogonal reference system with axes parallel to the cross frame and with origin in the intersection of the segments of the frame. It



Fig. 1.5 multi-perspective view of walking with exoskeleton and walker without the treadmill returned by azure Kinect multicamera system: a) view camera 1, b) view camera 2, c) view camera 3, d) virtual reconstruction of the 3D point cloud with Azure Kinect skeleton (in black) in overlay

is then possible to calculate for each camera the rototranslation matrix which allows the representation of coordinates of each joint with respect to a common reference system, aligned and superimposed on the Vicon Reference System, as shown in Figure 1.9.

1.6 Data collection and processing

A total amount of two subjects have been collected and the acquired data is saved on local memory. The post-processing is performed with an Asus Tuf Gaming FX504 with Nvidia Ge Force 1050 graphics card and intel core i7 8th generation processor. The post-processing algorithm extracts the RGB and depth images saved in the .mkv video and returns the skeleton of the subject using the Azure Kinect Body Tracker neural network³. Multiplying each Joint returned by the Body Tracker by the rototranslation matrix obtained in the calibration phase, it is possible to express each skeleton with respect to the same reference system, as shown in

³<https://docs.microsoft.com/en-us/azure/kinect-dk/body-sdk-download> (last access 01/01/2022)



Fig. 1.6 Dynamic Calibration of Vicon Cameras

Figure 1.10. As explained during the calibration section of the Kinect system, this reference system also coincides with that of Vicon.

Now that we have the information obtained from each Kinect respectively, it is possible to proceed with the fusion of the measurements. This allows us to obtain a more accurate overall measurement and reduce the uncertainty. For our purposes, we have defined the uncertainty of the single joint as proposed in [82]. In [82], the authors propose an estimate of



Fig. 1.7 Points selection (blue stars on the yellow cross frame) in the RGB image for the first Azure Kinect Calibration

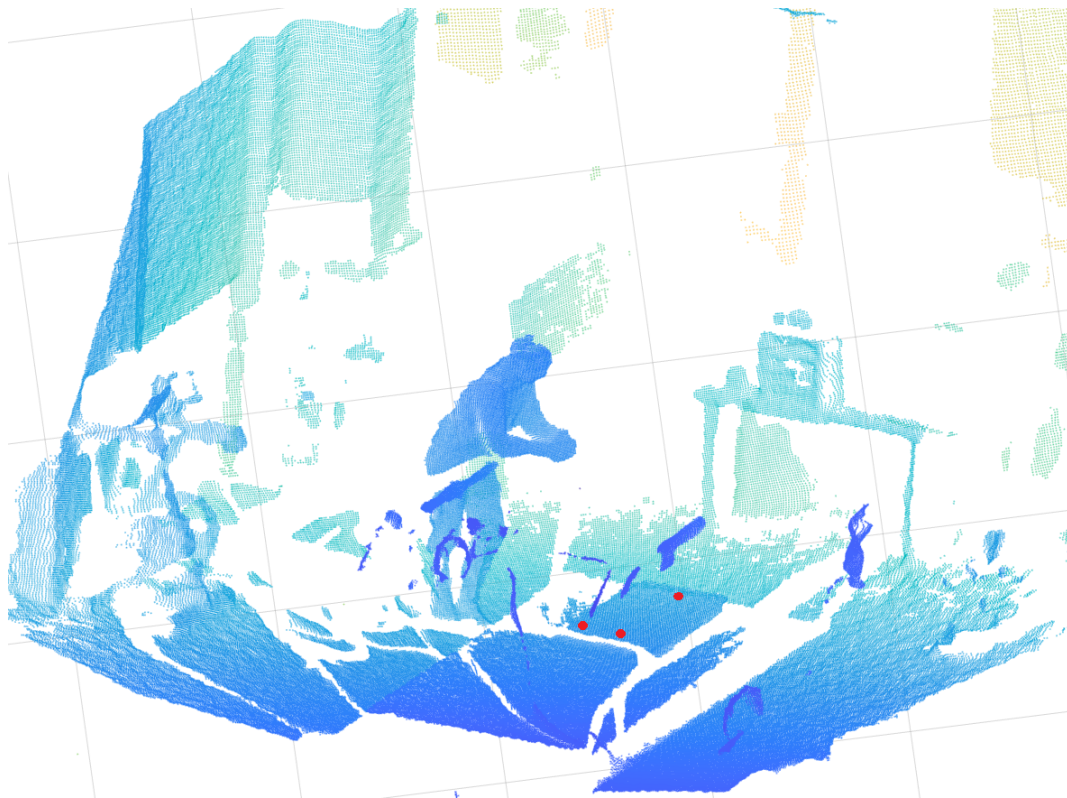


Fig. 1.8 Pointcloud representation of the 3D remapping in the depth image of the 2D points (in red) selected in the RGB image

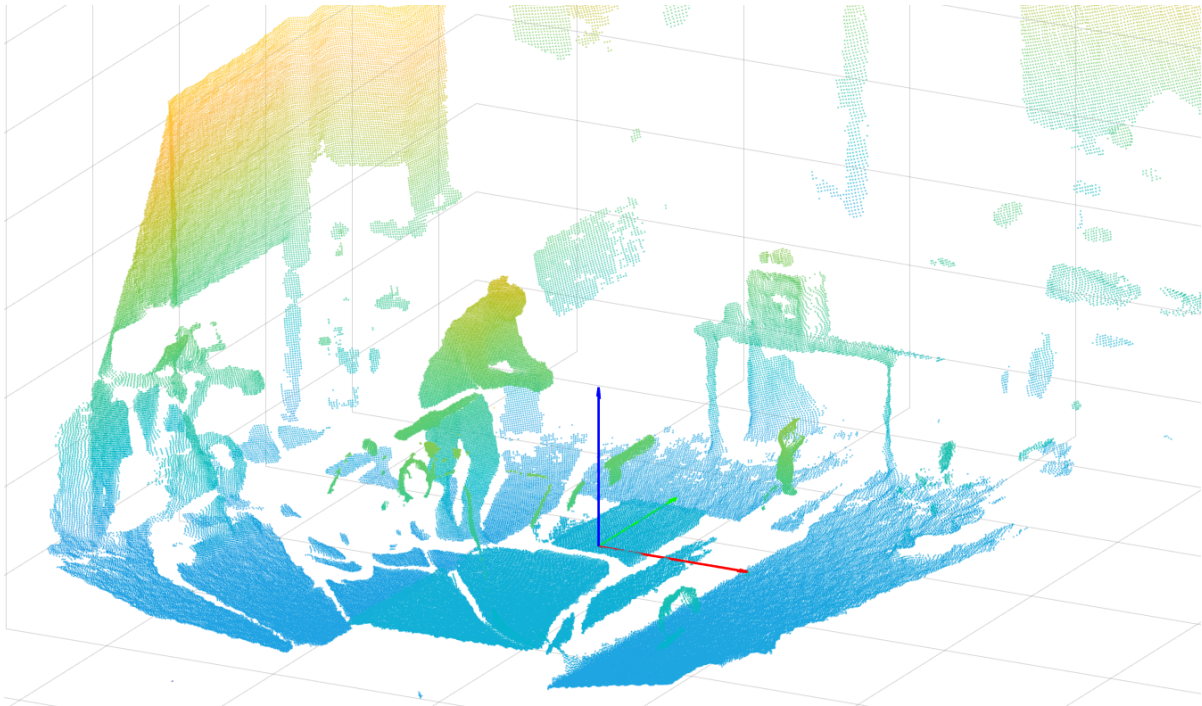


Fig. 1.9 3D Pointcloud representation of the first Azure Kinect field of view with respect to the new reference frame

the uncertainty as a function of the distance from the median skeleton, obtained by calculating the median between the skeletons of the different cameras. In this way, the uncertainty of each joint coordinate is computed as a function of the distance from the corresponding coordinate of the median skeleton. Among the skeletons returned by the markerless multicamera system, it is then possible to extract a median skeleton by collecting all the median joints computed as the median values of the x, y , and z coordinates. By repeating this operation for all coordinates of all joints you can obtain a median representation of the acquired skeleton. This median skeleton allows the calculation of the uncertainty of each joint for each skeleton. The uncertainty associated with a specific joint of a specific skeleton is estimated as the Euclidean distance between the joint and its counterpart in the median skeleton.

1.6.1 The Kalman Filter sensor fusion

As described by Pasinetti et al. [82], by defining a kinematic and dynamic model that governs the skeletal motion, it is possible to introduce a predictive factor within the measurement fusion. We then obtain a sensor fusion model via the Kalman filter. The model used by the filter predicts where each joint will be in the next frame by knowing its position velocity and acceleration conditions from the previous frames. The measurement instead will report the

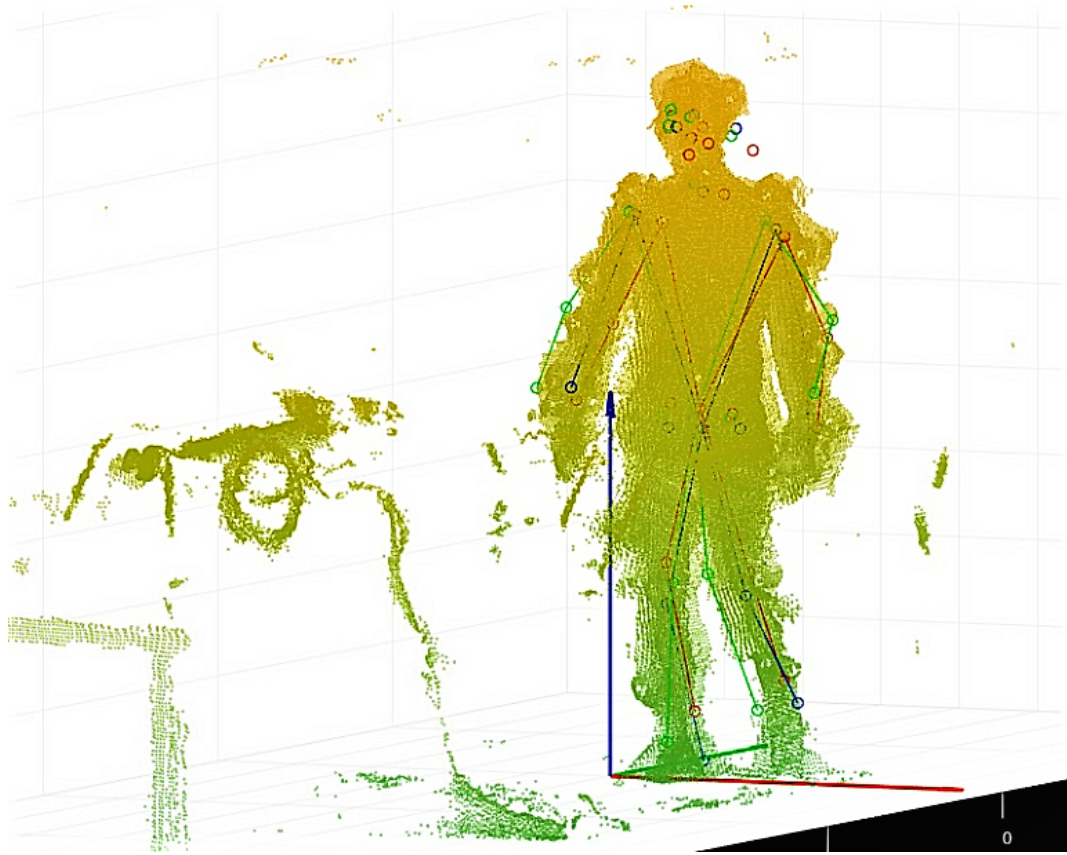


Fig. 1.10 Display of 3D overall point cloud and the Skeletons reconstructed from the three Azure Kinects

value returned by the measurement system. Considering the prediction, the measurement and their related uncertainties a weighted fusion between the two pieces of information takes place.

1.6.2 The model

As described in [82], the post-processing of the markerless system data considers a variable amount of Kinects each to assess the difference in performances depending on the total amount of cameras. Similarly to the work proposed by Pasinetti et al. [82], in general, our model takes into account N nodes and M Azure Kinects. Each node describes a trajectory n described by a vector X^n which reports the position, velocity and acceleration of the node.

$$X^n = [x_p^n, x_v^n, x_a^n] \quad (1.1)$$

Knowing the position, velocity and acceleration of a frame it is possible to predict the position by solving an ordinary differential equation of motion for each one of the Cartesian coordinates and thus obtaining a simple equation for the uniformly accelerated motion :

$$x_k = Ax_{k-1} + w_k \quad (1.2)$$

which in the Matrix has the A matrix equal to the $Diag(A_1, \dots, A_1) \in R^{9N \times 9N}$ where

$$A_1 = \begin{pmatrix} I_3 & I_3\Delta t & I_3\Delta t^2/2 \\ 0 & I_3 & I_3\Delta t \\ 0 & 0 & I_3 \end{pmatrix} \in R^{9N \times 9N} \quad (1.3)$$

The vector x_k reports the 3D position, velocity and acceleration of each skeleton node at the frame k, and Δt is the interval of time within two frames, while w_k is the process noise, represented as a random variable with a normal distribution.

In this way, we can implement the following update equation and the model then returns the a priori estimate of the state \hat{x}_k^- at step k.

$$\hat{x}_k^- = A\hat{x}_{k-1} \quad (1.4)$$

It is then possible to proceed with the estimation of its uncertainty expressed and propagated as follows

$$P_k^- = AP_{k-1}A^T + S_w \quad (1.5)$$

P_k^- is the a priori estimate error covariance while P_{k-1} is the a posteriori estimate error covariance at frame k-1. S_w is the model noise covariance and results to be $\in R^{9N \times 9N}$ as reported in [82]

$$S_{w_n} = \begin{pmatrix} I_3\Delta t^4/4 & I_3\Delta t^3/2 & I_3\Delta t^2/2 \\ I_3\Delta t^3/2 & I_3\Delta t^2 & I_3\Delta t \\ I_3\Delta t^2/2 & I_3\Delta t & I_3 \end{pmatrix} S_{w_i} \quad (1.6)$$

S_{w_i} is a constant and is equal to $0.05^2(m/s^2)^2$ as reported in [76],[13].

Each Azure Kinect returns a skeleton and measures each frame k. The measurement vector $y_k \in R^{3NM}$ reports the positions along x,y, and z of all the joints for each skeleton and is represented by the following model:

$$y_k = Cx_k + v_k \quad (1.7)$$

The matrix $C \in \mathbb{R}^{3NM \times 9N}$ represent the measurement matrix and is created iteratively filling a null matrix with an Identity matrix $I_{3 \times 3}$ in those columns related to the node taken into account and the raw of the related Azure Kinect. The vector x_k is the vector of positions, velocities and accelerations of each node per skeleton. The v_k is the measurement noise and is reported as a random variable. It is independent of the model noise and has a normal distribution.

$$C = \begin{pmatrix} C_1 \\ \vdots \\ C^m \\ \vdots \\ C^M \end{pmatrix} \quad (1.8)$$

where each $C^m \in \mathbb{R}^{3N \times 9N}$

$$C^m = \begin{pmatrix} C_1^m \\ \vdots \\ C_n^m \\ \vdots \\ C_N^m \end{pmatrix} \quad (1.9)$$

Following the measurement update equations are as follows

$$K_k = P_k^- C^T (C P_k^- C^T + S_{v_k})^{-1} \quad (1.10)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (y_k - C \hat{x}_k^-) \quad (1.11)$$

$$P_k = (I - K_k C) P_k^- \quad (1.12)$$

where the state error covariance matrix is

$$P = \text{Diag}(P_1, \dots, P_n, \dots, P_N) \quad (1.13)$$

and each P_n is

$$P_n = \begin{pmatrix} \sigma_p^2 & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & \sigma_v^2 & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & \sigma_a^2 \end{pmatrix} \in \mathbb{R}^{9N \times 9N} \quad (1.14)$$

The $\sigma_p^2, \sigma_v^2, \sigma_a^2$ parameters are respectively the variances of the position, velocity and acceleration of the n^{th} node.

1.7 Results

The results obtained with the markerless 3D system were compared with the Vicon marker-based system. The comparison proposes an analysis of the skeletons reported by the two different systems by calculating the distance between each joint of one model and its counterpart reported in the other model. As can be shown in the Figure 1.11, when the skeleton detected by Kinect (illustrated in blue) deviates considerably from the skeleton of Vicon (illustrated in black), the distance between the same joints increases. An example of this can be observed in the sub-figure further to the right which reports huge errors. In this case, the error is expressed as the distances between the corresponding joints of the two skeletons. This distance is visualized by a red line that links the related joints. These lines should have zero length if the reported skeletons coincide and the error is null. In contrast, if incorrect positioning of the joints occurs, a red segment appears with a length proportional to the error.

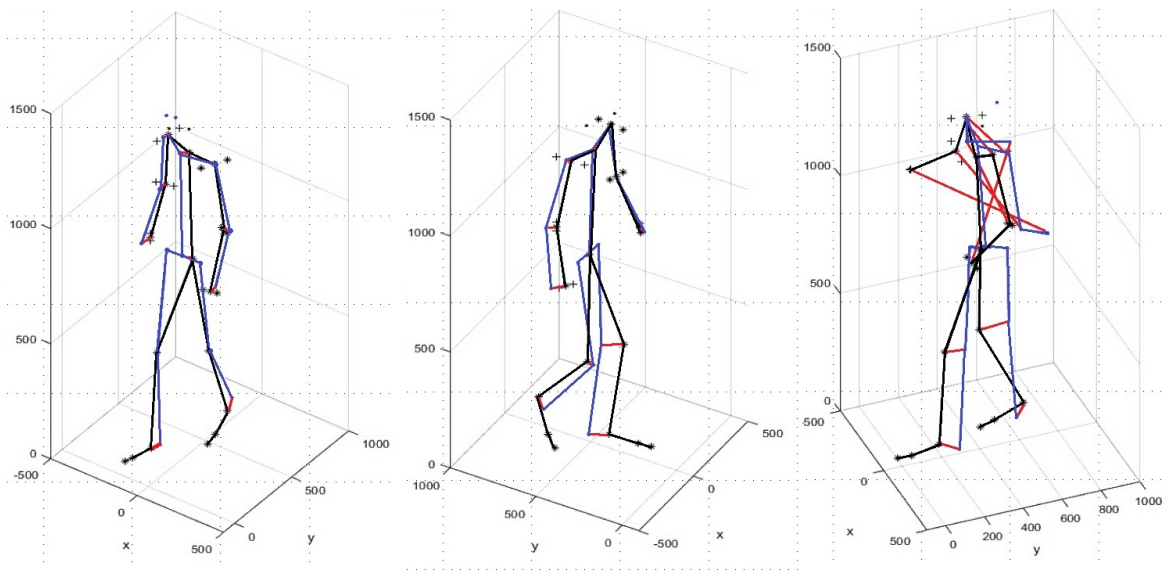


Fig. 1.11 Joint to Joint Euclidean Distance (reported in millimetres and coloured in red) between the Vicon marker-based skeleton reconstruction (in black) and the markerless Azure Kinect system skeleton reconstruction (in blue)

The following eleven joints were taken into account for the comparison:

1. Pelvis
2. Right Knee

3. Right Ankle
4. Left Knee
5. Left Ankle
6. Right Shoulder
7. Right Elbow
8. Right Wrist (reported also as Right Hand)
9. Left Shoulder
10. Left Elbow
11. Left Wrist (reported also as Left Hand)

The Figure1.11 shows the two skeletons of the basic Vicon marker system, in black, and the Markerless system, in blue. The distance segments joining the respective joints of the two skeletons are shown in red. When large errors occur, such as reversal of skeleton orientation (right is confused with left due to an error in the neural network) or glitching, these red segments elongate proportionally to the difference between the corresponding joints.

During each test, all the distances were saved as shown in Figure1.12 and 1.14. In Figure1.12 is reported a generic test with one single Azure Kinect. You can see the large error that occurs between frame 45 and frame 70. During these frames, the subject raised his hands leaving the camera's field of view. The missing information then affected the Azure Kinect body tracker and caused the error to increase. For the remaining frames, the subject remained within Kinect's field of view at all times, but some glitches still occurred, as can be seen with the spike in frame 145. As part of this, the average error never exceeded 10 cm.

A further analysis returns the box plot error for each joint, previously reported in Figure1.14 as a simple graph. The Figure1.13 instead reports the statistical values that determine the distribution of each joint error. The graph 1.13 shows several rectangles, one for each joint. Within the rectangles, it is possible to see the median values reported in the Figure as red horizontal segments. These red segments separate the first quartile (Q1) from the third quartile (Q3). Hence, the variance of the measure is proportional to the height of the rectangle, and in this way, it is possible to roughly understand the error distribution for a particular joint.

It is, therefore, possible to observe that from the comparison of the Vicon marker-based system with a single Azure Kinect without Kalman filter only four joints have a median error

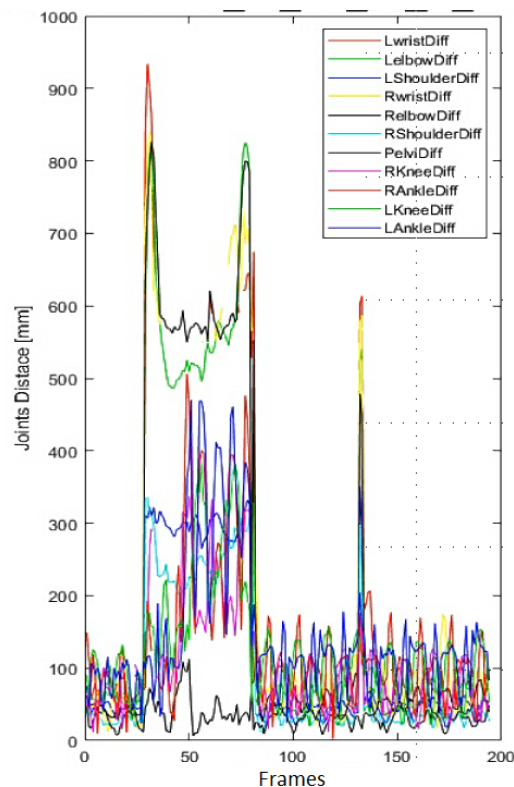


Fig. 1.12 Graph of all the Joint to Joint Euclidean Distance between the marker-based Vicon skeleton reconstruction and a single Azure Kinect Skeleton reconstruction

of less than 5 cm, five joints have a median error between 5 and 10 cm and two have error greater than 10 cm. Also noticeable is a large amount of outliers. These outliers are due to large errors in the neural network for several reasons, such as can not define well the skeleton of the person and its joints. Among the main reasons that influence this phenomenon, it has been noticed that occlusions, exit from the visual field and rapid movements play a fundamental role in the correct functioning of the neural network.

Objects such as exoskeletons, crutches, or orthopaedic walkers prevent the correct visualization of the person's profile and prevent the convolutional network from correctly extracting geometric features.

The exit from the visual field of some joints instead involves a total lack of some information. In this case, the neural network returns a result, but this result is the outcome of a prediction given by a model, unrelated to the real measure of the joint. If the lack of information is high, the model can no longer even determine the person's orientation, thus confusing right and left joints.

The velocity of motion, on the other hand, affects the correct determination of the joint along the direction of the motion. In fact, during several tests, as shown in Figure 1.11 by the

red segments, it was possible to observe that hands and feet had a relatively higher error in the velocity direction during the swing phases when the maximum speed was reached.

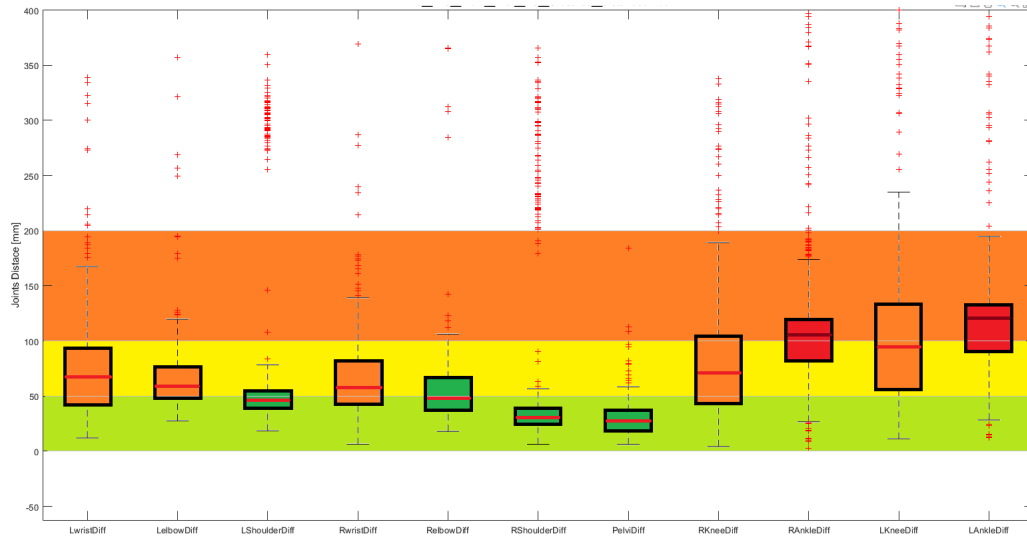


Fig. 1.13 Joint to Joint Euclidean Distance between the Vicon skeleton reconstruction and a single Azure Kinect Skeleton Reconstruction

The grap 1.14 and 1.15 report the results obtained along different tests using three Azure Kinects and introducing a sensor fusion with the Kalman filter. Immediately you can see a remarkable reduction of outliers and a decrease in the average error of the joint (except for the right which, unfortunately, was badly seen by the cameras due to an asymmetrical positioning of the cameras and occlusions by the bars of the treadmill).

Neglecting an initial settling phase of the Kalman filter in the first 20 frames, as shown in the figure 1.14, the result of the sensor fusion returned a better estimate of the skeleton than the tests performed with the single Azure Kinect. From Figure 1.15, it can be seen that the average error never exceeds 10 cm except for the right hand. In addition, the number of joints reporting an error of less than 5 cm rose to seven (reported as green boxplots in the Figure 1.15). A not insignificant improvement can also be seen in terms of glitching and unexpected network errors. The presence of outliers is greatly reduced and present only for some fast-moving or poorly visible joints.

1.8 Conclusions

In this chapter, a comparison of two motion capture systems is presented. In particular, a markerless Azure Kinect multi-camera system was compared with a standard marker-based goal system, Vicon nexus. The multi-camera markerless system featured a variable number

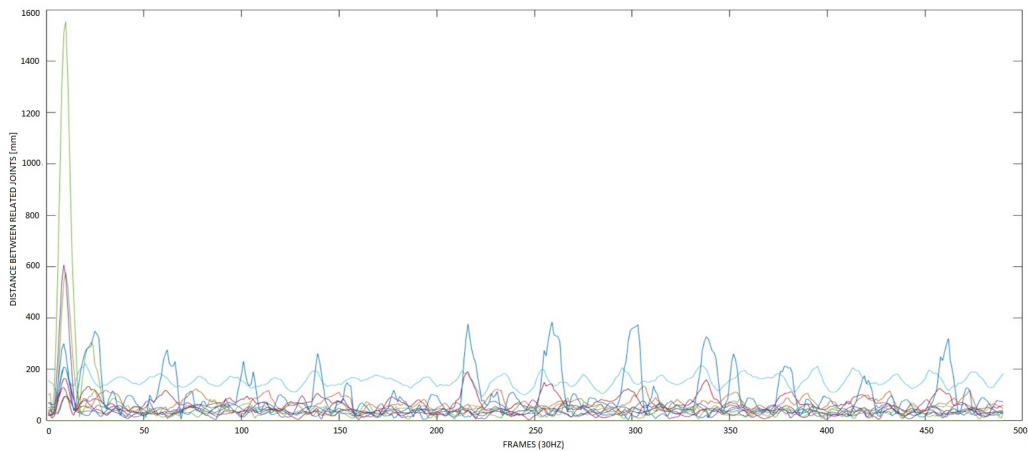


Fig. 1.14 Graph of all the Joint to Joint Euclidean Distance between the marker-based Vicon skeleton reconstruction and the markerless Azure Kinect Skeleton reconstruction with the Kalman Filter sensor fusion

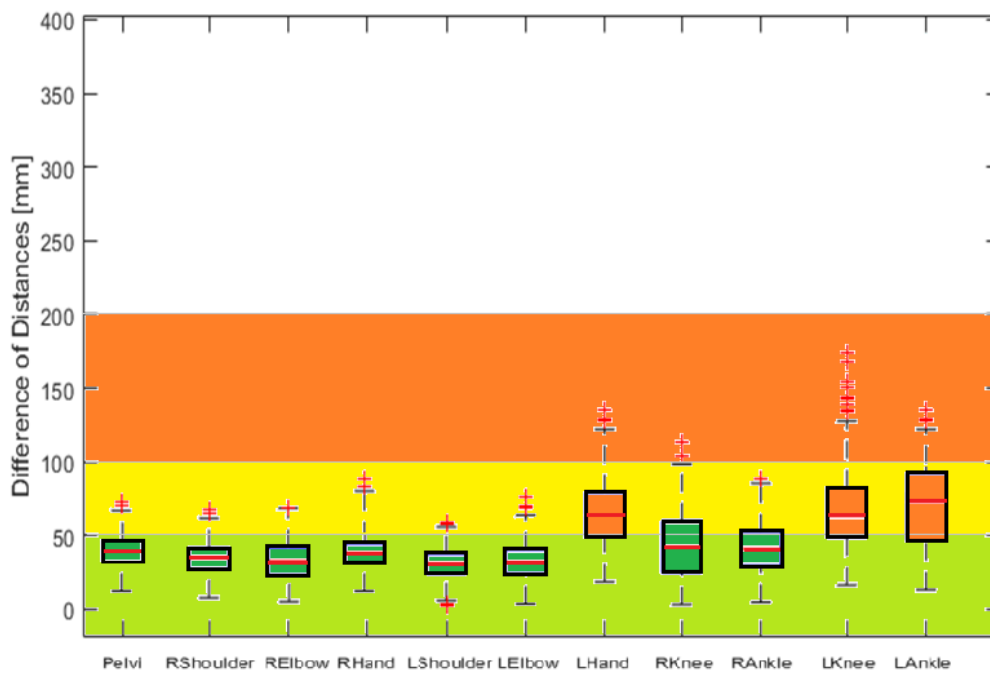


Fig. 1.15 Euclidean joint-to-joint distance of a subject walking without an exoskeleton, crutches and orthopaedic walker support between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction

of Azure Kinects cameras. A total of two subjects were monitored in different walking conditions, with and without the support of crutches, an orthopaedic walker or a robotic exoskeleton. The results obtained are promising and seem to improve considerably by introducing more cameras and sensor fusion models such as the Kalman filter. The introduction of visual obstacles, such as the robotic exoskeleton, crutches or orthopaedic walker, seems to create greater difficulties in joint identification, introducing greater measurement uncertainty and lower final accuracy. In particular, the presence of the robotic exoskeleton seems to introduce more uncertainty in the identification of the foot and ankle, as shown in Figure 1.16. The presence of the orthopaedic walker introduced greater uncertainty in the correspondence of hand joints and a general increase in glitches for some cameras, as shown in Figure 1.17. Finally, the presence of crutches introduced increased uncertainty for the joints of the hands, wrists and elbows, as shown in Figure 1.18.

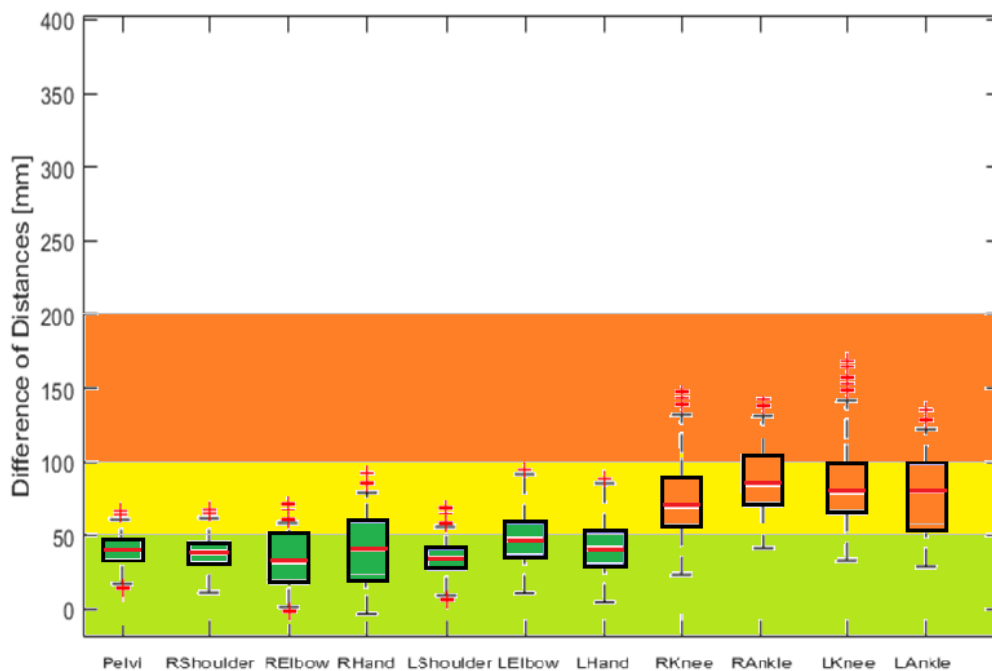


Fig. 1.16 Euclidean joint-to-joint distance of a subject walking with a lower limb exoskeleton between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction

Furthermore, as shown in Figure

Finally, it is important to remember that the acquisition system using azure Kinect and the convolutional neural network is further influenced by the colour of the clothing, the positioning and the number of cameras [125]. For these reasons, further tests and specific

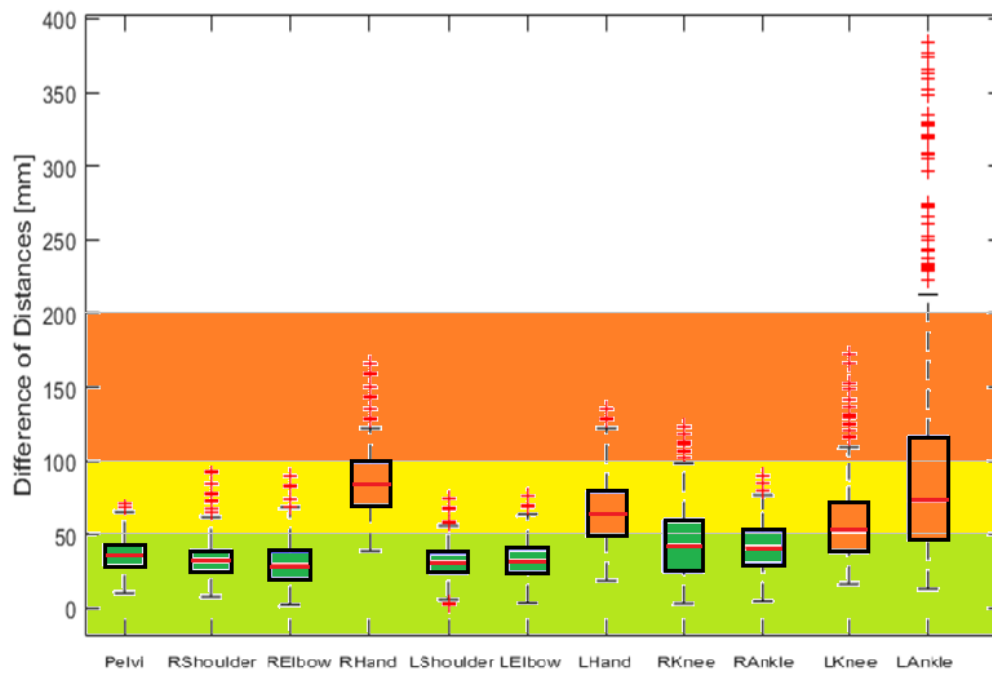


Fig. 1.17 Euclidean joint-to-joint distance of a subject walking with an orthopaedic walker between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction

developments of this project are in progress. These include the introduction of additional cameras, the analysis of performance by wearing different coloured clothing, the repositioning of the cameras and an increase in the number of subjects. Moreover, it will be considered a more suitable model for the Kalman filter which considers the joints and angular constraints.

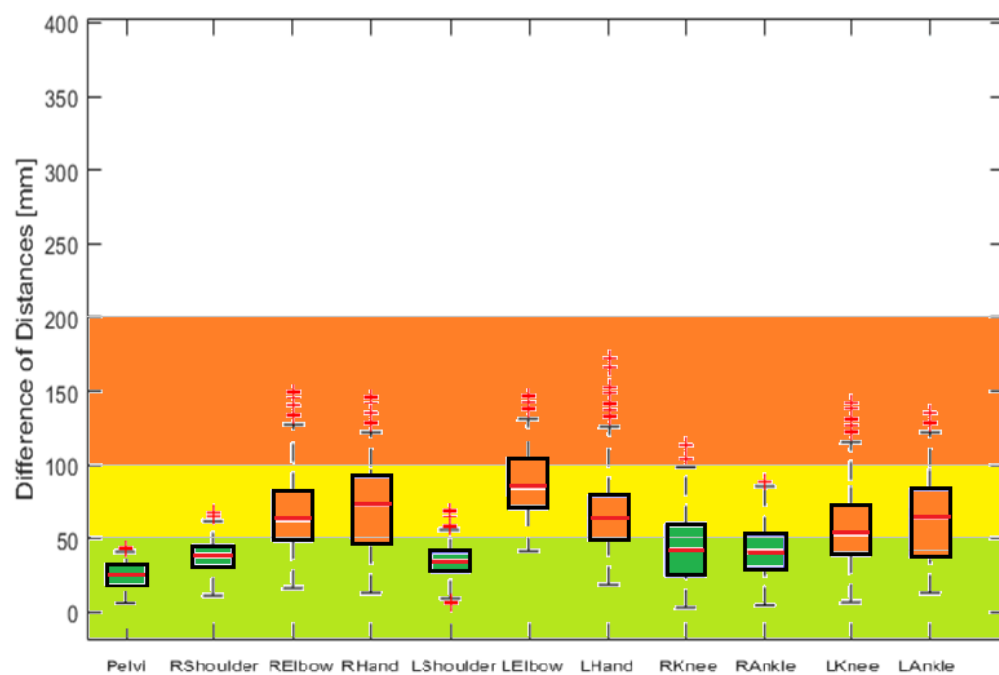


Fig. 1.18 Euclidean joint-to-joint distance of a subject walking with the support of crutches between the Vicon skeleton reconstruction and the three-azure Kinect skeleton reconstruction

Chapter 2

Estimation of the mass and inertia of the anatomical segments

2.1 Traditional approaches for the measurement of anthropometric parameters

The systems for measuring human biomechanical parameters have played a very important role in medical studies, the development of custom-made prostheses, and motion analysis. Among the most common methods, the anthropometric tables report measurements and references obtained from large samples of the population by the standards ISO 7250 and ISO 15535 [83]. Although they have been obtained by exploiting high standard measurement equipment and cutting edge technologies, the tables report statistical values valid to give general indications on a portion of the population and can not be directly applied to the individual. In this case, it is then necessary to rescale the values based on the subject main characteristics, such as height and weight. In addition, they do not often include all the parameters, leading to coarse approximations and large uncertainty[15].

Other approaches have pursued their practicality and simplicity, such as the measurement with the tailor's tape as shown in Figure2.1, reporting the anatomical circumferences and lengths of the subject and approximating the anatomical segments to simple geometric solids [110]. Also, in this case, the estimation of biomechanical parameters, such as volumes, masses and inertias, results rather coarse and with little repeatability, due to the manual measurement procedures that are strongly dependent on the operator's skills.

Another approach for estimating the volume and the mass is to dip each anatomical part in water [107]. However, despite the simplicity and high accuracy of the method, it results quite slow and time-consuming. Furthermore, on one hand, it results relatively easy while



Fig. 2.1 Measurement with the tailor's tape for the comparison of the methods

considering upper and lower limbs, on the other hand, it is difficult to implement for proximal or central segments, such as the trunk, shoulders and pelvis. Moreover, by giving a direct measurement of volume without first establishing any geometric feature, it can be extended for the calculation of the masses by knowing the densities but is not usable for estimating rotational inertias.

2.2 Technologies for the 3D virtual reconstruction

3D acquisition systems are now used in many fields thanks to the reliability and accuracy achieved by the sensors of which they are made. These devices perform measurements able

to reconstruct three-dimensional environments and render them as a virtual representation. From this representation, it is possible to extract a lot of information typically of geometric nature that allows the study of the characteristics of some target elements. Among these features, the study of the volume seems to be particularly investigated on large and small scales, from the estimation of morphological-environmental variations for land management to the estimation of small volumes for production control in industrial automation.

Each sensor has specific characteristics tailored with respect to the application for which it is required by choosing the definition parameters according to the needs, such as image resolution or depth accuracy over long distances. Lidar technology for example has a long-range but low image resolution. This sensor has been used for volume estimates for large portions of land or environmental applications, such as estimating tree height [80] or avalanche hazards in mountainous areas[20]. Moreover, Lidar technology is still very expensive and it is quite difficult to implement for simple low-cost applications.

Other devices for 3D human body reconstruction are Bodyscanners. Body scanners, such as Vitronic¹, and Artec², appear as cabins or portable devices that are sold as commercial products. They are designed to scan a person's body using simple reverse engineering techniques and appear to be particularly accurate and suitable for our application. However, they are particularly expensive and can not be considered for the development of a low-cost system. In addition, while the Vitronic cabin has an automated protocol for scanning, the less expensive Artec 3D Scanner requires specialized personnel to operate it.

Some industrial processes have automated production through vision and artificial intelligence technologies. For example, the work reported by [67] and [32] illustrates the development of a measurement system with neural networks for estimating food volume and dietary intake. However, the use of neural networks for this kind of application implies the training of the algorithm through a training dataset generated on purpose. This dataset is usually created from a pre-existing process. If this is not feasible, the creation from scratch of a training dataset is particularly time-consuming, eager for resources and cumbersome.

In the work reported by Pablo Davila et al. [29], a low-cost system for measuring anthropometric parameters through mesh reconstruction is proposed. The body is scanned with Kinect 360 and reconstructed in 3D using Skanect (Scanning Software By Occipital) and Blender (Blender Foundation). However, our work is not aimed at mesh reconstruction, and by aiming only at a simple measurement of masses and volume inertia we opted for minimizing the dependencies on external software. In any case, this work gave us a good starting point for choosing an affordable device that would suit our needs.

¹<https://www.vitronic.com/en-us/3d-bodyscan> [Accessed: 01/2022]

²<https://www.artec3d.com/> [Accessed 01/2022]

Another work that further supports the choice of Kinect for calculating volumes from scanning small objects is reported [5]. In particular, Andújar et al. describe the use of the Kinect camera in an agricultural setting for a weed volume estimation application. Even if the estimation is performed coarsely and aims only to give feedback for the automation of a maintenance process, the potential benefits of using such a device emerge.

For this reason, we have chosen to use Azure Kinect, the advanced version of Azure 360 and we then developed our processing algorithms based on the 3D Montecarlo integration approach as reported by Nicola Covre et al. in [24]. This algorithm estimates the volume of the anatomical segments starting from their 3D surface point clouds. For the body segmentation, instead, we exploited Open pose, an open-source convolutional neural network algorithm.

2.3 The Bullet project

In this chapter, we propose a low cost, automated and fast acquisition system compared with the state of the art proposed in 2.2. The system estimates the measurement of volumes, and geometric features using RGB and depth images. This allows the calculation of masses and inertias of the anthropometric segments providing the average density values reported on the anthropometric tables. The scanning system integration and validation are proposed as a part of the Bullet project. The Bullet project is a research activity funded by the European Eurobench project through a cascade funding scheme called FSTP (Financial Support to Third Parties).

The final goal of the Bullet project is the development of a biomechanical system to estimate the reaction forces acting on the joints of subjects walking with robotic exoskeletons and crutches. The development of such a measurement system allows for to prevent of health problems associated with the incorrect use of such exoskeletons and aims to prevent the overloading of the articulation which may cause joint pain and chronic diseases.

The autonomous scanning system consists of a sensorised hospital bed that allows to estimate of the volume, mass and inertia of the anatomical segments of the lying subject through the acquisition of an Azure Kinect time-of-flight camera and a computational unit that processes the information. This operation is done in several steps. First, the algorithm segments the 3D point cloud of the subject lying on the bed using Open Pose. Then, provided the anatomical densities of the anthropometric tables, it calculates the volumes, masses and inertias of the anatomical segments through numerical integration algorithms.



Fig. 2.2 Test of the Bullet project during exoskeleton-assisted walking with sensorised crutches and pressure matrices

2.4 Materials and Set up

The 3d volumetric scanner system is composed of a hospital bed equipped with a wooden frame to support a low deformation mattress, an aluminium frame for the installation of electronic devices, and sensitized with an azure Kinect camera to allow the acquisition of RGB and depth images, as reported in Figure2.3.

The time-of-flight camera should be positioned in the centre of the horizontal rod of the metal frame. It should be installed at a height of about 170 cm from the low deformation mattress as a compromise between the best image resolution and the minimum distance that

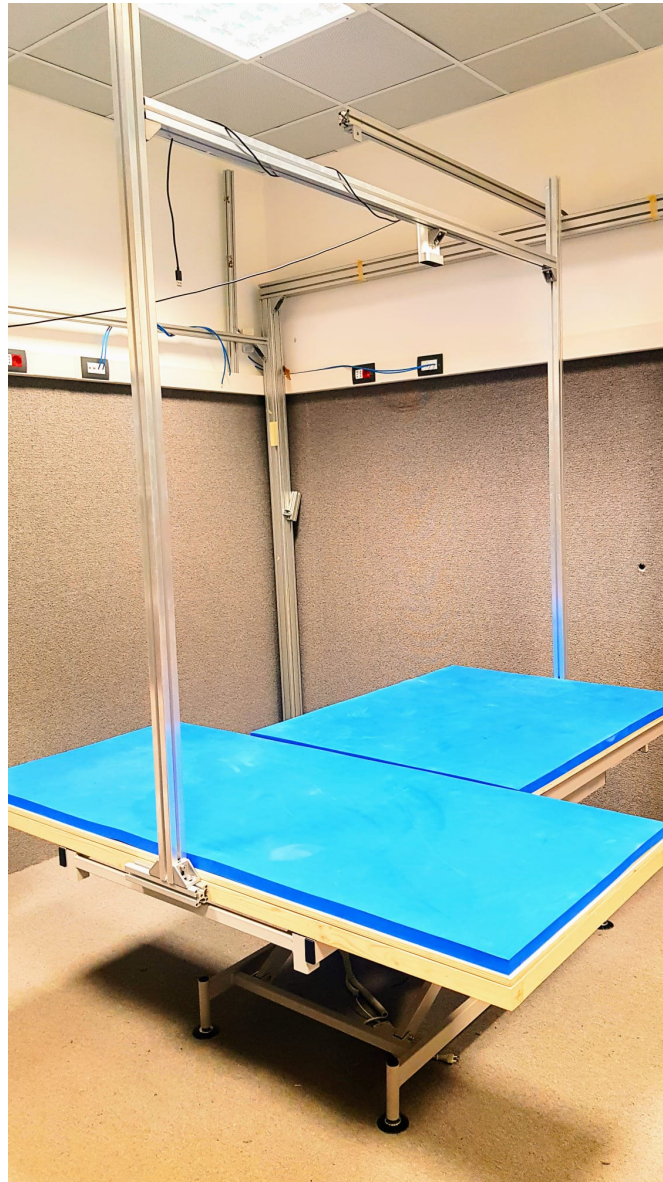


Fig. 2.3 Volumetric Scanner system Set up composed of the hospital bed, the horizontal wood frame with low deformation mattress, the vertical aluminium frame and the ToF Camera Azure Kinect

includes the bed surface and the subject in their entirety. The camera must also be oriented downwards with vision indicatively orthogonal to the plane where the subject lies. Even in the case where the camera is mounted slightly tilted (a few degrees), the algorithm will independently identify the plane of the bed and reposition the reference system.

The bed is designed with a pantographic mechanism operated by an electric motor to allow the lowering and raising of the platform thus helping the subject to get on and off. As

shown in Figure 2.4, a low-deflection mattress is mounted on a wooden frame arranged in a "T" shape. This allows the subject to spread his arms, facilitating the segmentation operation near the armpits, elbows, and wrists.

The low deformation mattress does not allow the subject to sink into it thus avoiding affecting the volume measurement.

The subject should be undressed or wearing tight-fitting clothing that minimizes bulkiness so as not to compromise volume measurement. The subject should wear an adherent swimming cap to reduce the volumetric artefact introduced by the presence of hair. Bald subjects may not use the swimming cap. The subject should not wear any black clothing or any colour that absorbs light. This requirement is dictated by the non-reflection of infrared emissions by which the time-of-flight camera performs the depth measurement. For this reason, light-coloured reflective clothing such as white, blue, or green is recommended.

The measurement procedure is divided as follows:

1. Explanation of the procedure to the subject, signing of consent forms and acknowledgement of data processing by what has been approved by the ethics committee. The subject is explained that he/she can withdraw at any time from the acquisition tests for any reason he/she deems appropriate and without having to explain the reason. The following steps are only possible if the first one is successful.
2. The subject is prepared according to what is indicated in the measurement protocol by letting him wear only clothes that do not compromise the measurement of volumes: undressed or with tight clothes of light colour, a swimming cap to contain the volume of hair.
3. Positioning of the subject on the bed with arms and legs spread as in the position shown in the Figure 2.4
4. RGB and depth image capture of the subject while exhaling air from lungs using Azure Kinect and reconstruction of the point cloud in 3D
5. Processing of the acquired data to separate the subject's body from the measurement environment and virtual segmentation of anatomical parts using Open Pose and 3D cutting sequences
6. Calculation of the length, volume, mass and inertia of each anatomical segment using an extension of the Monte Carlo algorithm for the estimated 3D volumes as reported and validated by Nicola Covre et al. in [24]



Fig. 2.4 Subject on the Hospital bed ready for the acquisition

2.5 Data Collection and Processing

Data Acquisition Once the subject is lying in position, she/he is asked to exhale all the air from her/his lungs so that such a volume is not considered part of the anatomical segment of the chest/abdomen. Once the lungs are emptied, the Azure Kinect acquisition code is started and instantly acquires the RGB and depth image.

Subject separation from the surrounding environment From the depth image, it is possible to reconstruct the point cloud that will allow the 3D segmentation of the anatomical segments. First, as reported in Figure 2.5 the algorithm loads the 3D point cloud of the reconstructed environment and uses a Ransac algorithm to search for planes. As shown in

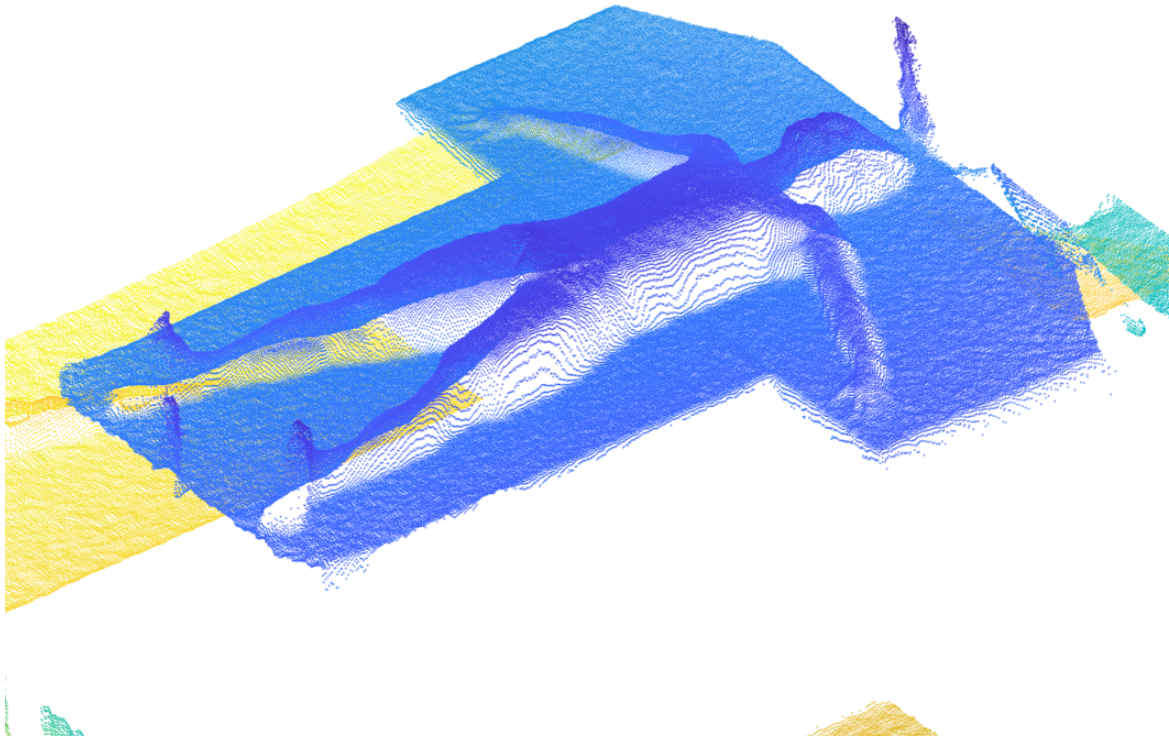


Fig. 2.5 Cloud of dots of the subject lying on the hospital bed

Figure 2.6 it finds the bed plane and selects three points approximately chosen points next to the upper left, upper right, and lower left corners.

These points all lie in the 3D plane and allow the reference system to be relocated to the bed plane. This operation permits isolating the subject from the rest of the environment discarding all the points located below the bed level, as illustrated in Figure 2.7.

Open Pose Joint Extraction and projection in the 3D space To proceed with the segmentation, it is necessary to first extract the coordinates of the joints using the Open Pose algorithm [17]. Open Pose is an open-source algorithm based on a convolutional neural network. Provided an RGB image, it returns the pose of the subjects it finds and the pixel coordinates of each one of their joints, as illustrated in Figure 2.8a.

By projecting the pixel coordinates into the depth image, it is possible to extract the 3D coordinates of each of the subject's joints as shown in Figure 2.9.

Separation of the anatomical segments Now that the point cloud of the subject's 3D surface has been separated from its surroundings and the 3D coordinates of each joint are known, it is possible to define cut sections to separate the anatomical segments. Each cut

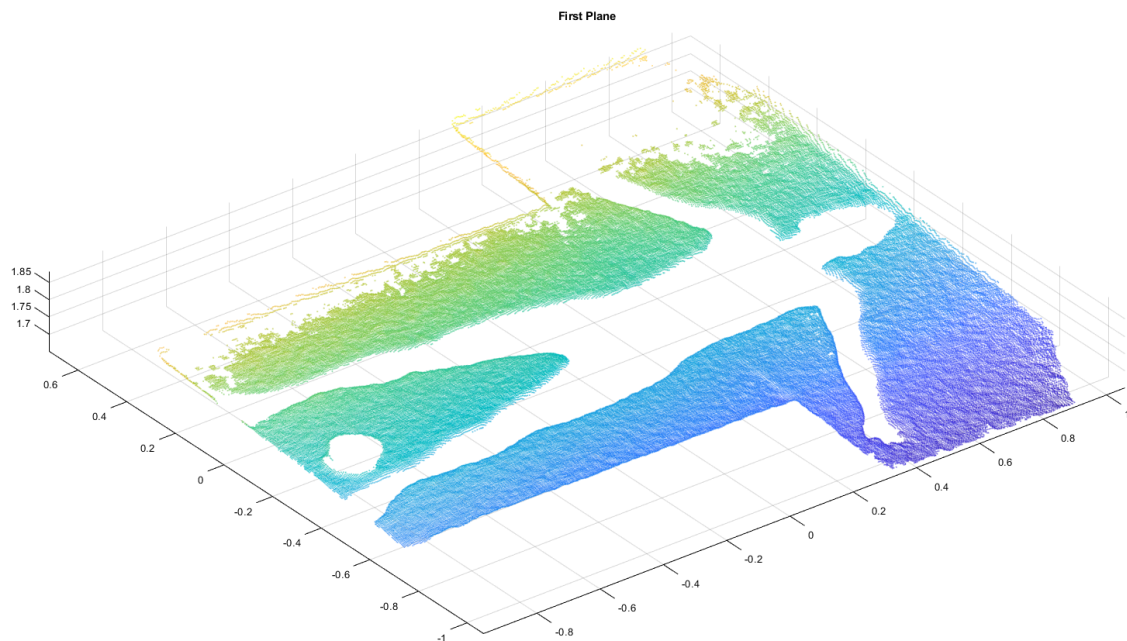


Fig. 2.6 Plane of the Hospital bed found by the Ransac algorithm

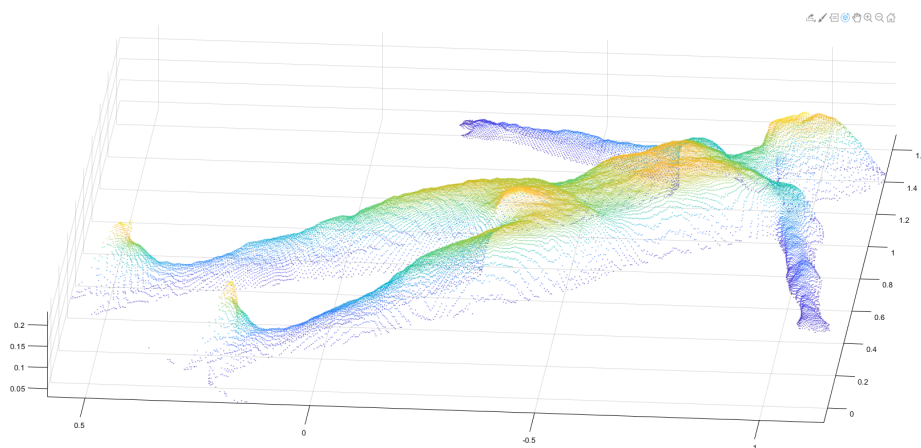


Fig. 2.7 Subject isolated from the surrounding environment

section is orthogonal to the bed plane and has a specific inclination with respect to the coronal plane of the subject. The system of reference is defined with the y-axis aligned with the nose-pelvis segment and the x-axis computed as the vector product between y and the normal outgoing from the bed plane. In this way, the reference system has the x-axis parallel to the intersection of the transverse plane with the coronal plane and it has the y-axis parallel to the intersection between the sagittal plane and the coronal plane.

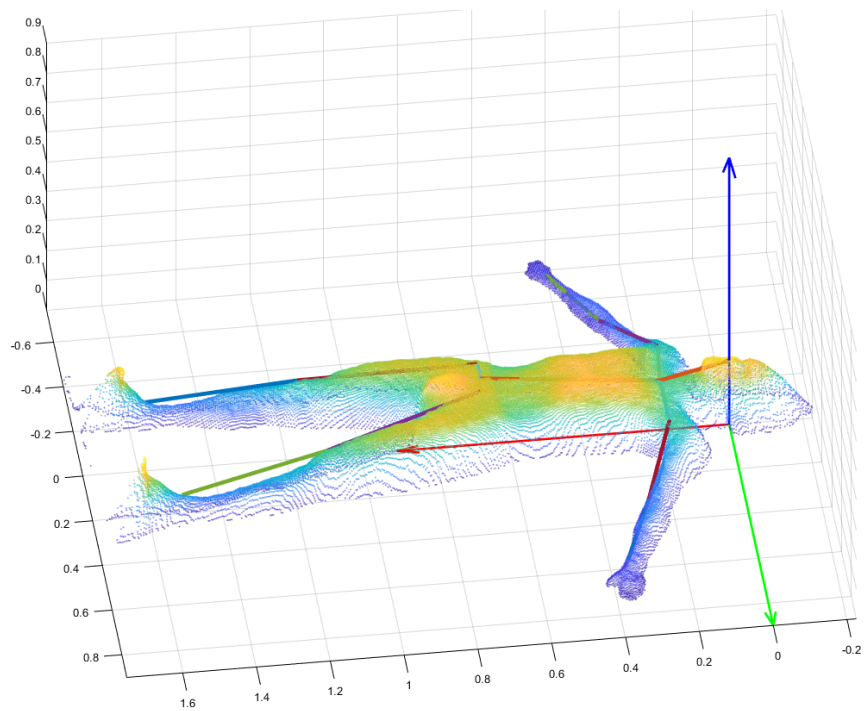
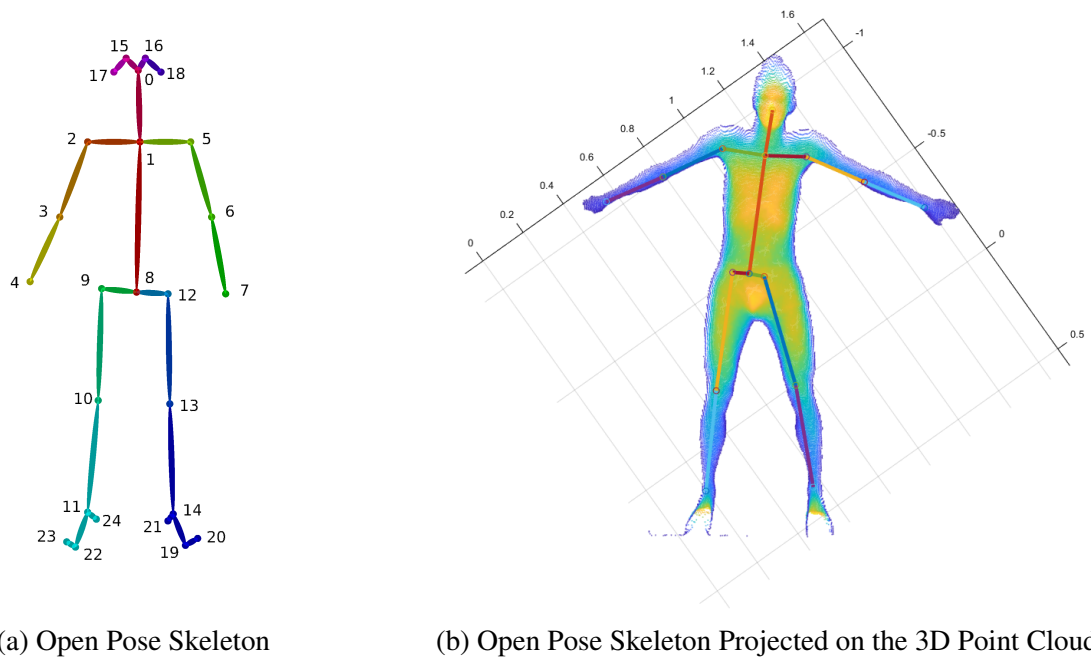
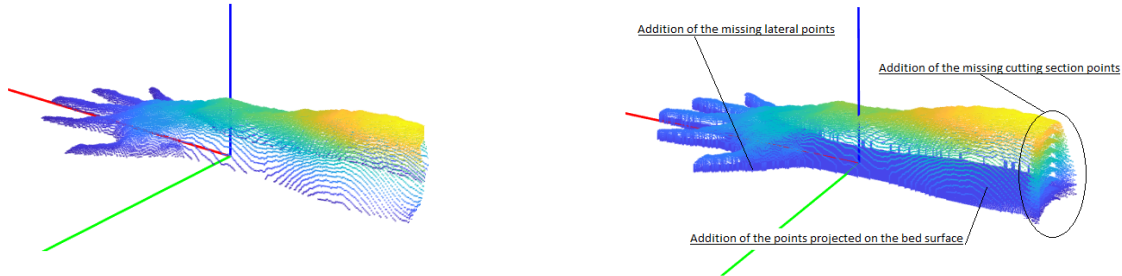


Fig. 2.9 Open Pose Skeleton merged to the 3D Point Cloud

The origin of the reference system is redefined each time a cutting sequence is carried out. The cutting sequence then translates the origin of the reference system to the proximal end of the anatomical segment, rotates it up to the cutting angle defined in [34] with respect to the

axis exiting the frontal plane and selects the points of interest by solving a simple inequality. Due to the irregularity of the virtual surface of the bed, as reported in Figure



(a) Original acquisition of the aight arm before applying the reconstruction algorithms

(b) Right Arm after applying the reconstruction algorithms

2.6 Methods for the volume estimation from a surface point cloud of 3D objects

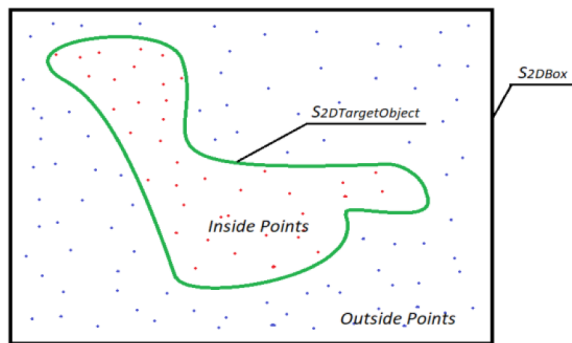
The calculation of the volume of a point cloud is a widely studied topic in literature because of the many applications it has. For example, Chang et al.[19] calculated the volume enclosed in a point cloud using the least-squares method and the slices approach. They achieved good accuracy, but only with regular solids and a homogeneous distribution of points. Similarly, Zhi et al.[131] applied the slices method and showed that the result is highly dependent on the quality of the point cloud, highlighting the difficulties of this method in calculating the volume of irregular shapes.

The Convex Hull algorithm [55] presents a different approach for calculating the volume from a discrete 3D representation of its surface. However, from the works of Xu et al.[120] and Bi et al.[11], it can be seen that the measurement is often overestimated. An improved version of the Convex hull algorithm was proposed by Lin et al.[66] They showed that this new version can also handle concave polygons. However, they only reported measurements of tree crowns as they could not highlight and validate the performance of the algorithm when working with small, detailed objects. Some other works instead opted for using Mesh reconstructing algorithms to first calculate the mesh and, from that, calculate the geometric properties of the object.

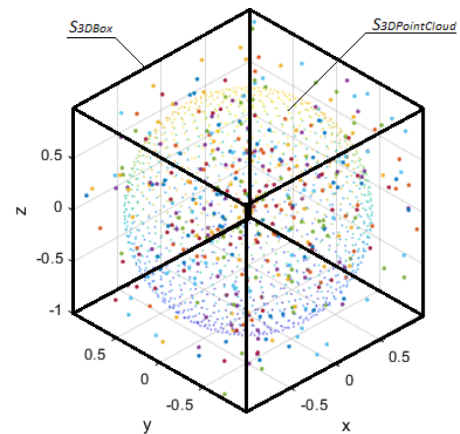
For example, the work of Lee et al.[62] proposes the calculation of the mesh from the acquired point cloud for computing the waste volume and monitoring purposes. This

approach is particularly effective for point clouds with homogeneous distributions, but it is a bit difficult in the case where the meshes have scattered or missing distributions. In addition, this approach relies completely on mesh calculation algorithms, such as Mesh Lab [93], and cannot pursue volumetric calculation in the case of discontinuous or open meshes.

In this section, a method for calculating the volume enclosed in the point clouds of the external surface of a 3D object is presented. The proposed method extends a method known as the Monte Carlo integration approach and does not use interpolation or mesh reconstruction algorithms. This algorithm seems to be able to calculate the volume of objects even in the presence of non-uniform point clouds, with sparse or even missing distributions. Furthermore, this approach allows the calculation of the volume of both regular and irregularly shaped objects.



(a) Application of the Monte Carlo integration approach to a 2D object (in green)



(b) Application of the Monte Carlo integration approach to the surface point cloud of a 3D object

The Montecarlo Integration approach The Montecarlo integration approach is in general a statistical algorithm that calculates the volume of the target object with a proportion of point distributions compared to a known volume. As shown in Figure 2.11a, this algorithm provides for the generation of a regular-shaped box around the target object. This box of known size and volume is used as a reference for the calculation of the unknown volume it contains. Then a large number of points are randomly generated with uniform distribution inside this box. Some of these points will also fall inside the target object while others outside. As the number of generated dots tends to infinity, the ratio of dots falling inside the target object to the total number of generated dots inside the box will tend to be the ratio of the unknown volume to the volume of the box. Therefore, by solving the following equation, the desired volume can be estimated:

$$\lim_{n \rightarrow \infty} \frac{n_{insidePoints}}{n_{TotalPoints}} = \frac{V_{targetObject}}{V_{BOX}} \quad (2.1)$$

with

$$n_{TotalPoints} = n_{insidePoints} + n_{outsidePoints} \quad (2.2)$$

hence

$$V_{targetObject} = \lim_{n \rightarrow \infty} \frac{n_{insidePoints}}{n_{TotalPoints}} V_{BOX} \quad (2.3)$$

2.7 The Monte Carlo-based volume estimation by exploding local cubes faces of a 3D surface point cloud

The Montecarlo algorithm relies on an affiliation criterion that determines which points belong to the inside and which to the outside of the target object. This criterion is easily defined when there is a continuous representation of the surface of the target object. However, it is particularly complicated when this surface is represented by a cloud of points ($S_{3DPointCloud}$), as shown in the Figure2.11b. In this section, we describe the affiliation criterion we have developed based on the explosion of cube's faces. The Monte Carlo-based volume estimation by exploding local cubes face of the surface point cloud of a 3D object is divided into the following steps as reported in appendix A:

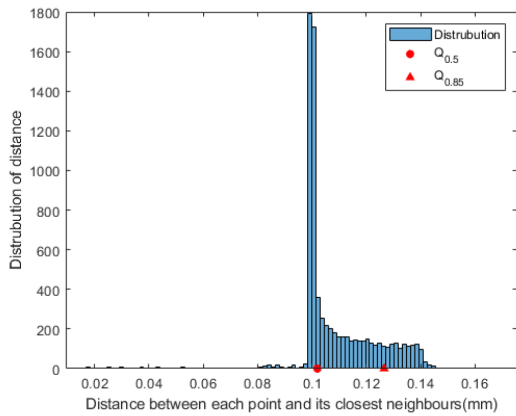
1. Re-orientation of the target object with PCA
2. Definition of the Box around the target object and computation of the reference volume
3. Cloud Analysis
4. Generation and affiliation of the points inside the Box
5. Computation of the volume of the Target object

Re-orientation of the target object with PCA At the origin, the Target object can be randomly oriented in space. For the same number of generated points, the Monte Carlo criterion is particularly efficient if the box of known volume is slightly larger than the Target object. For this purpose, it is convenient to reorient the object using principal component analysis along with its main directions.

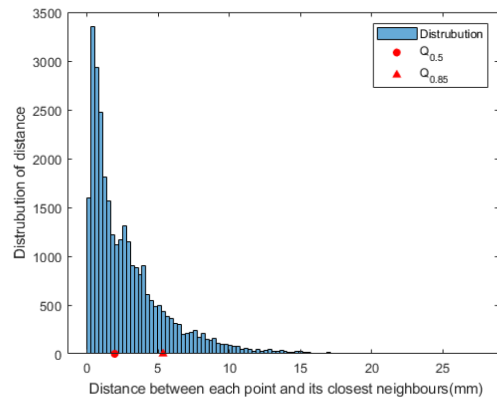
Definition of the Box around the target object and computation of the reference volume

The six minimum and maximum coordinate points in the three main directions x, y, and z are used to define the box around the target object. From these six points, the 8 vertices of the parallelepiped box enclosing the target object are created. The volume of the parallelepiped is then calculated and used as the reference volume for the final calculation.

Cloud Analysis The point cloud analysis is used to define the initial size of the cubes during the generated point cloud affiliation phase and is fundamental for the adaptation of the algorithm to the point cloud of the target object. The analysis of the point cloud begins by defining the average distance between each point of the cloud and its nearest neighbours. Depending on whether the distribution of the point cloud is more or less homogeneous, this average distance will vary point by point. In the case of a homogeneous point cloud, the histogram of distances will present a consistent average distance with a single high peak and a narrow distribution as shown in Figure2.12a. In the case of uneven distribution, with dense areas of the point cloud interspersed with scattered areas, the histogram of the distribution will be low and wide as shown in the Figure2.12b.



(a) Distribution of the Sphere surface point cloud - an example of a homogeneous point cloud



(b) Distribution of pokemon Mew - an example of a non-homogeneous point cloud

By carrying out several tests with different point clouds, it turned out empirically to initialize the edge of the cubes as follows:

$$l_{cube} = Q_{0.5} 3.5 \left(\frac{Q_{0.85}}{Q_{0.5}} - 1 \right); \tag{2.4}$$

where $Q_{0.5}$ is the quantile at 50 % of the distribution and $Q_{0.85}$ is the quantile at 85 % of the distribution.

Generation and affiliation of the points inside the Box In contrast to the original Monte Carlo algorithm, the point generation procedure is sequential, to ensure a homogeneous and total coverage of the space inside the box. The algorithm generates a cube around each generated point ($P_{generated}$) by initializing its dimensions as set out in the equation 2.4. An extrusion routine then begins and is performed for each one of the six faces of the cube. At each iteration, a face is selected and extruded in the direction normal to it as shown in Figure 2.13. If the extrusion encounters some points of the cloud (in Figure 2.13 reported in green) they are counted. An empirically selected threshold value between five and ten is chosen as reference. If the total number of points intercepted exceeds the threshold value the extrusion is repeated by taking a cube with smaller dimensions as follows:

$$l_{cube_{new}} = l_{cube_{old}} * 0.09 \quad (2.5)$$

A smaller cube will intercept fewer points of the cloud. The iteration is repeated until the total amount of intercepted points is lower than the threshold. Once the total amount of intercepted points is less than the threshold value, the algorithm proceeds with the clustering as shown in the Figure 2.13. Clustering is the operation that groups the intercepted points along the extrusion direction. The criterion for determining which points belong to the same cluster or another is established by a matter of distances. All points close to each other in the extrusion direction belong to the same cluster. Knowing that each extrusion takes place parallel to the axes of the reference system (x,y,z), this translates the issue into a simple difference between the coordinates relative to the extrusion axis. If this difference is small, the points are close and they belong to the same cluster ($P_{cluster}$), and vice versa if it is large.

Having defined the total number of clusters into which the points have been divided, the judgment related to the extrusion of the single face takes place and is saved. If the total number of clusters encountered is odd, the judgment is positive and the point is defined as internal to the target object. If the total number of clusters is even, the judgment is negative and the point is defined as external, as shown in Figure 2.13.

In the end, once each of the six faces has been extruded, the six judgments generate a final judgment equal to their mode. By repeating this routine for each point generated, the total amount of external points, the total amount of internal points and the total amount of generated points will be eventually defined and plotted, as shown in Figure 2.14.

Computation of the volume of the Target object The algorithm calculates the final volume enclosed in the target object as the reference volume of the box multiplied by the ratio of the total amount of internal points over the total amount of points generated, as expressed by the equation 2.3.

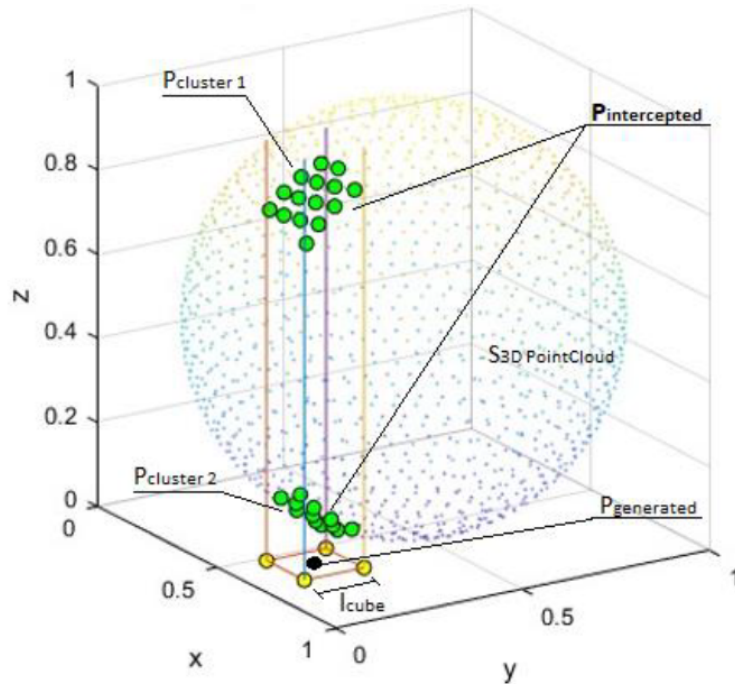


Fig. 2.13 Extrusion of one of the faces of the cube and interception of two clusters of points.

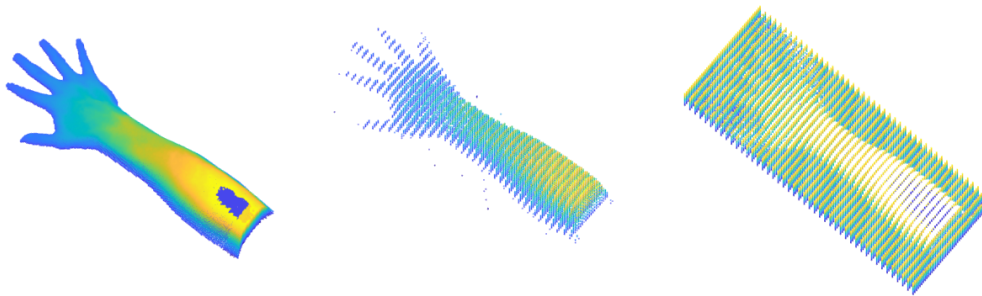


Fig. 2.14 Comparison between the surface point cloud of the 3D object(left), inner point cloud (centre) outer point cloud (right)

2.8 Results of the volume estimation algorithm

The Montecarlo algorithm for calculating the volume enclosed in surface point clouds of 3D objects has been validated on several objects with different shapes and point distributions. A total of nine surface point clouds were considered for validation, five generated from virtual meshes and four reconstructed from real objects. As it is reported in the first column in the table 2.1 on the left we have the representation of the virtual objects while in the table 2.2 are reported the real objects. The second column of both tables shows the surface

point cloud ($S_{3DTargetObject}$) obtained from the mesh, for the virtual objects, or from the 3D reconstruction of a Konica Minolta Vivid VI-9i 3D, for the real objects. The third and fourth columns in both tables report, respectively, the reconstruction performed by MeshLab and that performed by the Monte Carlo algorithm with the affiliation criterion by exploding local cube faces. The volumes have been computed in the *MATLAB_{R2020b}* environment and by using a MacBook Pro equipped with 2 GHz Intel Core i5 quad-core and with 16GB of RAM.

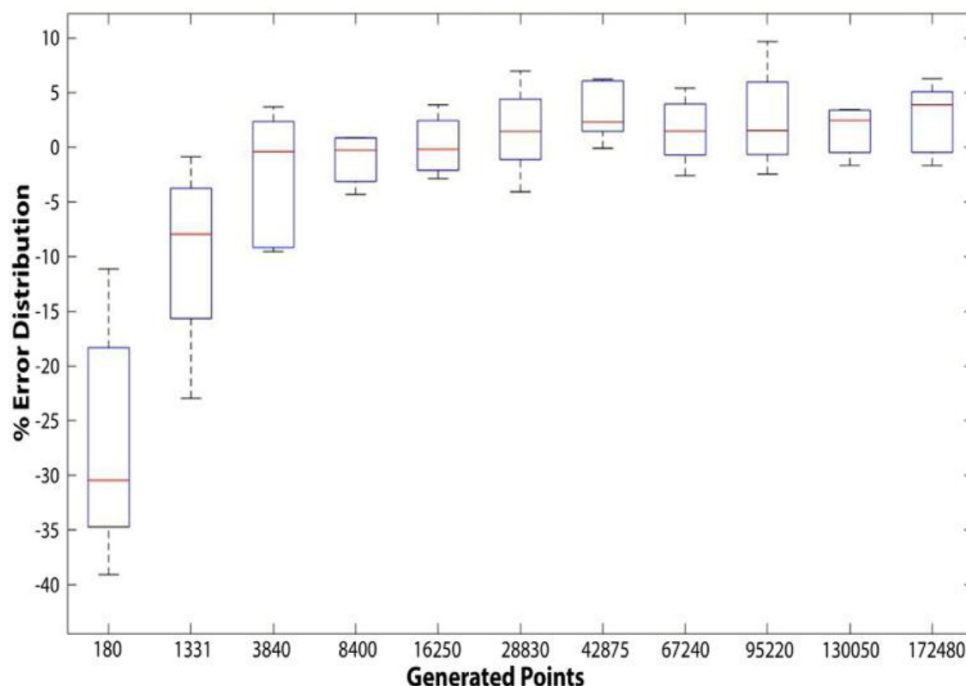


Fig. 2.15 Box plot of the Volume computation depending on the amount of generated points

As can be seen from the Figure 2.15, the volume calculation converges to the reference value as the number of generated points increases. The greater the number of points generated for the volume calculation, the more accurate the measurement and the more detailed the point cloud of internal points.

However, as the number of generated points increases, the volume computation time also increases, as shown in the Figure 2.16. Thus, a good criterion for establishing a good tradeoff between computation time and accuracy achieved is based on the convergence criterion. When the difference in the volume of the same object calculated with different amounts of points falls below a certain threshold we can stop and take the median of the last calculated volumes.

As can be seen by comparing the Montecarlo calculation method with the reference volume and with other standard volume calculation algorithms such as Meshlab, the local cube explosion criterion seems to report good results with both simple and complex shape

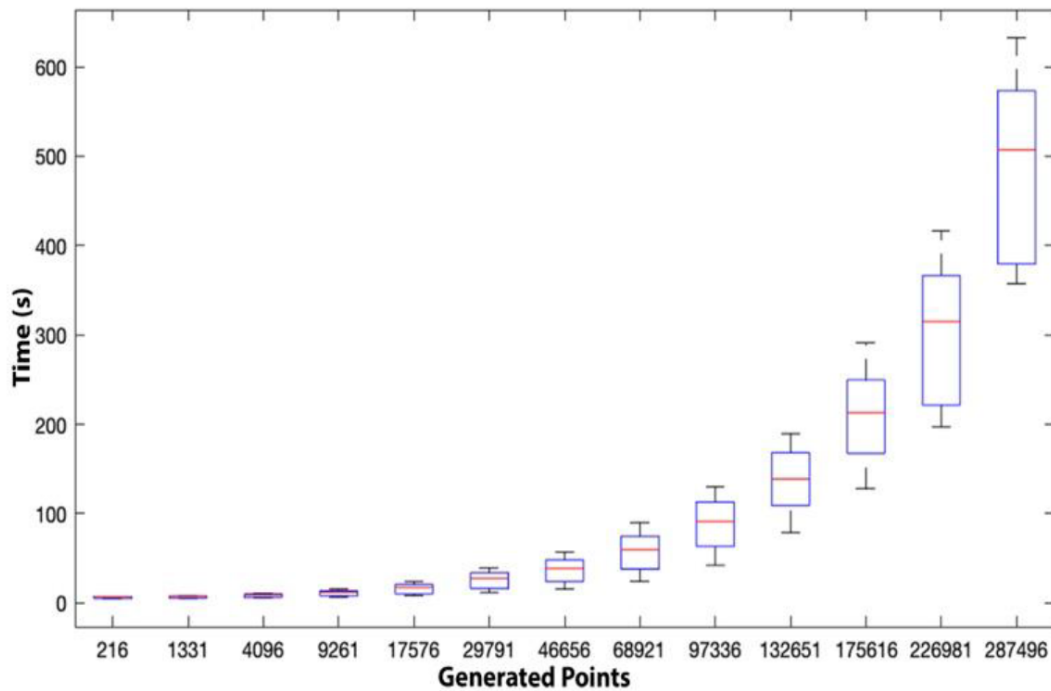


Fig. 2.16 Box plot of the computational time depending on the amount of generated points

figures. For Example, by taking into account the point cloud of the Pokemon Mew (reported in the fifth row of the table 2.1) and the sphere (reported in the second row of the table 2.1), it is possible to see that the volume measure is stable with both homogeneous and non-homogeneous meshes. We can easily observe that in the case of the Pokemon Mew there is an alternation of dense distributions in the eyes and sparse distributions on the belly. While, in contrast, the sphere is an example of homogeneous distribution. The two different types of distribution are further evidenced by the respective histograms, shown respectively in the Figures 2.12b and 2.12a .

2.9 Computation of Mass and Inertia

As reported in this chapter, the Monte Carlo algorithm performs a volume calculation by also returning the position of all points generated with homogeneous distribution within the surface point cloud, as illustrated in the last column of both tables 2.1 and 2.2.

Mass computation Assuming a uniform density and knowing the average density reported on the anthropometric tables [34], the mass of the anatomical segment can be calculated by multiplying the estimated volume by the average density.

Computation of the moment of inertia The moment of inertia of an object with non-zero mass measures its resistance to rotational acceleration with respect to a fixed axis. The calculation of the moment of inertia that is here proposed requires the knowledge of the cloud of internal points returned by the volume calculation with Monte Carlo through the explosion of local cubes, as reported in the central image of Figure 2.14 or the last column of the tables 2.1 and

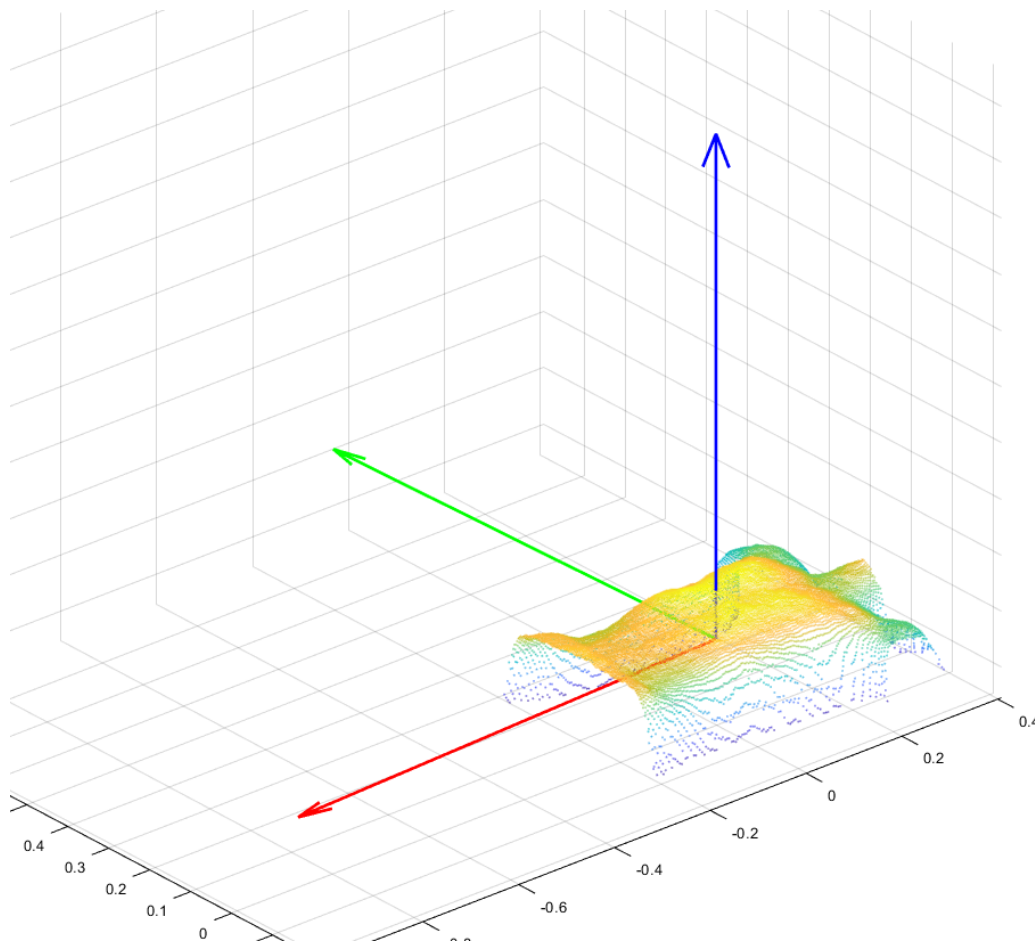


Fig. 2.17 Cloud of points on a subject's chest surface reoriented with principal component analysis and expressed with respect to a reference system with axes parallel to the three orthogonal directions of maximum variance and with origin in the centre of gravity.

Then, the moment of inertia with respect to each axis is computed as follows:

$$J_x = \sum m_i d_i^2 \quad (2.6)$$

where J_x is the moment of inertia with respect to the rotation around the x-axis, m_i is the mass of the i^{th} point of the inner points cloud and d_i is the orthogonal distance of the i^{th}

point from the x-axis. Repeating the same operation for the J_y and the J_z moments of inertia respectively referring to the y and z axes, it is possible to build the Matrix of Moments of Inertia with respect the main body axes as follow:

$$J = \begin{pmatrix} J_x & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & J_x & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & J_x \end{pmatrix} \quad (2.7)$$

2.10 Conclusions

This chapter presents the development and realization of an automatic system for the estimation of volumes, masses, and inertias of the anatomical segments. The system is composed of an acquisition and a post-processing component. The acquisition component consists of a hospital bed properly equipped and sensorised to allow the installation and control of an Azure Kinect depth camera. the processing component consists of two algorithms developed in Matlab. The first deals with the segmentation of the subject's body while the second calculates the volume, masses, and inertia. Once the acquisition of the RGB-D images of the subject is carried out, it is explained how to isolate its body from the surrounding environment. Then, the segmentation of the point cloud is illustrated starting from the pixel coordinates of the anatomical joints recognized by Open Pose. The calculation of volumes proposes an extension of the Monte Carlo algorithm to 3D point clouds. The validation of this algorithm is carried out using objects of known volume so that the uncertainty of the measurement can be well defined, unrelated to the uncertainty of the acquisition of the point cloud. The algorithm shows very promising results and returns a volume estimate with an error of less than 7% in a reasonable time. The final accuracy is consistent with state-of-the-art performance and a few percentage points lower than commercial systems, such as Vitronic and Artec2. However, the system is overall significantly cheaper than its commercial counterparts, costing approximately five to ten times less. In addition, the system allows the acquisition of disabled subjects who cannot stand and can calculate the volume of even partially incomplete point clouds relying only on custom made or open source core software. A more complete validation of the entire measurement system is currently underway, including the uncertainty associated with the point cloud acquisition itself and the estimate of inertia and masses.

Table 2.1 Results Comparison with Virtual Objects

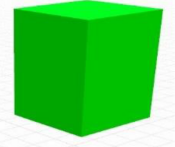
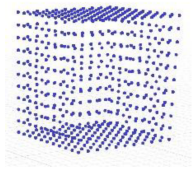
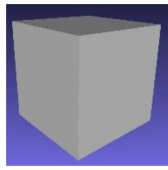
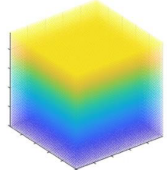
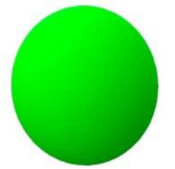
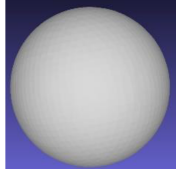
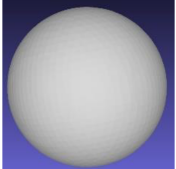
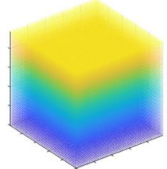



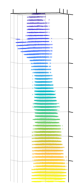



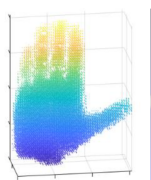
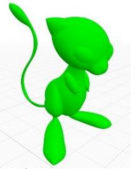
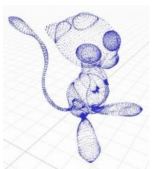
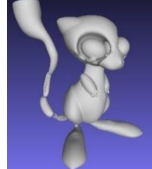
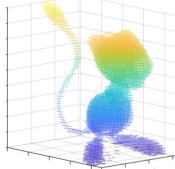



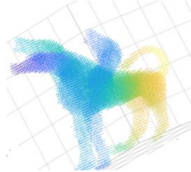



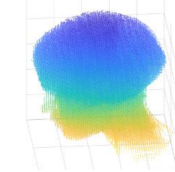

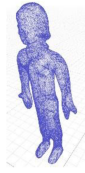

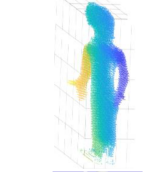



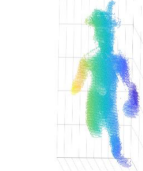
Original Model	Original PointCloud ($S_{3DTargetObject}$)	Meshlab reconstruction	Montecarlo output
			
			
			
			
			

Table 2.2 Results Comparison with real Objects

Original Model	Original PointCloud ($S_{3D}TargetObject$)	Meshlab reconstruction	Montecarlo output
			
			
			
			

Chapter 3

sEMG Sensor for monitoring the muscle activation and for Action Recognition

3.1 The Electromyography

Electromyography (EMG) measures the neuromuscular activity generated to activate the muscle fibres during the contraction [104]. These EMG signals are widely used in numerous biomedical applications, such as monitoring the motor activity of patients while walking with exoskeletons and performing rehabilitation [108], studying muscle synergies [52],[100], diagnosing neuromuscular disorders[35],[88], [58]. Still in the medical field, electromyographic sensors are also used outside the clinical environment in human-related applications, such as the control of wearable robotic prostheses, and action recognition [9] or visual feedback for mixed reality applications[111]. There are different types of EMG sensors. First of all, they can be divided into superficial sensors and intramuscular sensors. Surface electromyography (sEMG) sensors consist of non-invasive electrodes that acquire the electrical signal by resting on the skin. These sensors are considered non-invasive, are easy to use and do not pose a great risk of infection even in a non-sterile environment. However, this type of electrode cannot detect deep muscle activation, but only the more superficial ones whose signal reaches the surface. These sensors are also affected by cross-talking [36],[103] Cross-talking occurs when the electrode collects the electrical signal from several muscles and each contribution is no longer discernible. In addition, sEMG is strongly influenced by skin surface conditions, such as the presence of hair or skin sweating, and by environmental parameters, such as wearable devices for sensor support or electromagnetic disturbances due to mains voltage. These conditions can influence several parameters, such as surface contact impedance, electrode adhesion, electrode oxidation, electrode short-circuit and the presence of noise. However,

multiple solutions have been proposed to mitigate these problems and minimize the effects of noise sources that jeopardize signal acquisition as reported in [31], [73], [115],[91],[84].

Intramuscular EMG is characterized by intramuscular electrodes that are inserted directly into the subject muscle using a needle. The intramuscular electrode is little affected by cross-talk and external environmental conditions. It also allows the measurement of the activation state of the deepest muscles with higher signal quality [112]. In addition, if inserted and fixed correctly[95], there is no risk of detachment from the muscle, guaranteeing a good and consistent signal acquisition [46]. However, intramuscular electromyography is invasive and requires sterile instrumentation, appropriate dressings and experienced staff to insert and remove the electrodes. In addition, the intramuscular electrode placement can be long and tedious, causing stress for the subject involved.

Depending on the application, it is, therefore, necessary to carefully assess which type of electrode to use. In our case, for example, most of the muscles involved are superficial and we focused on using EMG sensors for acquisition. Among the most popular commercial sEMG sensors are MyoArmband [89],[9], DelsysTrigno¹ [58] and Biometrics EMGsensors². The Myo armband is a bracelet consisting of eight differential sensors that communicate via Bluetooth with a laptop device. It is particularly easy to wear and it is possible to find open-source algorithms that allow real-time data communication. The cost of the device is low, around a few hundred euros. However, Myo is a relatively simple device and does not allow a sampling rate higher than 200 Hz. Hence, it cannot acquire the full spectrum of EMG frequencies which ranges up to 500 Hz. But still, MyoArmband can be used for simple action classification [40] and muscle activity monitoring.

The Delsys trigno is a high-standard EMG sensor. It is usually sold as a multi-channel kit and has a separate differential sensor for each channel. It can acquire at a high sampling rate 4kHz [58] and has very good signal quality. Wireless communication takes place between the sensor and the charging case to which it is possible to connect the laptop via cable. The sensors are registered and charged via the same case. The Delsys Trigno is widely used in the biomedical field and is often considered as a reference for comparing different sEMG devices[86], combined with IMUs for biomechanical analysis and the assessment of physical activities [87] or used for specific clinical evaluations, such as in [58]. However, the cost of these sensors is particularly high, running into tens of thousands of euros for a basic kit. For this reason, not all research groups can take advantage of them and rely on cheaper solutions. After Myo Armband and Delsys Trigno, other sensors such as Biometrics or

¹<https://delsys.com/trigno/> accessed: Jan. 2022

²https://www.biometricsltd.com/surface-EMG-sensor.htm?gclid=CjwKCAiA8bqOBhANEiwA-sIIN2HDYwyOFaWKtSWi4tAp10MfmRHoccoAZ1d8HzEfCyVkkxbKhiwWIBoCOzEQAvD_BwE accessed : Jan.2022

Comet are somewhere in between in terms of cost and performance. Another big difference, however, is whether they have wired or wireless communication with the storage device. If on one hand the presence of cables simplifies the hardware components on-board and reduces its cost, on the other hand, it implies a greater complexity of use for the subject. The presence of cables generates a risk of obstacles to movement and requires greater attention during the acquisition. In addition, the presence of cables between the electrode and the acquisition device means that further treatments (analog or digital filtering) need to be taken into account while the signal processing occurs. This is due to the increase of noise from environmental sources, such as mains voltage or cable oscillation. Consequently, sEMG sensors have a fairly standard hardware architecture. Since the biological signal has an amplitude of a few millivolts, sEMG sensors consist of one or more amplification stages to increase the signal amplitude. To reduce the noise at 50 Hz they often present an analog notch filter that reduces the main component of the mains voltage. Sometimes this is also performed in post-processing with a digital notch filter to lighten the hardware components. A 10 Hz high-pass filter is often inserted to filter out low-frequency noise not related to muscle activity. Finally, to avoid aliasing during sampling, a low-pass filter is chosen according to the acquisition frequency. Usually, the cut-off frequency never exceeds 500 Hz, since the EMG signal does not present large components after this frequency.

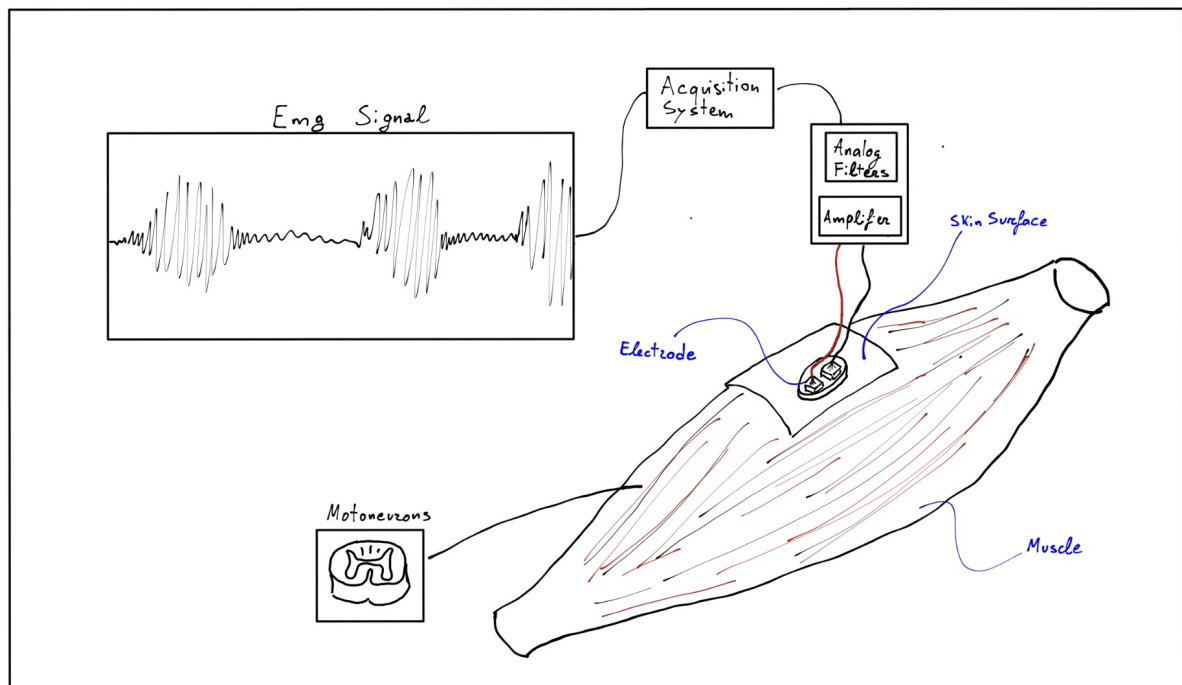


Fig. 3.1 Generic sEMG signal acquisition pipeline

3.2 Action recognition in the clinical Environment

Action recognition in the clinical domain is a widely studied aspect for a large number of applications, such as home care for personal support and safety [68], analysis of daily behaviour to diagnose inconsistencies or medical needs, gait analysis [78],[72],[74],[117] or action recognition for the control of intelligent robotic prostheses. Action recognition plays an important role in many different aspects and can be addressed in many different ways.

Action recognition with machine learning techniques has often been addressed using different sensors. Common approaches for action recognition include inertial measurement units[44], electromyographic sensors[16],[63], RGB cameras [98], and depth cameras [114][2]. All the different approaches have led to more or less good results depending on the type of actions chosen for classification, the type of algorithm trained and the amount of data acquired for training. In addition to the performance of the approach itself, the choice of the sensor is often characterized by the type of application for which it is designed, in terms of cost, measurement effectiveness, and convenience of use. On one hand, if the subject in question moves in a defined and limited space with few visual obstructions, it is possible to opt for a recognition system based on RGB or depth cameras. A common example can be found in the Ausilia project domestic apartment [85], where the subject is confined to a domestic environment consisting of a few rooms. Equipping each room with the appropriate number of cameras and being able to exploit the appropriate algorithms for image processing is possible to monitor the subject and recognize the actions. In the case of the Ausilia project, this allows the development of an intelligent home automation environment able to study and help the patients involved. In the future, this technology can be easily extended to give help and support to elderly subjects living alone. Other examples and applications that leverage computer vision for action recognition have been developed in the areas of on-site clinical analysis (such as the gait analysis on the treadmill) [51], rehabilitation [64], and gaming [102]. In all three areas, the subject does not make large movements and remains within a small, well-defined area of movement. On the perimeter of such area usually are disposed a variable number of cameras that enrich the whole vision of the subject from different perspectives.

On the other hand, it can be observed that the use of wearable sensors with wireless communication is particularly suitable in cases where the subject is free to move in a large space or the presence of visual occlusions for a camera-based system. Wireless inertial or electromyographic sensors are easily integrated into clothing or elastic bands that allow for subject adherence and wearability. In the case where movements can be easily associated with muscular patterns, where the raw data can be processed for clinical sport to study synergies with movement, or, as in the case of this research work, the information on the

state of muscular activation is further used as visual feedback for the therapist, it is very convenient to use sEMG sensors for the acquisition of data.

In addition, where actions do not involve large or rapid movements, inertial sensors are particularly penalised. Inertial sensors acquire accelerations and are poorly suited for the classification of small movements or stationary actions due to the low amplitude of the signal, which inevitably results in little significance and makes it more difficult to classify the gesture. In contrast, EMG sensors manage to monitor the muscles involved well, allowing classification of the gesture even in the case of static or semi-static actions. This is why inertial sensors and EMG sensors are often considered complementary in the literature [8].

3.3 Action recognition

sEMG pattern recognition suffers from different perturbations, some of which are caused by electrode shifts, anatomical variability, subject posture, muscle fatigue, and adhesion of the electrode to the skin [10]. Above all, the shift of the electrodes is one of the most deleterious perturbations. Unwanted rotations and translations of the electrodes often occur during the subject movement and the muscle contraction. It usually causes a remarkable deviation of signal features from those used to train the recognition algorithm. The resolution of such a problem represents a challenging achievement for all those applications which uses bracelet sensors. Furthermore, it defines a fundamental achievement towards a more robust classifier of practical use which permits better repeatability. Although approached from the image classification perspective with convolutional neural networks, a similar problem has been already faced by Szegedy et al. in [109] to increase the image classifier robustness. As proposed by LeCun et al. [61], data augmentation has led to remarkable improvements and it turns out to be as important as the network architecture choice. Jia Shijie et al. [99] investigated the impact of using a few different techniques for data augmentation. Data augmentation is widely used to strengthen the CNN image classifier performances and make them more robust to geometrical perturbations, such as rotation, translation and rescaling of the information. In the work proposed in [57] they augment the starting dataset by introducing new elements derived from image flipping and random cuts. However, these methods have been little tested in the field of EMG signals. Palermo et al. [81] focused on the repeatability of grasping operations for the control of a robotic hand by using sEMG signals and by taking advantage of the available NinaPro dataset, which is delsys sensor based. They considered the grasping action of 10 subjects and noticed that the model's accuracy decreased by more than 25 % over different repetitions. This evidence showed them how sensitive the model can be to this perturbations and, as a consequence less robust.

In addition to the problem description, their work does not address an actual solution, but it proposes a further analysis of the problem and suggests to mitigate the problem by creating a larger dataset considering more actions. Akhmadeev and Al. [3] used MyoTM Armband to monitor the muscle activation and used a State Vector Machine to classify the gestures. They decoded six hand motions from 7 subjects, and their model reached a classification accuracy around 92% during online validation. Moreover, they studied the model robustness by placing and removing the sensor device several times and by comparing the accuracy achieved. They obtained promising results in terms of classification stability but they did not propose any data augmentation to compensate the electrode shifting. Other recent results had been illustrated by Zanghieri et al. [129]. During the robustness assessment, they tested their classification algorithm on a self build dataset. On It appeared that the accuracy on new acquisitions dropped only of just 3.4 %, against the SoA results, which typically report a drop of more than 5.0%. They mainly focused on energy efficiency and well investigated the performances of the classifier over time. The system they proposed exploits a network based on the temporal convolution for gesture decoding and relies on a custom-made front-end, using standard electrodes gel-based. Li et al.[65] proposed a k-nearest-neighbors (KNN) to recognize 10 hand gestures developed as an online semi-supervised learning model. They trained the classifier by using a small amount of labelled samples and updating it by using unlabelled data. Two datasets were generated using different electrode configurations during three days of acquisitions. A total amount of one hundred samples for each hand motion were used to create the first dataset, while a total amount of three hundred for the second dataset. On the first day they collected labeled sample while the days later, when usually the electrode shift occurs, were acquired unlabelled data. They have chosen the wave length (WL) feature for their experimentation. The model robustness is evaluated using the data collected of the second and the third day. It returned a mean prediction accuracy of 65.8% on the first dataset and 91.6% on the second dataset. The recognition of hand gesture with the sEMG signals monitoring has been also investigated by Xu et al. [121] using an artificial neural network with a final vote based on the majority. They used fiftyfive consecutive samples for the classification of 6 gestures. An innovative approach, based on the synchronous reset gesture and the interpolated peak location (IPL), is described to detect and correct the electrode shift. Data are acquired by considering eight experimental acquisition of ten subjects to perform the gestures movement with different configurations of electrodes. The Device used for the data acquisition is a MYO armband sensor. Each samples is defined considering a 200 ms interval of time (40 sampling values) over which seven time-based features are extracted. The comparison on the prediction accuracy before and after the electrode shift correction shows an improvement of 22.7% over the reported tests: 72% with MYO standard electrode

configuration and 94.7% with position correction. Several works do not address a possible data augmentation to solve the electrode shift problem, but instead propose different models based on Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) or Gated Recurrent Units (GRUs) architectures to extract the sEMG temporal features and correlations. For example, the research proposed by Simao et al. compares the use of different Artificial Neural Networks (ANN) highlighting the superiority of the LSTM performances [101] with respect to the results achieved by a simple RNN. In fact, the LSTM appears to present a structure particularly suitable for evaluating temporal features between sEMG channels.

For all these reasons in our work different neural network models are trained and compared to classify hand gestures using sEMG forearm acquisitions and taking particular care of the network performances, such as the robustness and the accuracy. This is enhanced by a custom acquisition system tuned to the specific application. A data augmentation technique is proposed to simulate possible electrode shifting and the related perturbations. Finally, an ensemble architecture of specialized LSTM is proposed to deal with the classification of actions with similar muscular pattern activation. We chose LSTMs to focus on the dynamical characteristics of sEMG signals as this kind of Neural Networks(NN) is widely adopted by many human activity recognition (HAR) researchers for EMG classification and pattern recognition [41],[43],[126]. The training dataset is augmented by simple data augmentation techniques showing a significant improvement in network robustness. The results were experimentally verified by comparing the results with those reported in the literature and it will show that specialized neural networks for sequence classification are particularly suitable for this kind of application. Moreover, a simple LSTM appears to be more suitable than classical CNNs when using a low-power processor. This is due to the high computational demand of the latter. DNNs instead did not result able to perform well due to their structural simplicity unable to deal with time-varying features. In this case, it would be required a further computation of features with the risk of burdening the overall computational cost, introducing delays and compromising the real-time applications.

3.4 Hardware architecture and set-up

The prototype of the developed acquisition system consists of a fabric sleeve, stainless steel electrodes coaxial cables, a conditioning system, an analogue-to-digital conversion system and a Raspberry Pi unit for signal storage and processing. The textile sleeve is equipped with Velcro strips that allow the reuse and repositioning of the electrodes. The electrodes consist of stainless steel plates. As shown in Figure3.2, different geometries and arrangements were tried to maximize the signal-to-noise ratio and best visualize the acquired signal. In the end,



Fig. 3.2 Testing of different geometries and electrode arrangements to maximise the signal-to-noise ratio

a horizontal arrangement of electrodes was chosen. All the electrodes measure 20x20x1 mm and were produced manually by cutting, drilling, and polishing in the mechanical machining laboratory of the University of Trento. A small copper wire was soldered to each plate to allow it to be soldered to the coaxial cables. A plexiglass case cut with a laser cutter was then used to position each pair of electrodes at a fixed distance. A fabric strap was glued to the back of the plexiglass case to attach it to the Velcro of the sleeve. Due to the small amplitude of the acquired signal, any connection of the electrodes with the conditioning system is allowed via coaxial cables, wiring the central core with the electrode and the shield with the common reference. This reduces interference and noise from electromagnetic fields caused by the numerous devices in the test area.

The conditioning system consists of several amplification stages with a total gain factor of 500 and three different analog filters. The first is a selective filter that attenuates the 50 Hz signal component due to the large disturbance introduced by the main frequency. The second filter is a high-pass filter with a variable cut-off frequency at 5-10-15 Hz. The high-pass filter allows excluding the bulk of the low-frequency oscillatory components introduced by the movements of the subject and the oscillation of the cables. Finally, a low pass filter at 500 Hz allows selecting the information content of electromyographic signals by attenuating the high-frequency noise components and avoiding the introduction of the Aliasing effect during the sampling phase.

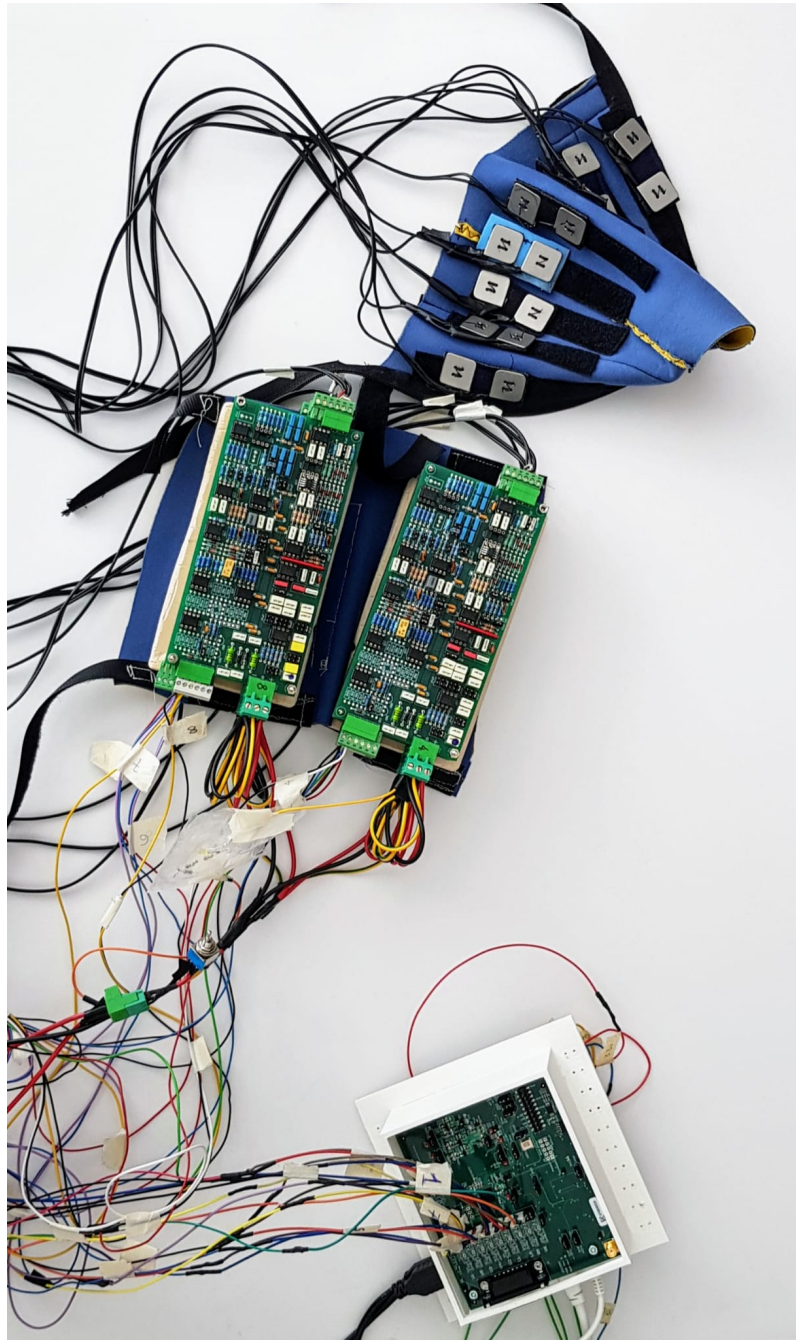


Fig. 3.3 Prototype of the 8-channel EMG acquisition system

3.4.1 Conditioning system

The conditioning system has been designed and developed by Ing. Antonio Selmo, in collaboration with the company Selmo ing. Antonio Franco, T4I. The system presents a standard structure and it is composed of a differential amplifier followed by a sequence of

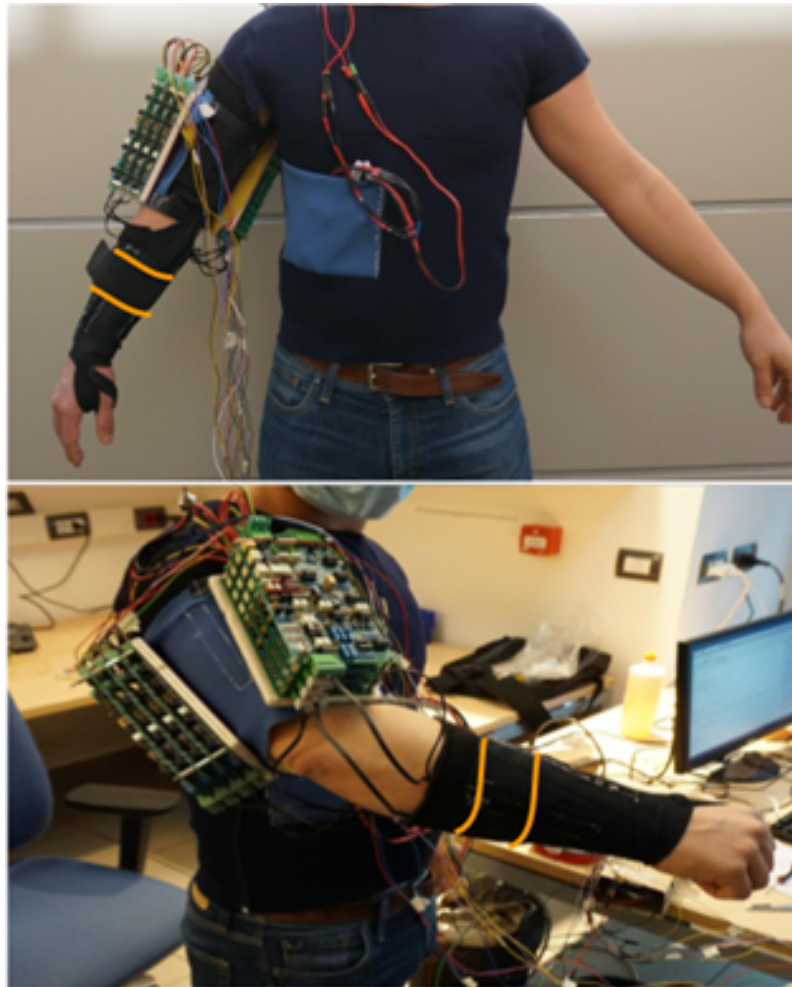


Fig. 3.4 Acquisition tests wearing the prototype of the developed system

common-mode amplification stages. The low pass filter has been placed to remove high-frequency noise and avoid aliasing problems during sampling. It has a two-stage Sellen & Key structure that allows obtaining a Bessel filter with a 5th order response and a cut-off frequency set at 500 Hz. This reduces the over-elongation in the step response that characterizes traditional Butterworth and Chebyshev filters. In addition, the circuit was equipped with a high-pass filter and a notch filter. The high pass filter always features a 5th order Bessel response and a variable cut-off frequency, 5Hz, 10Hz, 15Hz to remove low-frequency artefacts. The notch filter instead is specifically designed to reduce the main noise component at 50Hz, due to the mains voltage and propagated to all powered devices in the test area.

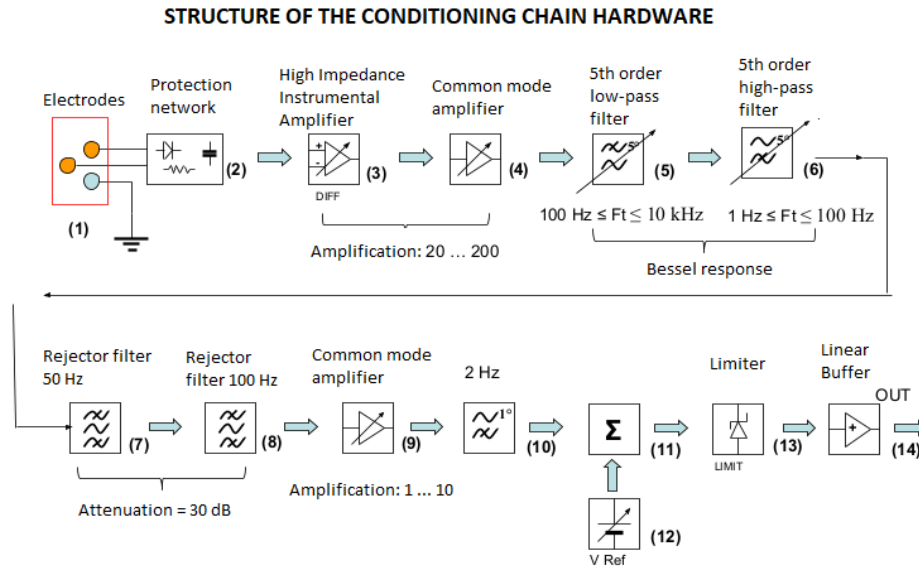


Fig. 3.5 Conditioning system architecture, designed and developed by Ing. Antonio Selmo

3.4.2 ADC and Processing Unit

The analog-to-digital conversion system was chosen based on acquisition performance and the maximum amount of input channels. The Texas instruments ads1298 device is a commercial 8-channel analog-to-digital converter (ADC) that allows for acquiring up to 32 Ksps, thus allowing to sample of each signal at 4 kHz at 16 bits. The device features a Delta-Sigma ADC architecture and can handle differential and single-ended inputs. The Spi communication protocol allows fast transfer of acquired data to the Raspberry processing unit. The processing and data storage unit consists of a Raspberry Pi 3 Model B board. The programming language for managing the ADC registers, setting the sampling rate, sample resolution, and data processing are Python 3.7 in a Raspbian environment (Raspberry Pi Os). In this context the Tensorflow open-source platform for machine learning and Keras libraries have been exploited to structure the architecture of the neural networks, to perform the training process and the validation.

3.5 Deep Neural Network Architecture

For the proposed work, different classical neural networks were taken into consideration, among which some simple fully connected layers DNNs, some convolutional neural networks (CNN), and some long short-term memory neural networks (LSTM). Following a preliminary analysis of the performances and based on what has been suggested in the literature, a long

short-term memory neural network was chosen with an architecture similar to that proposed by M. Simão et al. in [101]. Fully connected layer neural networks are the simplest example of a neural network that can be proposed and therefore are not specialized for specific inputs. For this reason, require pre-processed input data in which relevant characteristics have already been highlighted. The fully connected layers networks present a simple sequence of cascading layers in which each neuron of the next layer is connected to each neuron of the previous layer. For information, typically these networks present simple neurons whose function is composed of a linear and a nonlinear part. The linear part involves a summation of all the states of the neurons of the previous layer, each pre-multiplied by the corresponding weight of the network, the weights. Once the summation is done, a parameter called bias is added. The non-linear component follows the linear component and is characterized by an activation function. The activation function " ϕ " is a non-linear mathematical function that takes the output of the summation with the bias and returns the evaluated value, " $\phi(x)$ ". There are many activation functions, and each of them has specific characteristics. Among the most commonly used are the sigmoid function, the arctangent, the rectified linear unit, and related variants. During the process of training and learning in the case of supervised learning, the output of the neural network is compared with the ground truth supplied by the Training dataset. This difference characterizes the cost function. An optimization process will try to reduce the cost function by adjusting the parameters with a specific criterion. There are many optimization processes and for this work, the most traditional ones were chosen and evaluated, GD (Gradient Descent), SGD (Stochastic gradient descent), and ADAM (Adaptive optimization algorithms). Such networks are often cascaded to specialized networks once a pre-processing of the data extracts features from the provided input. For this reason, we have selected and calculated the most common features proposed in the literature [77], [123] such as Waveform Length, Mean Absolute Value, Zero Crossing and Slope Sign Change.

Convolutional neural networks (CNN) are specialized algorithms for image processing. As reported in the work of Meng et al. [71], [12], such networks have also been used for the analysis of EMG signals. The classical CNN structure presents a sequence of convolutional layers followed by fully connected layers. The input image is expressed in the form of a tensor or as an array of matrices (for example, in the case of RGB images the information is represented by three matrices of equal size, one for each colour channel). Hence, given an input image, the convolutional layers perform geometric feature extraction to allow the fully connected layers to classify it more efficiently. Convolutional layers extract geometric features using the convolutional filters. These filters, such as the Sobel filter, are represented as matrices whose dimensions and elements are selected according to their functionality. They work as image masks and filter the data, allowing only the desired information to

permeate. They usually look for simple geometric features such as horizontal and vertical lines, edges, triangles, squares or circles, but more elaborate features can also be sought with more elaborate approaches. Once the necessary information has been extracted, the image is compressed through a pooling operation (usually max pooling to keep only the most chromatically relevant information) that reduces its size by concentrating the information in a smaller matrix. This operation allows the most salient information to be highlighted and geometric information to be extracted on several levels. Once all the geometric-chromatic features have been extracted from the raw data (the input image), a flattening operation is performed and everything is passed to the fully connected network that will proceed with classification. In the case of sEMG biological signal classification, it is possible to exploit the CNN architecture by producing an RGB representation of certain signal characteristics. For example, by reporting a coloured representation of the wavelet transform of the signal. Unfortunately, CNN networks and data pre-processing require a high computational cost and make real-time implementation on small processing units difficult.

LSTM (long short-term memory) networks are specialised in the analysis of data sequences. They are an evolution of the traditional RNNs (Recurrent neural networks) overcoming the problem of vanishing gradient in the presence of long sequences. The LSTM networks have a gate architecture managed mainly by the presence of sigmoid and arctangent activation functions. This characteristic architecture allows the processing of the input at the time $t+1$ remembering the relevant information previously processed and forgetting the less important information. Although seemingly intricate, the LSTM architecture is simple and functional, allowing light pre-processing of the data without requiring a high computational cost. The low computational demand combined with the characteristic ability to handle data sequences makes LSTM networks particularly suitable for our application. To pursue a comparison within different networks, a few DNN, CNN, and LSTM architectures were trained and tested on a sub-portion of the same open-source EMG Ninapro dataset³ acquired with Delsys sensors and available online.

Considering different networks trained and tested on Nina Pro datasets, as shown in the Figure3.6, it can be observed that the Lstm networks scored a higher and more homogeneous accuracy than the CNN and DNN networks, which had a lower average accuracy with higher variance. The higher variance indicates how much the better the result changes as the depth of the network changes while keeping the underlying architecture (DNN, CNN, LSTM) unchanged. For these reasons, the best-performing LSTM network structure has been retained as shown in the Figure3.7.

³<http://ninapro.hevs.ch/node/123> accessed: 01/01/2022

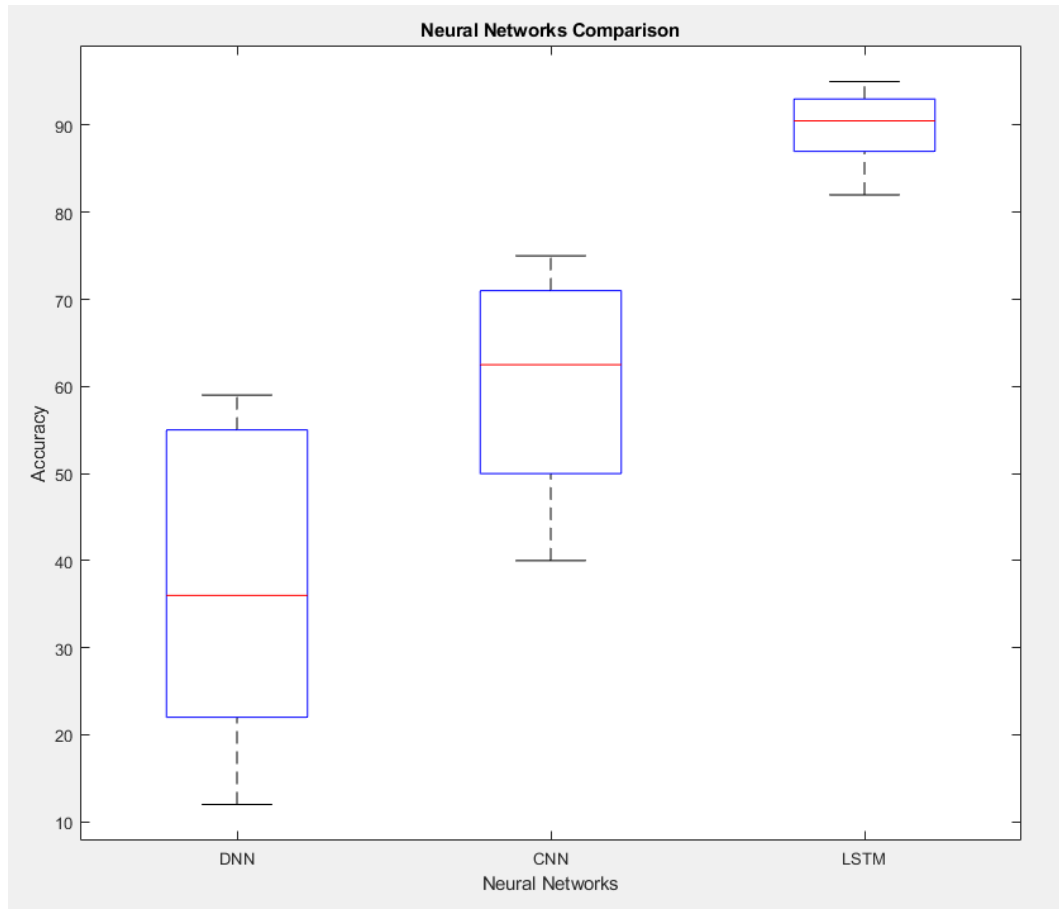


Fig. 3.6 Comparison on the results obtained by training different types of neural networks with the same sub-portion of the EMG open-source Nina Pro dataset of 8 forearm muscles

It is important to note that, unlike the architecture proposed [101], the network we propose introduces layers of batch normalisation downstream of each dense layer. With these additions, we have obtained a greater speed in processing the data and greater stability in classification as also suggested in the literature [45].

3.6 Generation of Dataset and Training

Following the definition of the neural network, we proceeded with the creation of the dataset for training and testing. The dataset was generated by considering a total of six static actions corresponding to a specific hand configuration. The chosen configurations (which define the action recognition classes) are the following:

- Hand closed (ch_)

```

# NETWORK ARCHITECTURE DEFINITION
verbose, epochs, batch_size = 1, 15, 32
#define lstm parameters from dataset size
n_timesteps, n_features, n_outputs = Xtrain.shape[1], Xtrain.shape[2], ytrain.shape[1]
model2 = Sequential()
#return_sequences=True
model2.add(Dense(50, input_shape=(n_timesteps, n_features)))
model2.add(BatchNormalization())
model2.add(Activation('tanh'))
#return_sequences=True
model2.add(LSTM(50, activation='tanh'))
model2.add(BatchNormalization())
model2.add(Dropout(0.2))
model2.add(Dense(n_outputs))
model2.add(BatchNormalization())
model2.add(Activation('tanh'))
model2.add(Dense(n_outputs, activation='softmax'))
opt = tf.keras.optimizers.Adam()
model2.compile(loss='mean_absolute_error', optimizer=opt, metrics=['accuracy'])
history2=model2.fit(Xtrain, ytrain, batch_size=batch_size, epochs=epochs, verbose=verbose, validation_split=0.2)
# evaluate model
loss, accuracy = model2.evaluate(Xtest, ytest, batch_size=batch_size, verbose=1)

```

Fig. 3.7 LSTM Network Architecture

- Hand relaxed (rx_)
- Open Hand (oh_)
- Closed hand with index-medium fingers raised (2_)
- Closed hand with index-medium-ring fingers raised (3_)
- Open hand with joined index finger and thumb (ok_)

Electrodes arranged radially in the sleeve were applied to the subjects' forearms, allowing the acquisition of EMG signals with the eight differential channels. Similar to the Myo armband device, this configuration does not bind the individual channels to a specific muscle bundle, but favours quick wearability of the sensor, taking into account cross-talking effects during training. One acquisition session involved holding each action statically for 20 seconds. A total of seven acquisition sessions were performed for each subject. By acquiring the signal at 4 Ksps a total of 3360000 samples (4000 samples per second x 20 seconds x 6 actions x 7 sessions) per subject were collected. A time window of 20 milliseconds with a 20 % overlap was chosen for data processing. Each window presents a sequence of 800 consecutive samples of the 8 acquired channels.

Representing the information obtained as a matrix of 8 channels x 800 samples, it is possible to proceed with the operations of noise cleaning and normalisation with respect to the mean value and standard deviation reported in the time window. To concentrate the information contained within the window, an average amplitude change (ACC) operation was applied by reducing the matrix size to 8x50. The data thus prepared is then appropriately labelled and added as an element of the Training and Testing dataset.

3.7 Results

Training the proposed neural network on 90 % of the generated dataset and testing it on random samples of the remaining 10% resulted in an average accuracy level of 90% and a variance of 1.47%.

However, performing further tests with acquisitions taken days apart and removing and reinserting the sleeve several times, a drop in performance was noted with an average accuracy of around 70%.

This is mainly due to the different acquisition conditions that occur each time the sleeve is removed and put on. Under these conditions, the electrodes may be slightly rotated clockwise or counterclockwise, introducing an unwanted component of the adjacent muscles due to the cross-talking effect. For this reason, we introduced an augmentation of the dataset taking inspiration from the augmentation techniques typical of image classification with CNN. In the case of object classification with images, it is common practice to augment the original dataset by introducing artificial perturbations to the data. These perturbations include rotation, cropping, zooming, and other modifications of the elements of the initial dataset to make the network more robust while testing on new inputs [39]. Similarly, we introduced an artificial perturbation of the original data by creating new elements of the dataset and simulating a cross-talking effect. For example, simulating cross-talking in the case of a clockwise rotation, each channel of the artificial acquisitions is created as a weighted combination of itself and the following channel considering a variable random multiplier (ranging from 0.9 to 1.1) of the two channels.

Looking at the confusion matrix, as shown in the Figure3.8, it can be seen that a significant proportion of the misclassification is due to the similarity of the patterns of several actions, in particular, "2_", "3_" and "ch_".

For this reason, we thought of introducing a new architecture, as shown in the Figure3.9, to deal with this similarity problem and help the network discern between similar actions. The proposed structure shows an ensemble of LSTM networks specialised according to the one-against-all criterion. This technique allows simplifying the classification problem into a

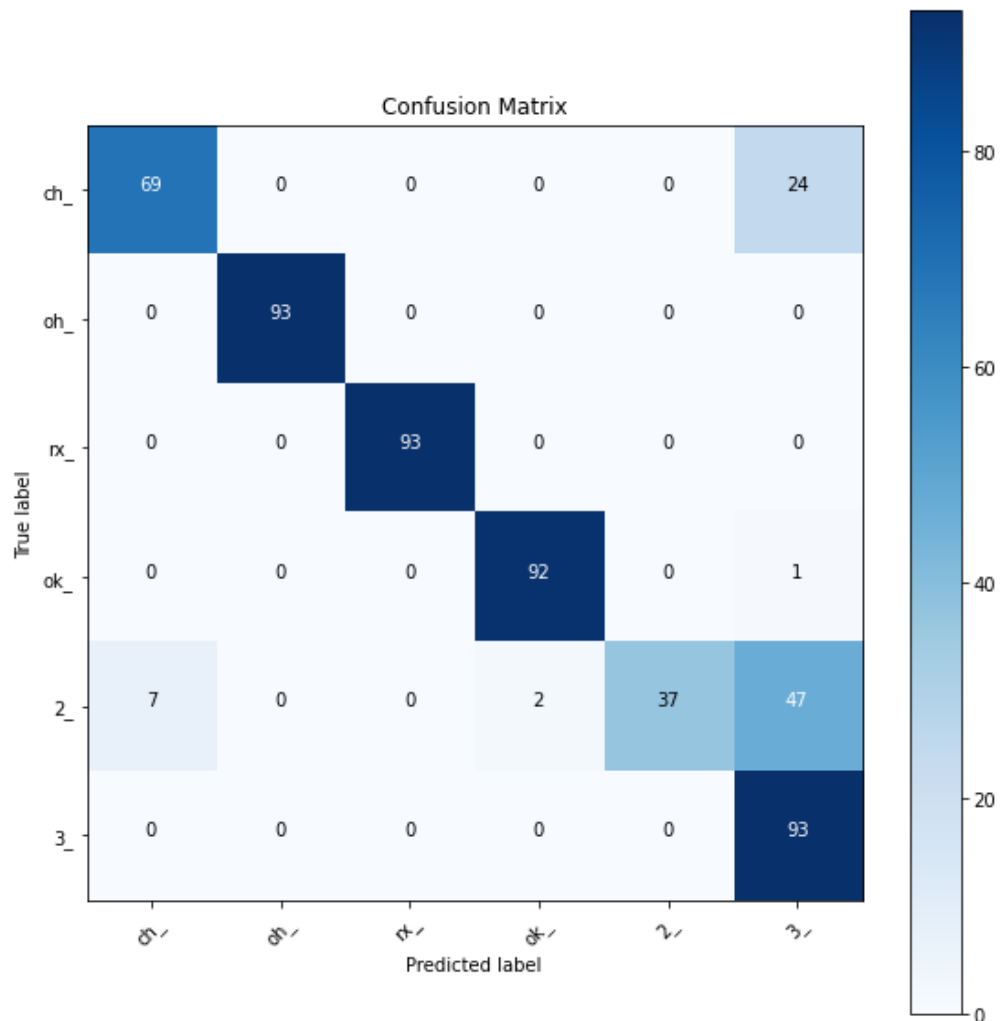


Fig. 3.8 Confusion matrix obtained by training and testing the LSTM neural network

binary choice that, if properly trained, guarantees higher reliability of the result. However, this does not yet guarantee high accuracy in the final overall classification, so, by training a further specialised network on similar actions, the classification is further checked and the final accuracy higher.

As can be seen from the Figure3.10, the result achieved with the introduction of the augmented dataset and the ensemble of specialized networks alone increased the accuracy of the network under test. In particular, by training the ensemble of networks with the augmented dataset it was possible to obtain an average accuracy higher than 85% along with the robustness assessment.

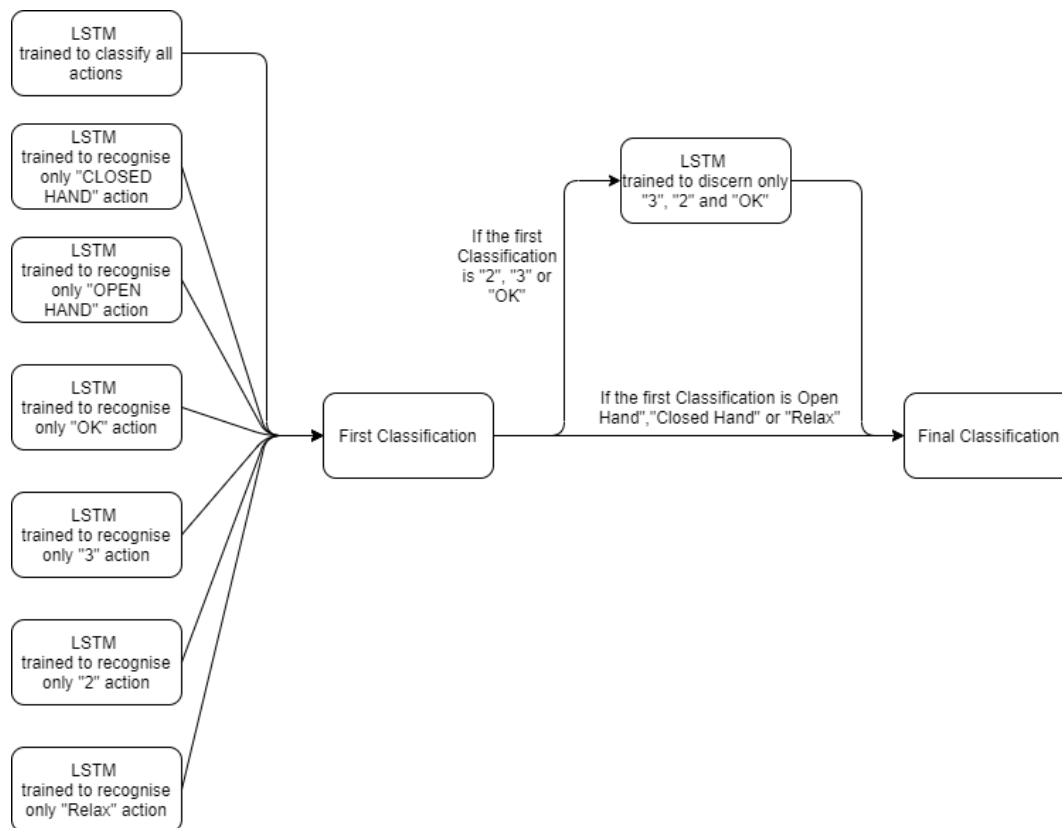


Fig. 3.9 Ensemble of specialized LSTM neural networks

Finally, purely for visualisation purposes, as shown in the Figure3.11, the classified actions were reproduced by a simple robotic arm controlled by Raspberry with PWM digital control, limited rotation servomotors (-180° , $+180^{\circ}$) and cable transmission ⁴.

3.8 Conclusions

This chapter reports the development of a versatile measurement system with reusable electrodes for the acquisition of electromyographic signals and the estimation of the activation state of the monitored muscles. In addition, this study reports the development of a classification algorithm for action recognition and control of a possible robotic prosthesis. The paper proposes a comparison of different neural networks proposed in the literature for the classification of EMG signals. The comparison is made by training DNN, CNN and LSTM networks respectively on the open-source and widely used Nina Pro dataset acquired with Delsys Trigno. The comparison shows that LSTM achieves significantly higher accuracy

⁴<http://www.miro.ing.unitn.it/gesture-recognition-with-lstm-classification-of-sEMG-signals-for-prosthetic-control> accessed:07/02/2022

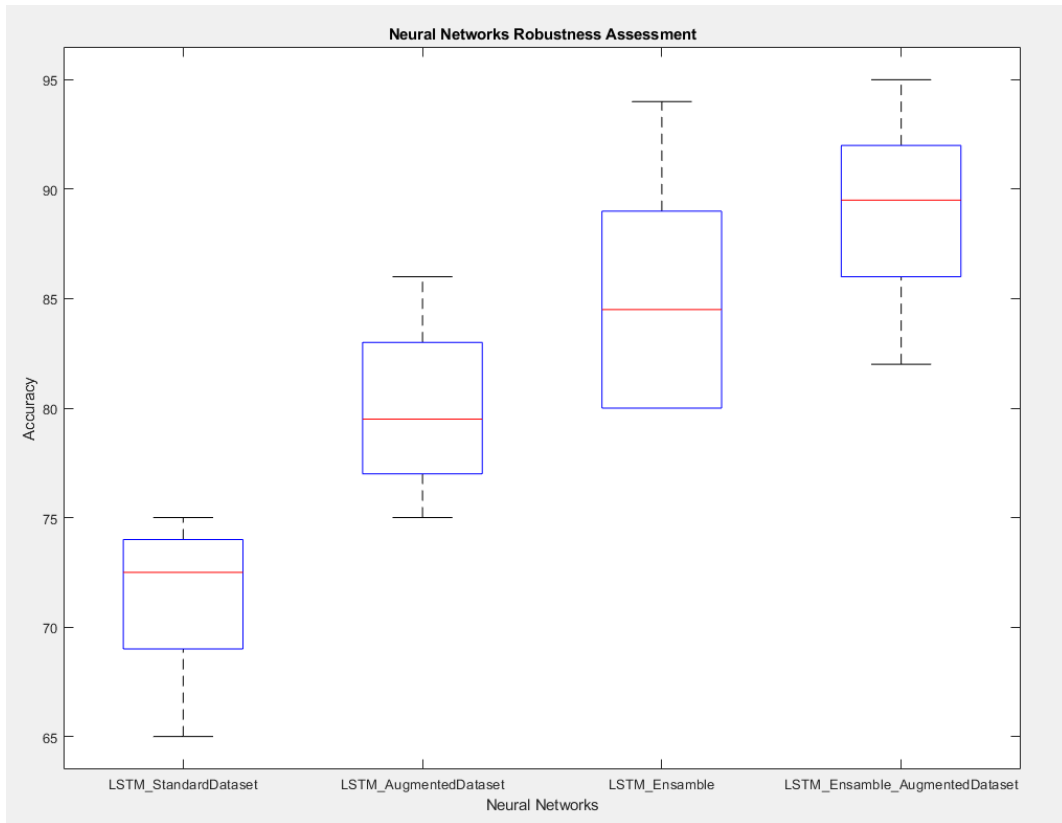


Fig. 3.10 Comparison of results obtained with specialised LSTM neural network ensemble and dataset augmentation

and is particularly suitable for the classification of electromyographic signals due to its low computational weight, confirming what has been reported in the literature [101]. Particular care has been taken regarding the robustness of the system. As reported in the state of the art, the classification of EMG signals is often perturbed by the repositioning of the sensor and by the cross-talking effect. Trained systems often lost tens of percentage points of accuracy when tested on samples acquired on different days or by repositioning the sensor with a slight displacement from the training dataset configuration. In contrast, even with slight changes of the electrode configuration, the system we proposed returns an accuracy loss lower than 10 % and an average accuracy score higher than 85 %. This is mainly due to the training dataset's augmentation, which artificially introduces the cross-talking and displacement artefacts, and to the characteristic structure of the action recognition pipeline. The action recognition pipeline presents a cascade of parallel LSTM architectures. Each LSTM classifier proposes a variant of the LSTM architecture proposed by M. Simão et al. [101] and introduces batch normalization within the layers as suggested [45] to achieve a greater speed in processing and

classification stability. Each architecture is specialized in specific patterns and this allows one to discern different actions which in contrast present similar muscle activation patterns.

The developed sensor is only presented using PTH (Pin through-hole) electronics components in a prototype phase. The final version is currently being miniaturised and will consist mainly of SMD (Surface Mount Device) components to reduce the size of the analogue electronics and be easily integrated into wearable robotics and clothing. The final version will feature an overall compact wireless system that is more robust to electromagnetic disturbances. Furthermore, it will be Arduino Nano compatible. In this way, a simpler and easier code interface will allow also inexperienced users to customize the code on their needs. The electrodes will be mounted directly on the conditioning board and the acquisition system will be placed above it using a shield. It will be powered by 1s LiPo batteries and transmit the acquired data via wireless communication using MQTT (Message Queue Telemetry Transport) or Bluetooth.

The development and the study of this prototype allowed us to set the bases for future applications in line with the gait cycle analysis reported in chapter one and the rehabilitative visual feedback reported in chapter four. In this case, the sensor will be relocated and will monitor the leg muscles to identify the gait phases automatically. The identification and subdivision of the gait phases are often performed manually or by exploiting a simple algorithm of pattern identification. The autonomous subdivision of gait phases instead could be of great importance for the gait analysis and may help the researchers by streamlining the process and correlating the forces with the involved kinematic configuration. Furthermore, the good performances achieved on the small arm muscles with similar patterns return a promising expectation of the performances that the system could achieve on the greater leg muscles while classifying different and more defined activation patterns.

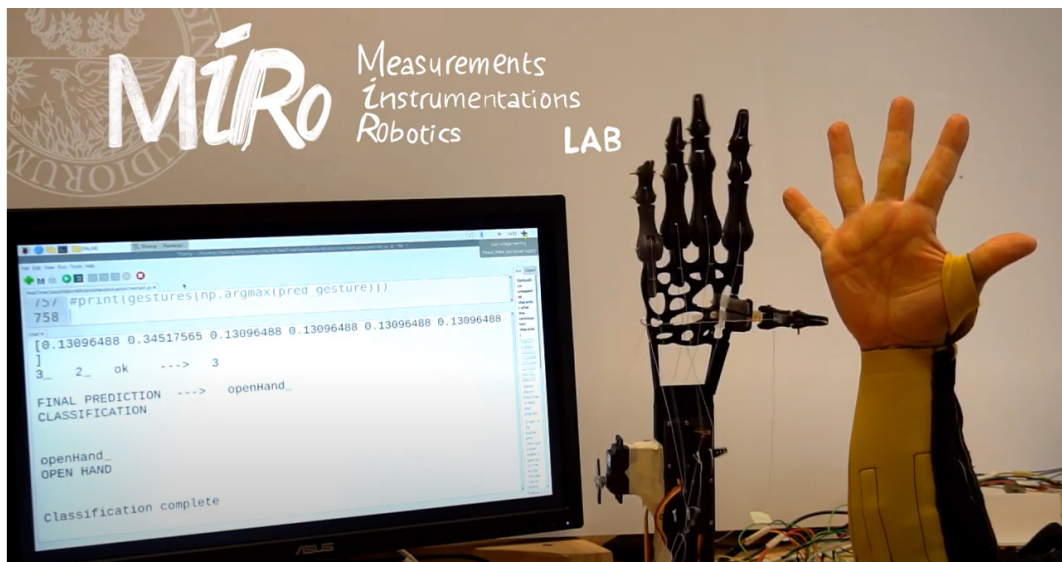


Fig. 3.11 Acquisition and classification of EMG signals in real-time for the control of a robotic limb

Chapter 4

Visual feedback through augmented virtuality and augmented reality

4.1 Mixed reality in the rehabilitative environment

Mixed reality assistance is a widespread technological innovation in the rehabilitation environment [128]. In fact, in recent times it has been possible to observe and evaluate multiple benefits that can be introduced through mixed reality and serious gaming [85], such as the possibility of introducing cutting-edge methods that improve the perception abilities of patients and therapists [105], the possibility of finding and visualising clinical data quickly and intuitively [18], and the partial transformation of boring and repetitive exercises into playful motor activities [56], [113] [7].

For example, as developed for the Ausilia project and reported in [128] and [85], the use of an adequately sensorised domotic apartment makes it possible to exploit augmented virtuality technologies to study patients during their daily activities. The domotic environment acquires information via wearable sensors, instrumented tools, and RGB-D cameras. Data are then passed to the local server via cable or Wifi using standard communication protocols such as MQTT or Ros networks. The virtual environment then retrieves the data saved on the server and presents it in an augmented virtual environment. With the support of HTC-VIVE technology, an immersive 3D animation is created. Such technology can recreate the recorded real environment by displaying data in the form of interactive graphs, reproducing the heartbeat sound as background audio, and by synchronizing all the information with the subjects' point clouds.

This allows the therapist to conduct an offline analysis and observe the patient's natural movements and behaviour without influencing him/her with his/her physical presence. In

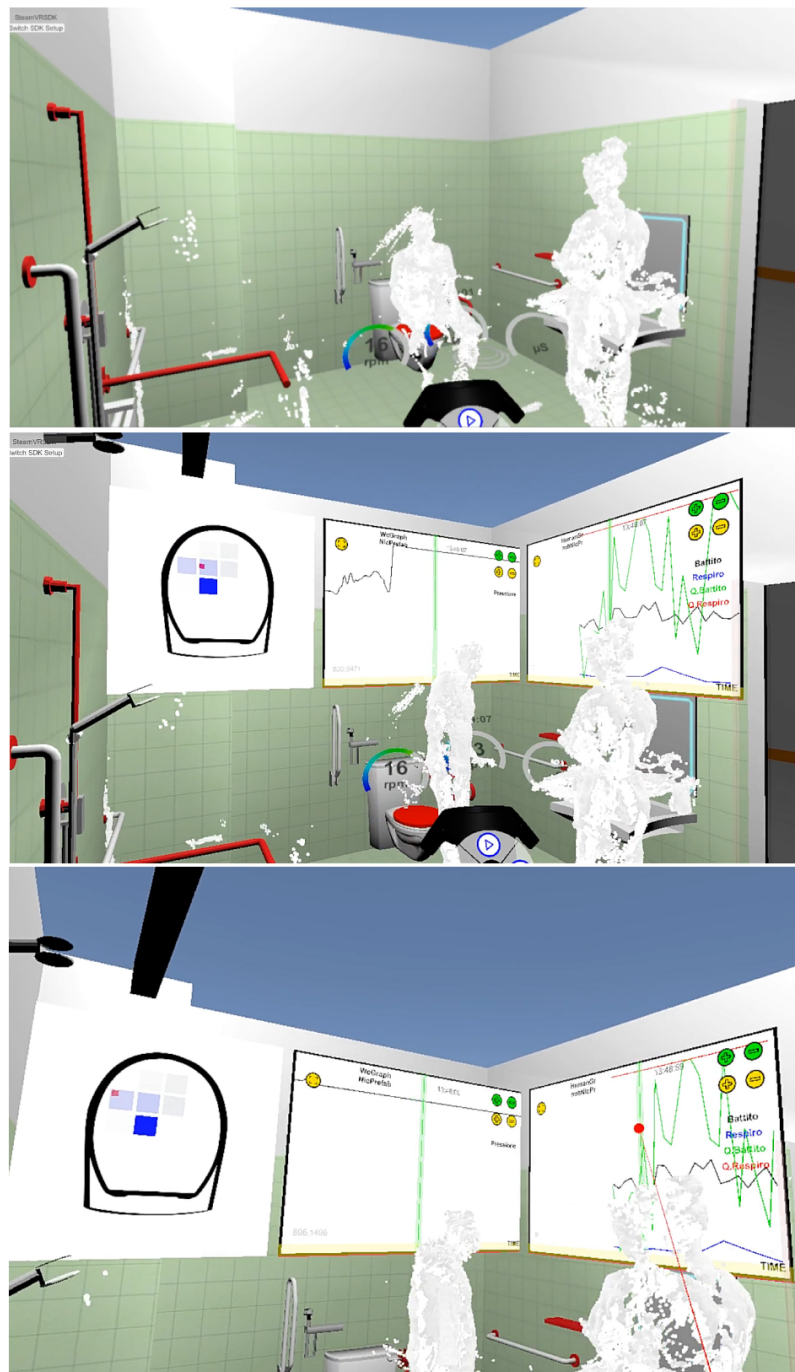


Fig. 4.1 3D Augmented Virtuality reproduction with HTC-VIVE of Ausilia domotic apartment with point clouds of the patient and the therapists. In the background on the left side are the graphs of the toilet pressure sensors and on the right side are the graphs of the patient's heartbeat and breathing.

addition, the opportunity to interact with the graphs (zooming, rewinding forward and backwards the recordings, the scenes and the animations) and the surroundings by moving around in the 3D environment allow the therapist to assess the scene from multiple perspectives while consulting objective parameters, such as heart rate, pressure on the toilet seat, and breathing. Finally, as the data is consulted offline, it is possible to scroll through the animation quickly, pausing to observe only the salient moments for the current analysis. In this way, the work time is considerably reduced and it compresses what would previously have been a work costing several days into a few hours of analysis.

In addition to remote patient observation, mixed reality also offers significant advantages in the field of rehabilitation. Together with robotic technologies and the development of wearable sensors, mixed reality makes it possible to reintroduce long and tedious rehabilitation exercises in the form of interactive and playful games, i.e. serious games [1]. Serious games seem to be similar to video games and may often be confused with them. Although they share the recreation of a playful environment, serious games are developed and designed with the help of physiotherapists and doctors based on the specific needs of the patients[38]. Serious games allow the patient to transform the repetitive movements of the rehabilitation into commands to interact with the game interface. This shifts the subject's attention and focuses it on achieving the game's goals. This ludic-motor component allows a new approach to rehabilitation activities, involving the subject more, longer and facilitating the repetition of movements [90], [1]. Thanks to mixed reality technologies, serious games, as well as video games, are nowadays abandoning the limits of the 2D screen and presenting increasingly immersive and interactive 3D environments to involve the user more deeply. Moreover, the need of introducing wearable sensors and cutting-edge technologies to convert the patient's movements into play commands has enabled quantitative data on the patient's exercises to be collected. Thus, by analysing the evolution of the patient's movements during rehabilitation, it is possible to assess the improvements and the performances of the therapy compared to other therapies and approaches.

Finally, this technology can be developed for telerehabilitation needs[42],[42],[124]. As the Covid-19 health emergency emphasised, many people often had to stop treatment due to limited mobility for the pandemic containment or personal limitations. Telerehabilitation offers a few good alternatives that would allow people to be monitored remotely from their homes for simple rehabilitation exercises. In this case, mixed reality would augment the therapist's clinical eye by allowing the patient to be observed in real-time and providing biomechanical data to conduct a careful and informed assessment. Similarly, the same system installed in a domotic environment can be used to support the daily activities of elderly people living alone. This would provide security and more help in case of need. Devices such as a

fall detector citede2017home [33] [79] can recognise a falling person and initiate procedures to call for help. Other applications, such as the one proposed by D'Agostini et al. [27], add simple visual support for mildly cognitive people to carry out everyday activities. Today, most of these applications are only at the prototype stage and involve many research groups.

4.2 Mixed reality for exoskeleton walking

All the robotic exoskeletons require practice and training for proper use. Wearing a robotic exoskeleton and using it without practice carries risks both for the subject and for the success of the exotherapy (or experimental trial tests). Being able to approach the correct movement in advance without the need to wear an exoskeleton makes it possible to divide the learning phases into steps, reduces the complexity of each step and makes it easier for the subject to use it. In addition, for research purposes with healthy subjects, the possibility of performing tests without an exoskeleton, realistically simulating the movements, reduces the risk of compromising the tests and considerably streamlines the preparation procedure for the biomechanical evaluation. As a result, this simplifies the set-up of the subject by also solving some of the issues related to the marketisation of motion capture systems and reduces the testing time. Being able to exploit mixed reality technologies to teach the use of an exoskeleton allows intuitive visual comparisons to be pursued.

For these reasons, we propose the study and development of an augmented reality system both for supporting and learning to walk with an exoskeleton and also to return visual feedback on a few simple human movements. For this purpose, we developed an augmented reality environment to allow the assessment of the subject's learning without the need to wear the robotic exoskeleton. The tests compared this new learning method with the traditional approach and focused on walking on a treadmill in situ, taking as reference the gait recording data of an experienced subject who is used to walking with robotic exoskeletons. To evaluate the final performance of each subject the pose is recorded and an off-line comparison follows. The proposed analysis is based on the comparison of key factors typically used in gait analysis and widely reported in the literature. Our paper reports the set-up and the material used to conduct the tests, the structure of the tests and the calculation of the key factors. The results of the comparisons with the traditional learning method and the conclusions drawn from the work are then presented.

4.3 Set up and material

The system is structured by three basic components:

1. A multicamera Azure Kinect markerless system for the real-time 3D pose acquisition of the subject with the shortest possible delay.
2. An MQTT communication protocol to transmit the 3D coordinates of the monitored subject.
3. Microsoft Hololens, an augmented reality reproduction device to allow the subject to visualise his pose and compare it with a guiding Avatar.

Along with the other supporting devices, there are :

1. A treadmill is used to simplify the analysis by walking on-site.
2. A multicamera Vicon Nexus system, an offline marker-based acquisition system that allows high-frequency and high-accuracy acquisition of the subject's pose, is used for the validation in post-processing.
3. A laptop device is used to control the data flow and perform post-processing.
4. An exoskeleton H3 for the ghost avatar movement recording



Fig. 4.2 Set up consists of six Vicon cameras, three Azure Kinect cameras, a treadmill and an AsusTUF Gaming Fx504 series laptop

4.4 Virtual Environment Set Up

The augmented reality environment is structured in scenes and is entirely rendered by Microsoft HoloLens 2. It is important to stress that only the user wearing the HoloLens can see the generated animations. As shown in Figure 4.3, the first scene presents a simple introductory screen with a brief description of the test and allows the user to start by pressing a button.

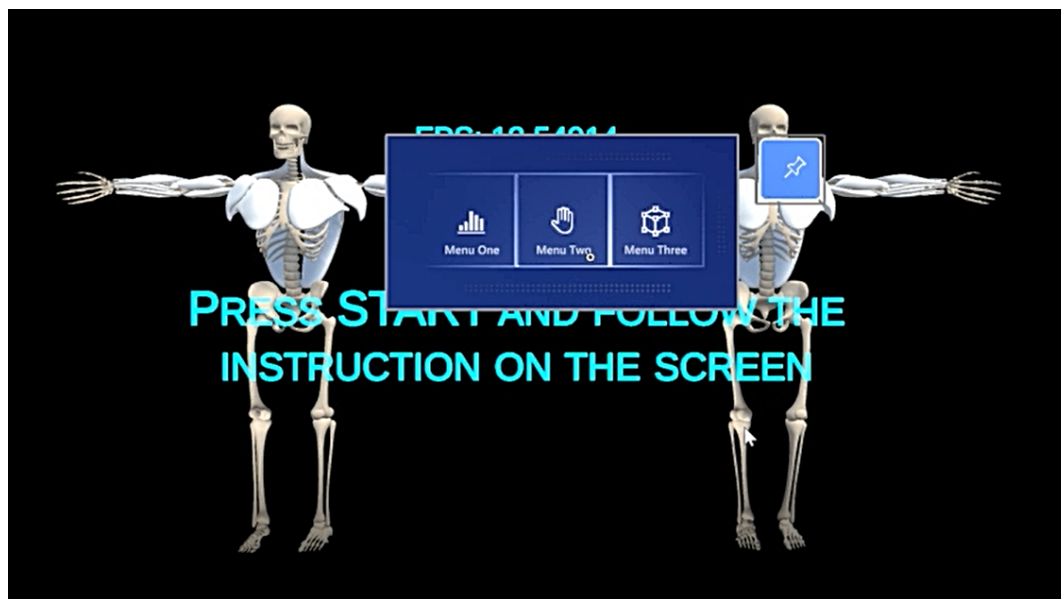


Fig. 4.3 Introduction Scene

Once the button is pressed, the test scene displays two avatars placed side by side a few meters in front of the user, as shown in Figure 4.4. On the left, the subject avatar reproduces the movements of the user in real-time, while the ghost avatar, on the right, reproduces an old recording of an experienced user, who used to walk with an exoskeleton.

4.4.1 Avatars Animation and motion control

Both avatars are generated with Blender 3.0¹ from open source models available online². The present musculature was instead drawn by hand starting from different orthogonal perspectives of the relative 2D anatomical representations of each muscle. The virtual environment was generated in Unity and all the avatar's movements are controlled using

¹<https://www.blender.org/> last access: 02/02/2022

²<https://www.turbosquid.com/Search/3D-Models/free/skeleton> last access: 02/02/2022

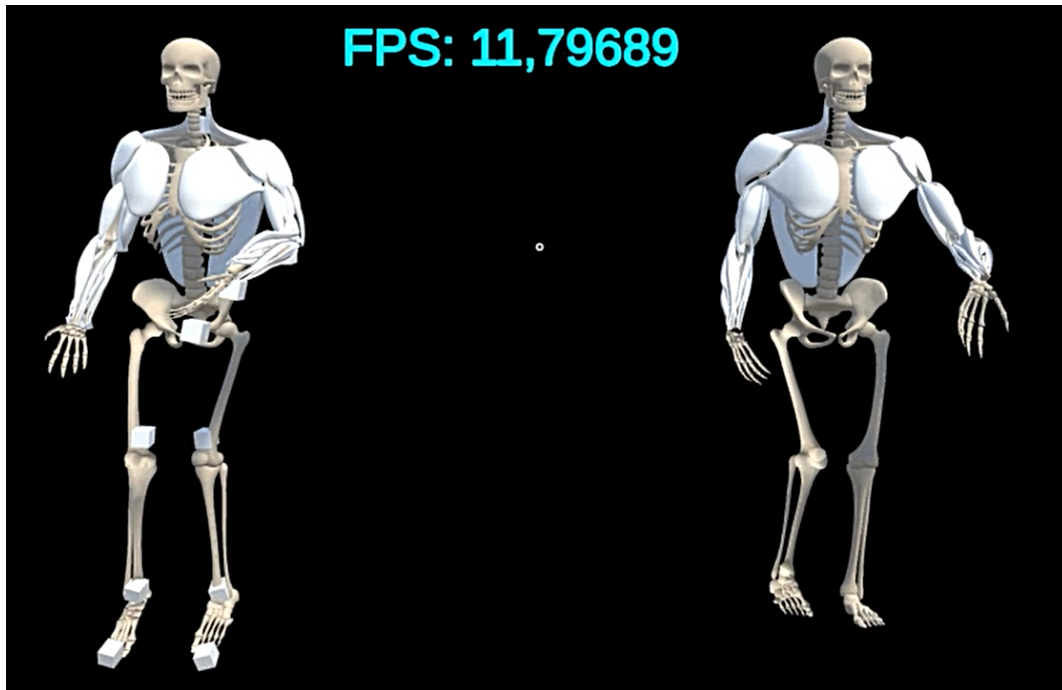


Fig. 4.4 Test scene with the two avatars. On the left side, there is the subject avatar, that reproduces the user's movements in real-time. On the other side, there is the ghost avatar, that reproduces the recorded movements of an expert pilot of robotic exoskeletons

animation rigging tools and inverse kinematics chains³ [14]. The project can be compiled for different platforms. By introducing the Mixed Reality Toolkit (MRTK) module it is possible to compile the project for Microsoft HoloLens 2 enabling the interaction with augmented reality tools, such as the buttons to start/stop events or the possibility to deal with the canvas, which allows the movement and resizing of the virtual objects in the 3D space. An additional module to manage the Message Queue Telemetry Transport (MQTT) asynchronous communication protocol allows reading of the data sent by the acquisition device. The acquisition device is an Asus Tuf Gaming Fx504 series laptop computer with 8 Gb of dedicated graphic memory and an 8th generation i7 core. It is connected with three Azure Kinects via a USB 3.0 connection cable. The Azure Kinects are arranged as shown in the Figure and capture the subject walking on the treadmill. The acquisition device processes the data coming from the cameras with the Azure Kinect Body tracker, a convolutional neural network already presented in the first chapter. Azure Kinect body tracker returns the 3D coordinates of each joint of the monitored subject with respect to each camera reference system. As presented in the first chapter, a preliminary calibration of each of the cameras

³<https://docs.unity3d.com/Packages/com.unity.animation.rigging@0.2/manual/index.html> last access: 02/02/2022

allows the relative rototranslation matrices to be found. By multiplying each joint by the rototranslation matrix of the corresponding camera, it is possible to express all the skeletons in the same reference system. Similarly, as described in the first chapter, Vicon is calibrated to share the same reference system.

4.4.2 The subject avatar and the ghost avatar

The subject avatar is a virtual character that reproduces the user's movements in real time and allows immediate comparison with the ghost avatar next to it. The subject avatar uses the skeleton acquired by the Kinect multi-camera system to reconstruct the subject's pose and move accordingly. After expressing all the skeletons into the same reference system using matrix rototranslation operations, the coordinates of the pelvis are subtracted, allowing to lock the avatar pelvis in a precise point of the space. This allows a more effective assessment of movement by focusing attention on the limbs and neglecting the overall translational movement of the subject.

The ghost avatar acts as a guide for the user. The ghost avatar is animated similarly to the subject avatar, but, instead of real-time acquisitions, it uses a previously acquired recording. In the case of exercises with exoskeletons, a recording of an experienced user who knows how to reproduce the movements correctly and without errors is used. In the case of telerehabilitation exercises, the ghost avatar can be animated by recording the guiding movements made by a therapist or by an interactive real-time acquisition.

4.5 Acquisition and processing at 30 fps data display

The acquisition system is structured in the same way as described in Chapter 2. The system is composed of several Azure Kinect cameras connected to a single laptop. Similarly, the calibration phase is carried out using a white cross placed in the origin of the common reference system. Unlike the acquisition system described in chapter 2, the system works in real-time and the data acquired by the cameras are immediately multiplied by the rototranslation matrices calculated during the prior calibration phase. Each camera then returns a skeleton as an array of x,y, and z coordinates for each joint expressed with respect to the common reference system. A quick procedure encodes the float array into a string message that is then immediately sent via the MQTT protocol. All sent skeletons are published at the MQTT broker. The MQTT broker is Eclipse Mosquitto. Eclipse Mosquitto is an open-source program that allows the user to use its laptop device as a broker. In the MQTT protocol, the broker handles the publishing and reading of messages. All devices registered as "publishers"

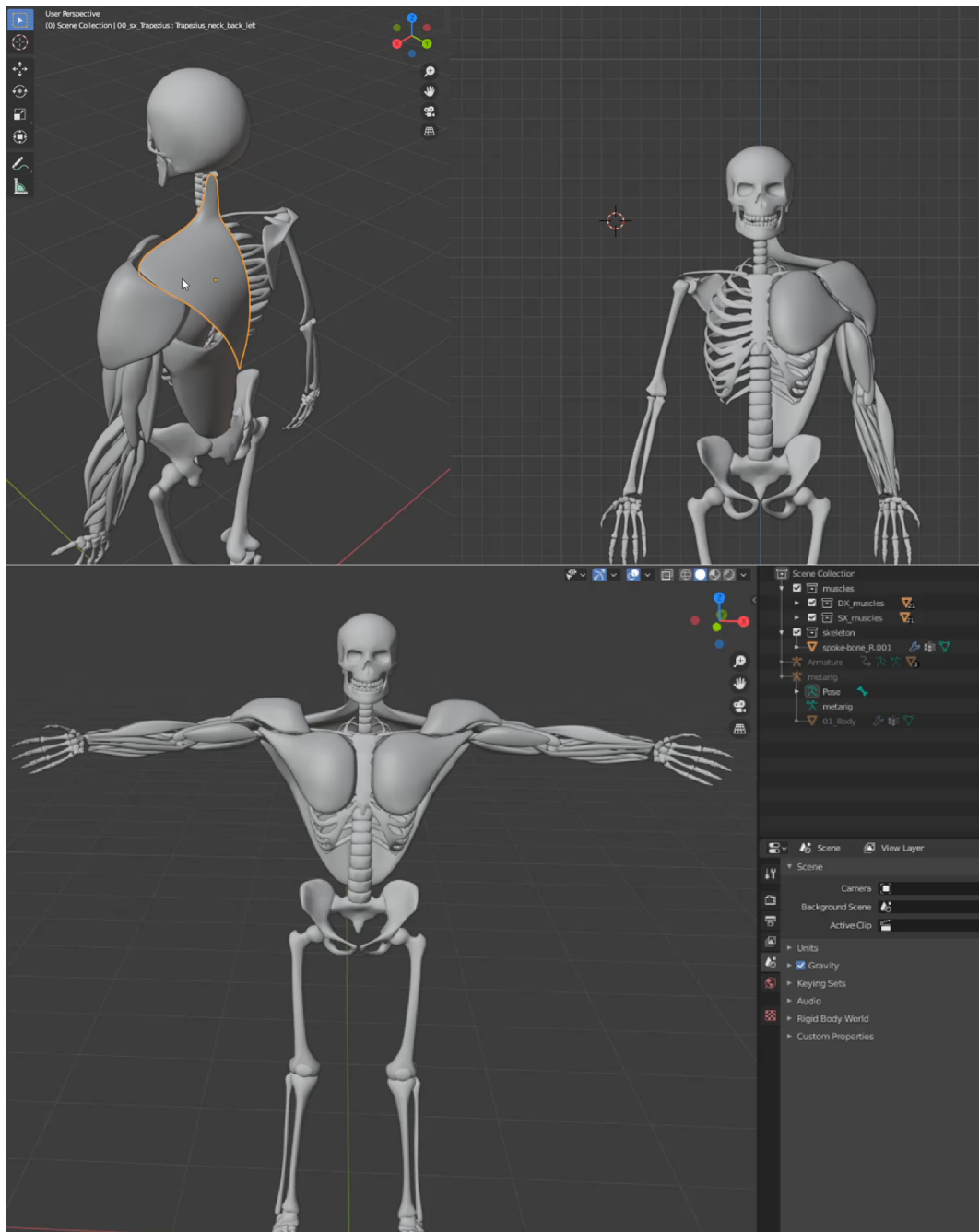


Fig. 4.5 Generation of the avatar's musculature in the blender environment

can connect to the broker and publish messages in a particular channel. All "subscribed" devices can read the published messages if they are subscribed to the channel. In this way the virtual environment created in Unity and compiled for Hololens 2 allows the wearable device to receive and read the messages. Once the capture device publishes the encoded message immediately the application running on Hololens reads and decodes it. Once the

message has been decoded into the array format of the original x,y, and z coordinates of the joints, the information is used to move the avatar. As errors in the Kinect markerless system can occasionally affect pose estimation, additional data processing is included to fluidify the final animation and avoid annoying glitching in avatar movement. In the presented case, a simple moving median is introduced on the individual coordinate triplets provided by the three cameras. Unlike the moving average, the moving median makes it possible to recognize and exclude the measurements that deviate most from the distribution, thus isolating outliers and avoiding most of the glitches. However, better results can be obtained by introducing a more elaborate information fusion, such as with a Kalman filter or by increasing the number of cameras.



Fig. 4.6 User's view during walking trials wearing Microsoft HoloLens 2

4.6 Test and trials

The data collection is structured in different walking trials with five different subjects. Each subject is allowed to observe how the avatar performs the movements while walking with the exoskeleton. Then, each subject performs different walks of two minutes each in two different conditions. In the first condition, the subject has to emulate walking with the exoskeleton without the aid of the exoskeleton and without the HoloLens (thus without the help of the Augmented Avatar). In the second condition, the subject is allowed to walk with the support of HoloLens guiding him/her through visual feedback between the subject avatar

and the ghost avatar. During each trial, data on the subject's pose is saved and then analyzed offline in post-processing for comparison between walks.

4.7 Postprocessing and skeleton comparison

The final objective of the tests is to determine whether the augmented reality guide allows the subject to emulate the walk proposed by the ghost avatar. To this end, the different walks are compared by analyzing the Spatio-temporal parameters of the gait.

The data analysis is structured as follows:

1. First, gait factors suitable for the analysis of lower limb movement and statistical comparison are selected from the literature[69].
2. Each distribution is checked using a normality check to see if the two main conditions for conducting the two-sample t-test are satisfied: gaussianity and heteroscedasticity [26].
3. The two-sample t-tests are conducted by taking the distributions two by two. First by comparing the normal walk, then the walk supported by the AR tools, with that of the expert user.

In particular, the comparison is expected to show that the expert user's walk is similar to that achieved with the aid of AR tools, while it differs from the walk without AR tools according to the skill of the test subject.

4.7.1 The Spatial-temporal gait parameters

The comparison between the different walks is performed by selecting and comparing the distributions of some Spatial-temporal gait parameters reported in the literature, such as described in [69]. The Spatial-temporal gait parameters chosen are particularly suitable for the characterization of the movement of the lower limbs and the two-sample t-tests statistical analysis reported later[54].

All parameters are calculated by rescaling all joint coordinates with respect to the subject's height, defined as the distance between the top of the head and the ground.

The Spatial-temporal gait parameters chosen are as follows:

1. The cadence is defined as the number of steps the subject takes per unit of time
2. The instantaneous speed, calculated as the distance travelled by the subject per unit of time

3. The Step length is defined as the distance by which each foot is in front of the other
4. The stride length is defined as the length by which each foot moves forward in a gait cycle
5. The maximum knee angle is defined as the maximum angle drawn by the segments linking the head of the femur, the knee, and the ankle
6. The minimum knee angle, defined as the minimum angle drawn by the segments linking the head of the femur, the knee, and the ankle



Fig. 4.7 Set up acquisition with the test subject wearing HoloLens 2 to emulate the movement of the expert user in exoskeleton walking. The colours highlight the height of the subject and some spatiotemporal parameters used for gait assessment.

4.7.2 The Normality Checks

The normality tests are used to determine whether the main characteristics required by the statistical test are fulfilled. In particular, they try to establish if the quantity and type of samples acquired present a Gaussian bell-shaped pattern and heteroscedasticity.

The tests mainly conducted are of two types, graphical/illustrative and statistical. The graphical illustrative tests include the QQ-plot and the frequency distribution or histograms. As shown in Figure 4.9 and Figure 4.10, these tests are mainly qualitative and allow visual

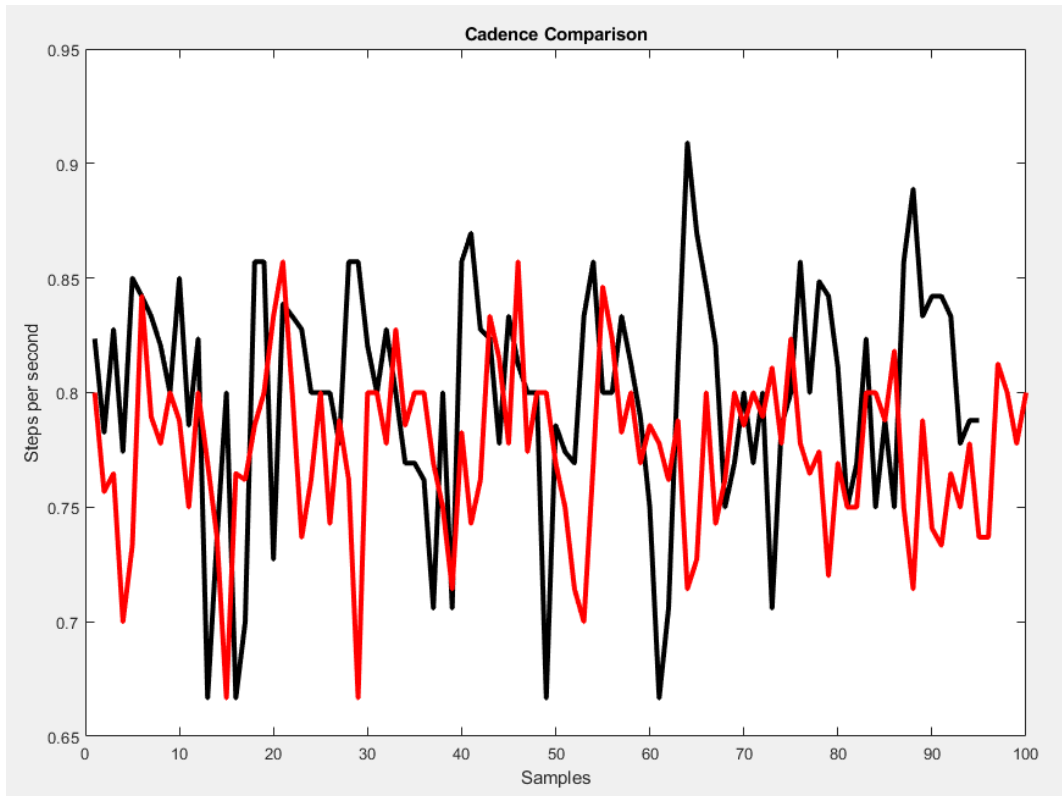


Fig. 4.8 In red, the cadence of the expert user, and in black the cadence of the test subject using Hololens

comparison between the analyzed distribution and a Gaussian distribution with the same mean and variance.

D’agostino-Pearson omnibus, Shapiro-Wilk test or Kolmogorov-Smirnov test The statistical normality tests, such as the D’agostino-Pearson omnibus, Shapiro-Wilk test or Kolmogorov-Smirnov test, compare the distribution under investigation with a normal distribution using a statistical test.

All the normality tests and the two-sample t-tests were performed with the related tools and statistical functions developed in the Matlab R2019b environment.

4.7.3 The two-sample T-tests

As reported in the literature [54] and the MathWorks site ⁴, the two-sample t-test is a statistical analysis comparing the mean of two Gaussian distributions, as shown in Figure4.11. The test has a null hypothesis, H_0 , for which the means of the two distributions are statistically equal (by default at 95%). The function developed in Matlab requests as input the array of the two

⁴<https://es.mathworks.com/help/stats/ttest2.html>

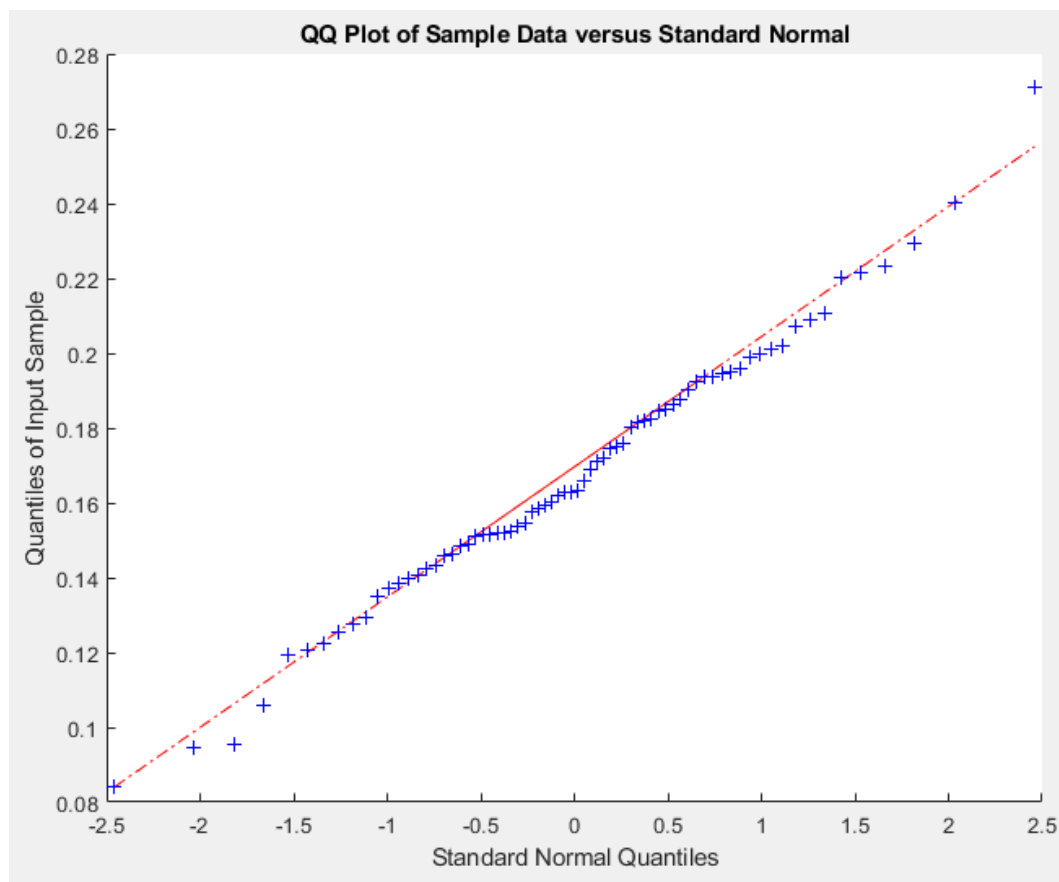


Fig. 4.9 Q-Q plot of the distribution obtained by calculating the maximum knee angle reached during each gait cycle

distributions to be compared and returns the binary result of the test, h , and the p-value, p . The binary result h results in 0 if the test accepts the null hypothesis H_0 , 1 if it rejects H_0 .

4.8 Results

The test reported significant results for most of the selected gait factors. In many cases, it was possible to notice in Figure 4.12, that significant differences between the distributions have been obtained with and without the HoloLens support compared to the gait of the experienced user. In particular, the most significant Spatio-temporal factors were the minimum and maximum angle drawn by the knee during the walk, and the step length scaled to the subject's height. As it is possible to see in the Figure 4.12, the subject without virtual guidance shows a difference between the averages of the distributions of about 20 degrees. The same subject walking with the help of HoloLens, on the other hand, manages to reduce the difference between the averages to a few degrees, reproducing the target movement more

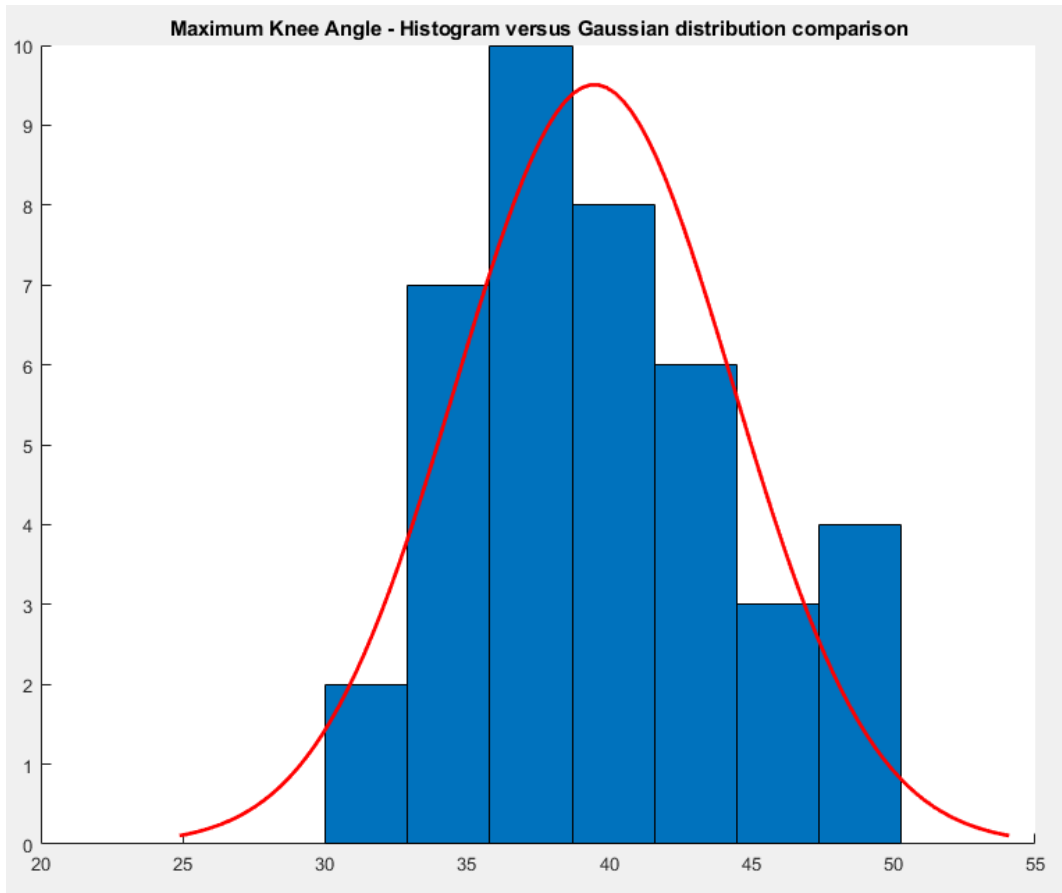


Fig. 4.10 Maximum Knee Angle - Histogram versus Gaussian distribution comparison

faithfully. Other factors, such as the instantaneous speed of the subject, were somewhat less significant since all subjects performed all tests on the treadmill at the same speed and with a constant gait.

As shown in Figure 4.13, the cadence of the subject also gave good results and it can be seen that the subject wearing HoloLens was able to reproduce target patterns by more accurately reproducing the walk of the experienced user. Compared to walking without a reference in the AR, it was noted that the subject walked with a cadence proportional to his height. In this case, the cadences were statistically equal only if the test user had a similar height to the experienced user. On the contrary, if the subject had to follow the reference of the ghost avatar, the subject easily synchronized the cadence by accelerating or decelerating the movements and extending or reducing the stride length.

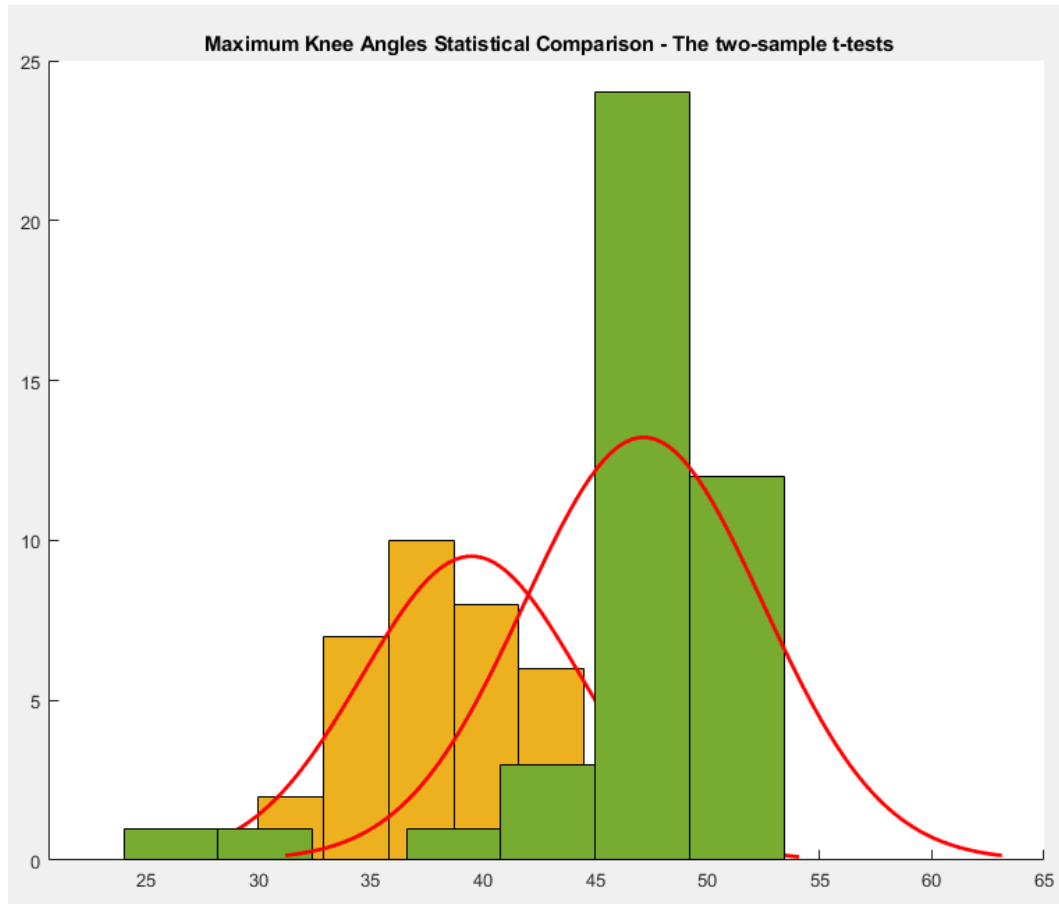


Fig. 4.11 Maximum Knee Angles Statistical Comparison using the two-sample t-tests - On the right, the distribution of the spatiotemporal parameter of an experienced user is shown in green. On the left, the distribution of the subject without AR support is reported in yellow.

4.9 Conclusions

The work proposed in this chapter illustrates the development and validation of a Mixed reality tool to support rehabilitation exercises and walking with an exoskeleton. This chapter describes how the system was developed using open-source libraries and programs. Afterwards, a brief validation on the effectiveness of such a tool in helping inexperienced users of robotic exoskeletons to walk similarly to an experienced user without the need to wear the robotic device is reported. The acquisition system was developed with a markerless multicamera system, composed of three Azure Kinect cameras communicating at 30 fps with an MQTT communication protocol.

The validation aims at comparing the walk of an expert user with different test subjects, who perform the gait with and without the support of the AR tools. The analysis carried out using standard statistical tests compares the distribution of some standard spatiotemporal

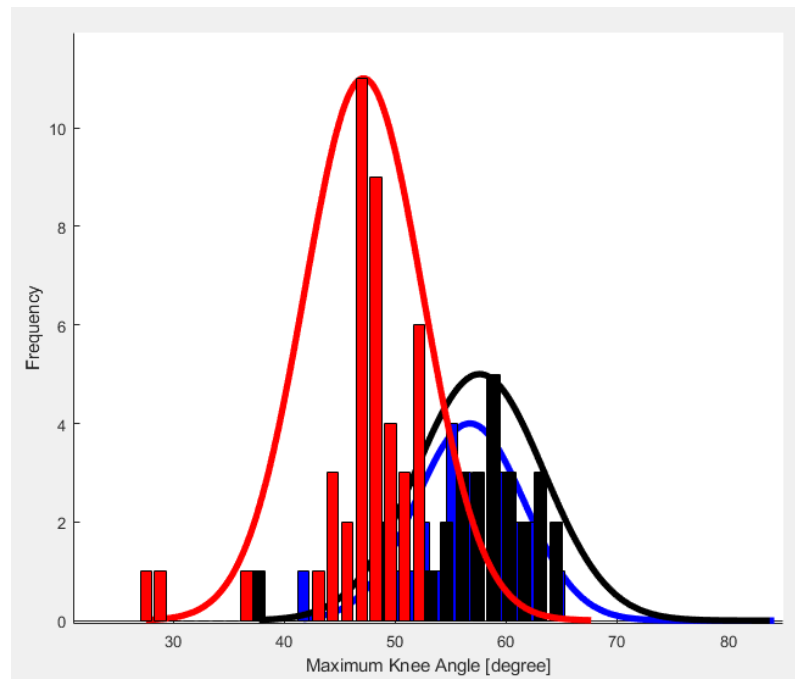


Fig. 4.12 Maximum Knee Angles graph Comparison between the expert user (in black), the subject supported with (in blue) and without (in red) AR tools

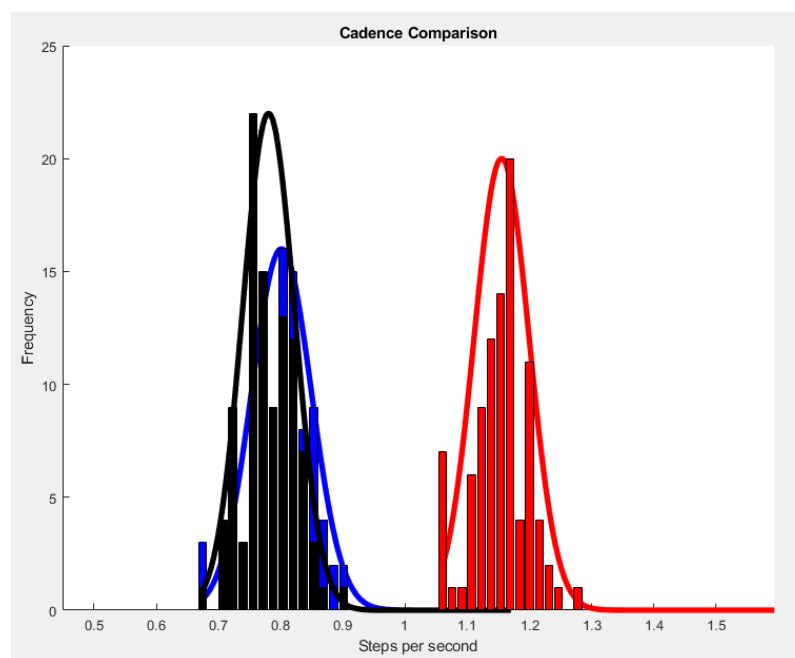


Fig. 4.13 Comparison of the cadence distributions of the expert user (in black) with the cadence distribution of the subject with Hololens (in blue) and the subject without Hololens (in red)

parameters reported in the literature. Overall, the reported results are promising and the statistical tests on the spatiotemporal parameters highlighted the improvements obtained through the use of AR tools. However, the validation was limited to the analysis and comparison of only five subjects. Moreover, the multi-camera acquisition system composed of only three Azure Kinects does not allow a sufficiently high accuracy to provide high-quality avatar movement and to calculate additional spatiotemporal parameters. In this case, it is necessary to rely on the Vicon Nexus acquisition and pursue further comparisons. Therefore it would be interesting to repeat the experiment by comparing the results on more subjects and with a larger number of acquiring cameras. Finally, using a larger number of subjects, it would be interesting to study how learning through the use of the developed AR tools evolves, compared to a sample of subjects who learn the use of the exoskeleton through simple use.

Conclusions

The manuscript reports the research, development, and validation of engineered and automated measurement systems for innovative applications in the field of rehabilitative robotics and biomechanical gait assessment. The work is divided into four projects, respectively reported in four chapters. Each chapter addresses some critical issues described in the state of the art by proposing and validating cutting-edge solutions, tools, and procedures.

In particular, the first chapter deals with the problems related to human pose estimation and kinematic skeleton reconstruction. Focusing on the most challenging aspects of preparation procedures and the limitations related to the current marker-based gold standard motion capture systems, the chapter proposes the development, and the study of a breakthrough markerless solution. The study compares a gold standard Vicon Nexus marker-based system with a custom-made multi-camera system, based on the acquisition and elaboration of three Azure Kinects data. First, the comparison shows the strengths of the markerless system in terms of cost and time effectiveness during the subject preparation and for the subsequent post-processing data treatment. Second, the work compares the accuracy achieved by the two pose acquisition systems. The comparison conducts a spatiotemporal synchronized joint-to-joint data analysis to evaluate the absolute euclidian difference between the measurement performed by the two systems. From this analysis, it can be observed that the accuracy increases remarkably with the number of cameras and that the absolute average error drops down from 15 cm to 5 cm by moving from a single-camera system to a multi-camera system. Furthermore, it could be observed that the measurements result more stable by introducing a Kalman sensor fusion filter than by calculating a simple average or median between the acquired poses. A final consideration lays the foundations for a more in-depth study of performances by using a markerless system with a larger number of cameras and a more advanced model for the Kalman filter.

The second chapter deals with the requirements of preliminary anatomical measurements. In particular, the need to evaluate volumes, masses, and inertias of anatomical segments in a reliable, affordable, rapid, and automatic way is addressed. After illustrating the main state-of-the-art solutions and approaches, an automated camera-based measurement system

is proposed supported by the novel algorithm for the volume computation is presented. The measurement system consists of a hospital bed equipped to acquire RGB-D images by using an Azure Kinect camera. The camera is installed facing downwards and acquires a 3D scan of the subject's body. Both RGB and depth images are then processed by a local computer which uses artificial intelligence techniques to identify and segment the subject's point cloud. The segmentation isolates the anatomical components into independent point clouds. At this point, the chapter introduces the custom-made algorithm that estimates the volumes and, subsequently, the masses and the inertias. In particular, the proposed algorithm extends the classical Monte Carlo method for computation of the areas of irregular surfaces to estimate the volume enclosed in the acquired surface point clouds without the burden of interpolation or mesh calculation. A study of the performance of the algorithm is then reported, testing the computational efficiency and measurement accuracy on objects of different shapes, sizes, and complexity. The results show that the algorithm can calculate the required volumes in reasonable times and with an average error of less than 7% even for complex and folded surfaces. It is also shown that the output of the algorithm returns a cloud filled with homogeneously distributed points. This cloud is particularly suitable for the calculation of the mass and the inertia assuming that the object in question has a zero density gradient across the whole domain. Having derived the density of the anatomical segment from state-of-the-art anthropometric tables, the chapter concludes the description of the measurement system by reporting a traditional calculation of mass and inertia with a discrete approximation of the standard integral formulas.

The third chapter focuses on the research and development of a real-time EMG acquisition system to meet the needs of usability and on-site data consultation, without the need to wait for post-processing operations. Emphasising the role and studies conducted on EMG signal acquisition in the rehabilitation and biomechanical fields, the chapter reports on the development and validation of a wearable device. The device consists of an eight-channel electromyographic sensor integrated into a wearable garment and complemented by instant data processing for information visualization and muscle pattern recognition. Afterwards, an extensive discussion and comparison of the machine learning algorithms taken into account are carried out, with a special focus on recursive neural networks. Finally, a particular LSTM (Long-short Term Memory neural network) architecture is proposed. As can be observed from the results, the LSTM neural networks are well suited for the recognition and classification of static hand gestures by collecting muscle signals of the forearm. The proposed research places special care on showing the robustness performance of the system by augmenting the training dataset, structuring the classification algorithm as an enable of specialized sub-networks and rising the accuracy performance from an average of 70% to

85% on newly acquired samples over different tests and pursued on different days. Finally, it is shown how this device can be applied in the fields of rehabilitation/collaborative robotics, control of intelligent prostheses and data analysis through pattern recognition. For this purpose, a simple application of robotic prosthesis control, developed by a few master course students, is chosen to visualize the entire pipeline and to show the effectiveness of the algorithm for action recognition and information feedback.

The last chapter proposes the development and validation of a mixed reality application for the guidance of robotic exoskeletons. The application exploits open-source programs, such as Unity and Blender, for the development of interactive virtual characters and environments. The characters thus created are animated and presented in augmented reality to the user wearing Microsoft HoloLens 2. In particular, the user learns and emulates the correct walk with an exoskeleton by comparing his movement with that of an expert user. This feedback is returned to the user through a visual comparison between the two avatars that respectively reproduce the real-time movements of the test subject and the recorded movements of the expert user. A multi-camera system consisting of three Azure Kinects, similarly to what was proposed in the first chapter, allows offline and real-time acquisition and rendering of the human pose by sharing data with an MQTT protocol. Finally, a simple study with a small number of subjects is proposed to give an initial quantitative assessment of the potential of the application. The study simply introduces a statistical comparison between the spatiotemporal parameters of the gait cycle, such as the stride, the cadence, or the maximum and minimum angle drawn by the knee, to evaluate how much the visual feedback in augmented reality can help the test subject. The results obtained are promising, and already with a small number of subjects, it is possible to observe how the subjects assisted by the augmented reality application can reproduce the walk of the expert user better and faster with respect to the unassisted subjects. The study, presented here in a preliminary phase, is still being developed and deepened with a larger number of subjects, a larger number of spatiotemporal parameters compared, and a better pose capture system.

Collectively, the arguments presented in the four chapters of this work apply emerging technologies to the practices and tools used every day by therapists and researchers in the field of rehabilitation and health care. The topics covered outline a common thread that concerns the entire line of measurement and processing of the information involved, proposing changes to the sensors responsible for data acquisition, the data processing methods and finally their display mode. Comparing the numerous aspects that characterize these tools, it is clear that the proposed solutions aim to simplify the work of human operators by improving the accuracy, automation and repeatability of the acquisition and processing of information and

providing intuitive feedback with more complete and immersive visualization of the collected data.

References

- [1] Abdelkader, G., Hocine, N., et al. (2011). Mixed reality serious games for post-stroke rehabilitation. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 530–537. IEEE.
- [2] Ahmad, Z., Illanko, K., Khan, N., and Androustos, D. (2019). Human action recognition using convolutional neural network and depth sensor data. In *Proceedings of the 2019 International Conference on Information Technology and Computer Communications*, pages 1–5.
- [3] Akhmadeev, K., Rampone, E., Yu, T., Aoustin, Y., and Le Carpentier, E. (2017). A testing system for a real-time gesture classification using surface emg. *IFAC-PapersOnLine*, 50(1):11498–11503.
- [4] Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., and Arnrich, B. (2020). Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors*, 20(18):5104.
- [5] Andújar, D., Dorado, J., Fernández-Quintanilla, C., and Ribeiro, A. (2016). An approach to the use of depth cameras for weed volume estimation. *Sensors*, 16(7):972.
- [6] Antifakos, S. and Schiele, B. (2002). Bridging the gap between virtual and physical games using wearable sensors. In *Proceedings. Sixth International Symposium on Wearable Computers*,, pages 139–140. IEEE.
- [7] Athanasiou, A., Antoniou, P., Pandria, N., Astaras, A., Nizamis, K., Mitsopoulos, K., Praftsiotis, A., Arvanitidis, T., Apostolou, T., Magras, I., et al. (2021). Affect detection in rehabilitation using wearable robotics, multiple biosensors and serious gaming: a concept using the neurosuitup platform.
- [8] Bangaru, S. S., Wang, C., Busam, S. A., and Aghazadeh, F. (2021). Ann-based automated scaffold builder activity recognition through wearable emg and imu sensors. *Automation in Construction*, 126:103653.
- [9] Benalcázar, M. E., Jaramillo, A. G., Zea, A., Páez, A., Andaluz, V. H., et al. (2017). Hand gesture recognition using machine learning and the myo armband. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1040–1044. IEEE.
- [10] Benatti, S., Farella, E., Gruppioni, E., and Benini, L. (2014). Analysis of robust implementation of an emg pattern recognition based control. In *BIOSIGNALS*, pages 45–54.

- [11] Bi, Y., Qi, L., Chen, S., Li, L., Liu, S., et al. (2013). Canopy volume measurement method based on point cloud data. *Science & Technology Review*, 31(27):31–36.
- [12] Bird, J. J., Kobylarz, J., Faria, D. R., Ekárt, A., and Ribeiro, E. P. (2020). Cross-domain mlp and cnn transfer learning for biological signal processing: Eeg and emg. *IEEE Access*, 8:54789–54801.
- [13] Bishop, G., Welch, G., et al. (2001). An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8(27599-23175):41.
- [14] Bouvier-Zappa, S., Dionne, O., and Hunt, D. (2019). Advanced use cases for animation rigging in unity. In *ACM SIGGRAPH 2019 Studio*, pages 1–2.
- [15] Brodin, E. (2016). *Anthropometric diversity and consideration of human capabilities*. Chalmers University of Technology.
- [16] Cao, T., Liu, D., Wang, Q., Bai, O., and Sun, J. (2020). Surface electromyography-based action recognition and manipulator control. *applied sciences*, 10(17):5823.
- [17] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2019). Openpose: realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1):172–186.
- [18] Cecco, M. D., Fornaser, A., Tomasin, P., Zanetti, M., Guandalini, G., Ianes, P., Pilla, F., Nollo, G., Valente, M., and Pisoni, T. (2017). Augmented reality to enhance the clinician’s observation during assessment of daily living activities. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 3–21. Springer.
- [19] Chang, W.-C., Wu, C.-H., Tsai, Y.-H., and Chiu, W.-Y. (2017). Object volume estimation based on 3d point cloud. In *2017 International Automatic Control Conference (CACCS)*, pages 1–5. IEEE.
- [20] Chrustek, P., Weżyk, P., Kolecka, N., Biskupič, M., Bühler, Y., and Christen, M. (2013). Using high resolution lidar data for snow avalanche hazard mapping. In *The Carpathians: Integrating Nature and Society Towards Sustainability*, pages 597–613. Springer.
- [21] Cifuentes, C., Braidot, A., Rodríguez, L., Frisoli, M., Santiago, A., and Frizzera, A. (2012). Development of a wearable zigbee sensor system for upper limb rehabilitation robotics. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 1989–1994. IEEE.
- [22] Collins, T. D., Ghoussayni, S. N., Ewins, D. J., and Kent, J. A. (2009). A six degrees-of-freedom marker set for gait analysis: repeatability and comparison with a modified helen hayes set. *Gait & posture*, 30(2):173–180.
- [23] Colyer, S. et al. (2018). A review of the evolution of vision-based motion a...
- [24] Covre, N., Luchetti, A., Lancini, M., Pasinetti, S., Bertolazzi, E., and De Cecco, M. (2022). Monte carlo-based 3d surface point cloud volume estimation by exploding local cubes faces. *ACTA IMEKO*, 11(2).

- [25] Covre, N., Nunnari, F., Fornaser, A., and De Cecco, M. (2019). Generation of action recognition training data through rotoscoping and augmentation of synthetic animations. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 23–42. Springer.
- [26] Cressie, N. and Whitford, H. (1986). How to use the two sample t-test. *Biometrical Journal*, 28(2):131–148.
- [27] D’Agostini, J., Bonetti, L., Salem, A., Passerini, L., Fiacco, G., Lavanda, P., Motti, E., Stocco, M., Gashay, K., Abebe, E., et al. (2018). An augmented reality virtual assistant to help mild cognitive impaired users in cooking a system able to recognize the user status and personalize the support. In *2018 Workshop on Metrology for Industry 4.0 and IoT*, pages 12–17. IEEE.
- [28] D’Angelo, L. T., Neuhaeuser, J., Zhao, Y., and Lueth, T. C. (2013). Simple-use—sensor set for wearable movement and interaction research. *IEEE Sensors Journal*, 14(4):1207–1215.
- [29] Dávila, P., Carrera, E., Jara, O., and Bassantes, H. (2019). Design of a low cost 3d scanner for taking anthropometric measurements. In *International Conference on Applied Human Factors and Ergonomics*, pages 971–978. Springer.
- [30] De Cecco, M., Fornaser, A., Tomasin, P., Zanetti, M., Guandalini, G., Ianes, P., Pilla, F., Nollo, G., Valente, M., and Pisoni, T. (2017). Augmented reality to enhance the clinician’s observation during assessment of daily living activities. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 3–21. Springer.
- [31] De Luca, C. J., Kuznetsov, M., Gilmore, L. D., and Roy, S. H. (2012). Inter-electrode spacing of surface emg sensors: reduction of crosstalk contamination during voluntary contractions. *Journal of biomechanics*, 45(3):555–561.
- [32] Dehais, J., Anthimopoulos, M., Shevchik, S., and Mougiakakou, S. (2016). Two-view 3d reconstruction for food volume estimation. *IEEE transactions on multimedia*, 19(5):1090–1099.
- [33] Dubey, R., Ni, B., and Moulin, P. (2012). A depth camera based fall recognition system for the elderly. In *International Conference Image Analysis and Recognition*, pages 106–113. Springer.
- [34] Dumas, R. and Wojtusch, J. (2018). Estimation of the body segment inertial parameters for the rigid body biomechanical models used in motion analysis. in: Müller b., wolf s.(eds) handbook of human motion.
- [35] Elamvazuthi, I., Duy, N., Ali, Z., Su, S., Khan, M. A., and Parasuraman, S. (2015). Electromyography (emg) based classification of neuromuscular disorders using multi-layer perceptron. *Procedia Computer Science*, 76:223–228.
- [36] Farina, D., Merletti, R., Indino, B., and Graven-Nielsen, T. (2004). Surface emg crosstalk evaluated from experimental recordings and simulated signals. *Methods of information in medicine*, 43(01):30–35.

- [37] Fornaser, A., De Cecco, M., Tomasin, P., Zanetti, M., Guandalini, G., Gasperini, B., Ianes, P., Pilla, F., and Ghensi, R. (2018). Augmented virtualized observation of hidden physical quantities in occupational therapy. In *2018 International Conference on Cyberworlds (CW)*, pages 423–426. IEEE.
- [38] González-González, C. S., Toledo-Delgado, P. A., Muñoz-Cruz, V., and Torres-Carrion, P. V. (2019). Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system. *Journal of biomedical informatics*, 97:103266.
- [39] Han, D., Liu, Q., and Fan, W. (2018). A new image classification method using cnn transfer learning and web data augmentation. *Expert Systems with Applications*, 95:43–56.
- [40] He, S., Yang, C., Wang, M., Cheng, L., and Hu, Z. (2017). Hand gesture recognition using myo armband. In *2017 Chinese Automation Congress (CAC)*, pages 4850–4855. IEEE.
- [41] He, Y., Fukuda, O., Bu, N., Okumura, H., and Yamaguchi, N. (2018). Surface emg pattern recognition using long short-term memory combined with multilayer perceptron. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5636–5639. IEEE.
- [42] Holden, M. K., Dyar, T. A., Schwamm, L., and Bizzi, E. (2005). Virtual-environment-based telerehabilitation in patients with stroke. *Presence: Teleoperators & Virtual Environments*, 14(2):214–233.
- [43] Huang, D. and Chen, B. (2019). Surface emg decoding for hand gestures based on spectrogram and cnn-lstm. In *2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI)*, pages 123–126. IEEE.
- [44] Hwang, I., Cha, G., and Oh, S. (2017). Multi-modal human action recognition using deep neural networks fusing image and inertial sensor data. In *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 278–283. IEEE.
- [45] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR.
- [46] Jochumsen, M., Waris, A., and Kamavuako, E. N. (2018). The effect of arm position on classification of hand gestures with intramuscular emg. *Biomedical Signal Processing and Control*, 43:1–8.
- [47] Kan, Y.-C. and Chen, C.-K. (2011). A wearable inertial sensor node for body motion analysis. *IEEE Sensors Journal*, 12(3):651–657.
- [48] Kanko, R., Laende, E., Selbie, S., and Deluzio, K. (2020). Inter-session repeatability of theia3d markerless motion capture gait kinematics. biorxiv 2020.06. 23.155358.
- [49] Kanko, R. M., Laende, E. K., Davis, E. M., Selbie, W. S., and Deluzio, K. J. (2021a). Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of biomechanics*, 127:110665.

- [50] Kanko, R. M., Laende, E. K., Strutzenberger, G., Brown, M., Selbie, W. S., DePaul, V., Scott, S. H., and Deluzio, K. J. (2021b). Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. *Journal of Biomechanics*, 122:110414.
- [51] Kastaniotis, D., Theodorakopoulos, I., Theoharatos, C., Economou, G., and Fotopoulos, S. (2015). A framework for gait-based recognition using kinect. *Pattern Recognition Letters*, 68:327–335.
- [52] Kieliba, P., Tropea, P., Pirondini, E., Coscia, M., Micera, S., and Artoni, F. (2018). How are muscle synergies affected by electromyography pre-processing? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4):882–893.
- [53] King, K., Yoon, S. W., Perkins, N., and Najafi, K. (2008). Wireless mems inertial sensor system for golf swing dynamics. *Sensors and Actuators A: Physical*, 141(2):619–630.
- [54] Kiss, R. M., Kocsis, L., and Knoll, Z. (2004). Joint kinematics and spatial–temporal parameters of gait measured by an ultrasound-based system. *Medical engineering & physics*, 26(7):611–620.
- [55] Klette, G. (2010). A recursive algorithm for calculating the relative convex hull. In *2010 25th International Conference of Image and Vision Computing New Zealand*, pages 1–7. IEEE.
- [56] Koutsiana, E., Ladakis, I., Fotopoulos, D., Chytas, A., Kilintzis, V., Chouvarda, I., et al. (2020). Serious gaming technology in upper extremity rehabilitation: scoping review. *JMIR Serious Games*, 8(4):e19071.
- [57] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.
- [58] Kugler, P., Jaremenko, C., Schlachetzki, J., Winkler, J., Klucken, J., and Eskofier, B. (2013). Automatic recognition of parkinson’s disease using surface electromyography during standardized gait tests. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5781–5784. IEEE.
- [59] Kulig, K. and Burnfield, J. (2008). The role of biomechanics in orthopedic and neurological rehabilitation. *Acta of bioengineering and biomechanics / Wrocław University of Technology*, 10:3–14.
- [60] Lau, H. and Tong, K. (2008). The reliability of using accelerometer and gyroscope for gait event identification on persons with dropped foot. *Gait & posture*, 27(2):248–257.
- [61] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* (2015). May; 521(7553): 436 10.1038/nature14539.
- [62] Lee, Y., Cho, S., and Kang, J. (2012). A study on the waste volume calculation for efficient monitoring of the landfill facility. In *Computer Applications for Database, Education, and Ubiquitous Computing*, pages 158–169. Springer.

- [63] Lei, L., Tao, Y., Jing-na, J., Ying, L., and Zhi-peng, L. (2011). Feasibility of surface electromyography signal acquisition for action recognition. *Chinese Journal of Tissue Engineering Research*, 15(22):4103.
- [64] Li, N., Dai, Y., Wang, R., and Shao, Y. (2015). Study on action recognition based on kinect and its application in rehabilitation training. In *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, pages 265–269. IEEE.
- [65] Li, Q. X., Chan, P. P., Zhou, D., Fang, Y., Liu, H., and Yeung, D. S. (2016). Improving robustness against electrode shift of semg based hand gesture recognition using online semi-supervised learning. In *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 1, pages 344–349. IEEE.
- [66] Lin, W., Meng, Y., Qiu, Z., Zhang, S., and Wu, J. (2017). Measurement and calculation of crown projection area and crown volume of individual trees based on 3d laser-scanned point-cloud data. *International Journal of Remote Sensing*, 38(4):1083–1100.
- [67] Lo, F. P.-W., Sun, Y., Qiu, J., and Lo, B. (2018). Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients*, 10(12):2005.
- [68] Lopez-Nava, I. H. and Muñoz-Meléndez, A. (2019). Human action recognition based on low-and high-level data from wearable inertial sensors. *International Journal of Distributed Sensor Networks*, 15(12):1550147719894532.
- [69] Macellari, V., Giacomozzi, C., and Saggini, R. (1999). Spatial-temporal parameters of gait: reference data and a statistical method for normality assessment. *Gait & Posture*, 10(2):171–181.
- [70] McGinnis, R. S. and Perkins, N. C. (2012). A highly miniaturized, wireless inertial measurement unit for characterizing the dynamics of pitched baseballs and softballs. *Sensors*, 12(9):11933–11945.
- [71] Meng, H., Wang, J., Lei, C., and Zhang, H. (2021). Action recognition and correction by using emg signal for health sports. *Internet Technology Letters*, 4(3):e241.
- [72] Meng, M., She, Q., Gao, Y., and Luo, Z. (2010). Emg signals based gait phases recognition using hidden markov models. In *The 2010 IEEE International Conference on Information and Automation*, pages 852–856. IEEE.
- [73] Mewett, D. T., Nazeran, H., and Reynolds, K. J. (2001). Removing power line noise from recorded emg. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 2190–2193. IEEE.
- [74] Mijailovic, N., Gavrilovic, M., Rafajlovic, S., Đuric-Jovicic, M., and Popovic, D. (2009). Gait phases recognition from accelerations and ground reaction forces: Application of neural networks. *Telfor Journal*, 1(1):34–36.
- [75] Mirek, E., Rudzińska, M., and Szczudlik, A. (2007). The assessment of gait disorders in patients with parkinson’s disease using the three-dimensional motion analysis system vicon. *Neurologia i neurochirurgia polska*, 41(2):128–133.

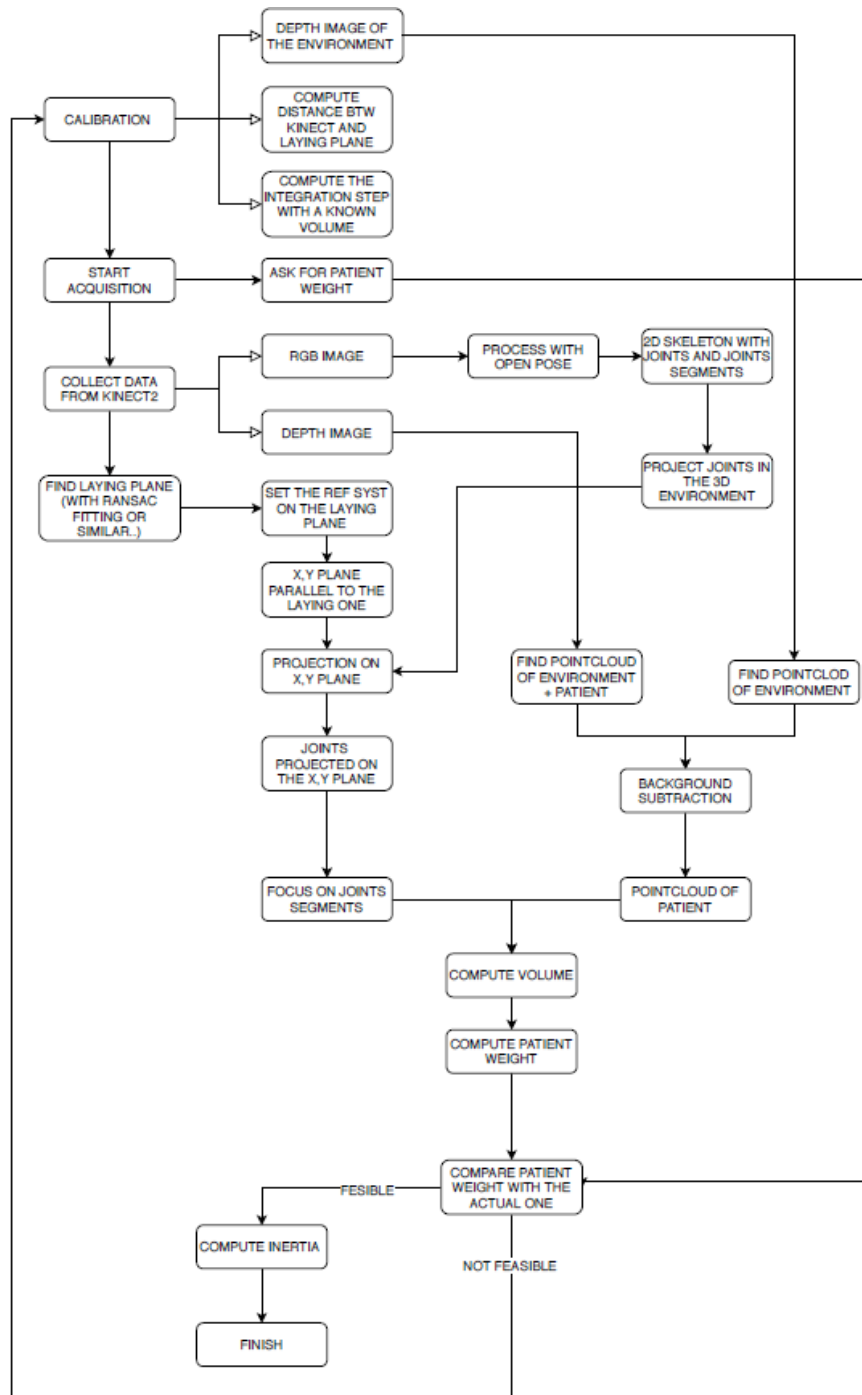
- [76] Moon, S., Park, Y., Ko, D. W., and Suh, I. H. (2016). Multiple kinect sensor fusion for human skeleton tracking using kalman filtering. *International Journal of Advanced Robotic Systems*, 13(2):65.
- [77] Mukhopadhyay, A. K. and Samui, S. (2020). An experimental study on upper limb position invariant emg signal classification based on deep neural network. *Biomedical signal processing and control*, 55:101669.
- [78] Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., and Yagi, Y. (2015). Similar gait action recognition using an inertial sensor. *Pattern Recognition*, 48(4):1289–1301.
- [79] Nguyen, V. D., Le, M. T., Do, A. D., Duong, H. H., Thai, T. D., and Tran, D. H. (2014). An efficient camera-based surveillance for fall detection of elderly people. In *2014 9th IEEE Conference on Industrial Electronics and Applications*, pages 994–997. IEEE.
- [80] Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote sensing of environment*, 56(1):1–7.
- [81] Palermo, F., Cognolato, M., Gijsberts, A., Müller, H., Caputo, B., and Atzori, M. (2017). Repeatability of grasp recognition for robotic hand prosthesis control based on semg data. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1154–1159. IEEE.
- [82] Pasinetti, S., Nuzzi, C., Covre, N., Luchetti, A., Maule, L., Serpelloni, M., and Lancini, M. (2020). Validation of marker-less system for the assessment of upper joints reaction forces in exoskeleton users. *Sensors*, 20(14):3899.
- [83] Pheasant, S. and Haslegrave, C. M. (2005). *Bodyspace: Anthropometry, ergonomics and the design of work*. CRC press.
- [84] Piskorowski, J. (2012). Powerline interference rejection from semg signal using notch filter with transient suppression. In *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pages 1447–1451. IEEE.
- [85] Pisoni, T., Conci, N., De Natale, F. G., De Cecco, M., Nollo, G., Frattari, A., and Guandalini, G. M. (2016). Ausilia: assisted unit for simulating independent living activities. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–4. IEEE.
- [86] Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H., and Atzori, M. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. *PloS one*, 12(10):e0186132.
- [87] Poitras, I., Biemann, M., Campeau-Lecours, A., Mercier, C., Bouyer, L. J., and Roy, J.-S. (2019). Validity of wearable sensors at the shoulder joint: Combining wireless electromyography sensors and inertial measurement units to perform physical workplace assessments. *Sensors*, 19(8):1885.
- [88] Preston, D. C. and Shapiro, B. E. (2012). *Electromyography and neuromuscular disorders e-book: clinical-electrophysiologic correlations (Expert Consult-Online)*. Elsevier Health Sciences.

- [89] Rawat, S., Vats, S., and Kumar, P. (2016). Evaluating and exploring the myo armband. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, pages 115–120. IEEE.
- [90] Rego, P., Moreira, P. M., and Reis, L. P. (2010). Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *5th Iberian conference on information systems and technologies*, pages 1–6. IEEE.
- [91] Salman, A., Iqbal, J., Izhar, U., Khan, U. S., and Rashid, N. (2012). Optimized circuit for emg signal processing. In *2012 International Conference of Robotics and Artificial Intelligence*, pages 208–213. IEEE.
- [92] Saponara, S. (2017). Wearable biometric performance measurement system for combat sports. *IEEE Transactions on Instrumentation and Measurement*, 66(10):2545–2555.
- [93] Scarano, V., De Chiara, R., Erra, U., et al. (2008). Meshlab: an open-source mesh processing tool.
- [94] Seel, T., Raisch, J., and Schauer, T. (2014). Imu-based joint angle measurement for gait analysis. *Sensors*, 14(4):6891–6909.
- [95] Semciw, A. I., Pizzari, T., Murley, G. S., and Green, R. A. (2013). Gluteus medius: an intramuscular emg investigation of anterior, middle and posterior segments during gait. *Journal of Electromyography and Kinesiology*, 23(4):858–864.
- [96] Shany, T., Redmond, S. J., Narayanan, M. R., and Lovell, N. H. (2011). Sensors-based wearable systems for monitoring of human movement and falls. *IEEE Sensors journal*, 12(3):658–670.
- [97] Shi, G., Chan, C. S., Li, W. J., Leung, K.-S., Zou, Y., and Jin, Y. (2009). Mobile human airbag system for fall protection using mems sensors and embedded svm classifier. *IEEE Sensors Journal*, 9(5):495–503.
- [98] Shi, Y., Tian, Y., Wang, Y., and Huang, T. (2017). Sequential deep trajectory descriptor for action recognition with three-stream cnn. *IEEE Transactions on Multimedia*, 19(7):1510–1520.
- [99] Shijie, J., Ping, W., Peiyi, J., and Siping, H. (2017). Research on data augmentation for image classification based on convolution neural networks. In *2017 Chinese automation congress (CAC)*, pages 4165–4170. IEEE.
- [100] Shuman, B. R., Schwartz, M. H., and Steele, K. M. (2017). Electromyography data processing impacts muscle synergies during gait for unimpaired children and children with cerebral palsy. *Frontiers in computational neuroscience*, 11:50.
- [101] Simão, M., Neto, P., and Gibaru, O. (2019). Emg-based online classification of gestures with recurrent neural networks. *Pattern Recognition Letters*, 128:45–51.
- [102] Sinpithakkul, C., Kusakunniran, W., Bovonsunthonchai, S., and Wattananon, P. (2018). Game-based enhancement for rehabilitation based on action recognition using kinect. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 0303–0308. IEEE.

- [103] Solomonow, M., Baratta, R., Bernardi, M., Zhou, B., Lu, Y., Zhu, M., and Acierno, S. (1994). Surface and wire emg crosstalk in neighbouring muscles. *Journal of Electromyography and Kinesiology*, 4(3):131–142.
- [104] Stegeman, D. and Hermens, H. (2007). Standards for surface electromyography: The european project surface emg for non-invasive assessment of muscles (seniam). *Enschede: Roessingh Research and Development*, 10:8–12.
- [105] Stocco, M., Luchetti, A., Tomasin, P., Fornaser, A., Ianes, P., Guandalini, G., Ty, J. F., Okahashi, S., Plopski, A., Kato, H., et al. (2019). Augmented reality to enhance the clinical eye: The improvement of adl evaluation by mean of a sensors based observation. In *International Conference on Virtual Reality and Augmented Reality*, pages 291–296. Springer.
- [106] Strutzenberger, G., Kanko, R., Selbie, S., Schwameder, H., and Deluzio, K. (2021). Assessment of kinematic cmj data using a deep learning algorithm-based markerless motion capture system. *ISBS Proceedings Archive*, 39(1):236.
- [107] Sukul, D. K., Den Hoed, P., Johannes, E., Van Dolder, R., and Benda, E. (1993). Direct and indirect methods for the quantification of leg volume: comparison between water displacement volumetry, the disk model method and the frustum sign model method, using the correlation coefficient and the limits of agreement. *Journal of biomedical engineering*, 15(6):477–480.
- [108] Sylos-Labini, F., La Scaleia, V., d’Avella, A., Pisotta, I., Tamburella, F., Scivoletto, G., Molinari, M., Wang, S., Wang, L., van Asseldonk, E., et al. (2014). Emg patterns during assisted walking in the exoskeleton. *Frontiers in human neuroscience*, 8:423.
- [109] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- [110] Tan, C.-W., Coutts, F., and Bulley, C. (2013). Measurement of lower limb volume: agreement between the vertically oriented perometer and a tape measure method. *Physiotherapy*, 99(3):247–251.
- [111] Togo, S., Matsumoto, K., Kimizuka, S., Jiang, Y., and Yokoi, H. (2021). Semi-automated control system for reaching movements in emg shoulder disarticulation prosthesis based on mixed reality device. *IEEE Open Journal of Engineering in Medicine and Biology*, 2:55–64.
- [112] Türker, K. S. (1993). Electromyography: some methodological problems and issues. *Physical therapy*, 73(10):698–710.
- [113] Vugts, M. A., Joosen, M. C., Mert, A., Zedlitz, A., and Vrijhoef, H. J. (2017). Serious gaming during multidisciplinary rehabilitation for patients with complex chronic pain or fatigue complaints: study protocol for a controlled trial and process evaluation. *BMJ open*, 7(6):e016394.
- [114] Wang, H., Song, Z., Li, W., and Wang, P. (2020). A hybrid network for large-scale action recognition from rgb and depth modalities. *Sensors*, 20(11):3305.

- [115] Wang, J., Tang, L., and Bronlund, J. E. (2013). Surface emg signal amplification and filtering. *International Journal of Computer Applications*, 82(1).
- [116] Wang, W.-W. and Fu, L.-C. (2011). Mirror therapy with an exoskeleton upper-limb robot based on imu measurement system. In *2011 IEEE International Symposium on Medical Measurements and Applications*, pages 370–375. IEEE.
- [117] Wei, P., Zhang, J., Tian, F., and Hong, J. (2021). A comparison of neural networks algorithms for eeg and semg features based gait phases recognition. *Biomedical Signal Processing and Control*, 68:102587.
- [118] Whittle, M. W. (1993). Gait analysis. In *The soft tissues*, pages 187–199. Elsevier.
- [119] Wu, D., Sharma, N., and Blumenstein, M. (2017). Recent advances in video-based human action recognition using deep learning: A review. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2865–2872.
- [120] Xu, W.-H., Feng, Z.-K., Su, Z.-F., Xu, H., Jiao, Y.-Q., Deng, O., et al. (2014). An automatic extraction algorithm for individual tree crown projection area and volume based on 3d point cloud data. *Spectroscopy and Spectral Analysis*, 34(2):465–471.
- [121] Xu, Z., Shen, L., Qian, J., and Zhang, Z. (2020). Advanced hand gesture prediction robust to electrode shift with an arbitrary angle. *Sensors*, 20(4):1113.
- [122] Yang, S. X., Christiansen, M. S., Larsen, P. K., Alkjær, T., Moeslund, T. B., Simonsen, E. B., and Lynnerup, N. (2014). Markerless motion capture systems for tracking of persons in forensic biomechanics: an overview. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2(1):46–65.
- [123] Yao, T., Gao, F., Zhang, Q., and Ma, Y. (2021). Multi-feature gait recognition with dnn based on semg signals. *Math. Biosci. Eng.*, 18:3521–3542.
- [124] Ye, M., Yang, C., Stankovic, V., Stankovic, L., and Kerr, A. (2016). A depth camera motion analysis framework for tele-rehabilitation: Motion capture and person-centric kinematics analysis. *IEEE Journal of Selected Topics in Signal Processing*, 10(5):877–887.
- [125] Yeung, L.-F., Yang, Z., Cheng, K. C.-C., Du, D., and Tong, R. K.-Y. (2021). Effects of camera viewing angles on tracking kinematic gait patterns using azure kinect, kinect v2 and orbbecc astra pro v2. *Gait & Posture*, 87:19–26.
- [126] Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. *Computers in biology and medicine*, 96:189–202.
- [127] Yu, H., Alaqtash, M., Spier, E., and Sarkodie-Gyan, T. (2010). Analysis of muscle activity during gait cycle using fuzzy rule-based reasoning. *Measurement*, 43(9):1106–1114.
- [128] Zanetti, M., Nollo, G., and De Cecco, M. (2018). Mixed reality for smart health. *IEEE Smart Cities Initiative*.

-
- [129] Zanghieri, M., Benatti, S., Burrello, A., Kartsch, V., Conti, F., and Benini, L. (2019). Robust real-time embedded emg recognition framework using temporal convolutional networks on a multicore iot processor. *IEEE transactions on biomedical circuits and systems*, 14(2):244–256.
- [130] Zhang, Z.-Q. and Wu, J.-K. (2011). A novel hierarchical information fusion method for three-dimensional upper limb motion estimation. *IEEE transactions on instrumentation and measurement*, 60(11):3709–3719.
- [131] Zhi, Y., Zhang, Y., Chen, H., Yang, K., and Xia, H. (2016). A method of 3d point cloud volume calculation based on slice method.
- [132] Zihajehzadeh, S., Lee, T. J., Lee, J. K., Hoskinson, R., and Park, E. J. (2014). Integration of mems inertial and pressure sensors for vertical trajectory determination. *IEEE transactions on Instrumentation and Measurement*, 64(3):804–814.



PSEUDO CODE

Algorithm 1 Pre-processing - Re-Orient the $S_{3DTargetObject}$ with PCA

Require: $S_{3DTargetObject}$
 $S_{3DTargetObject} \leftarrow$ coordinate matrix of the original point cloud $[X, Y, Z]_{N,3}$

- Compute Principal Component Analysis
- Compute the roto-translation matrix to orient the $S_{3DTargetObject}$ along the main axes
- Orient the $S_{3DTargetObject}$ throughout matrix multiplication

Algorithm 2 Pre-processing - Cloud Analysis

Require: $S_{3DTargetObject}$, $n_{neighbours}$
 $S_{3DTargetObject} \leftarrow$ coordinate matrix of the original point cloud $[X, Y, Z]_{N,3}$
 $n_{neighbours} \leftarrow$ total amount of considered neighbors

- Compute the Matrix of Distances "D". The generic Element $D_{i,j}$ is the Euclidean distance of the point i from the point j
- \forall point select only the closest $n_{neighbours}$ neighbors, draw the histogram distribution and compute
 - $Q_{0.5}$ = average distance of \forall Cloud point from its neighbors
 - $Q_{0.85}$ = Quantile at the 85 % of the distribution

Algorithm 3 Pre-processing - S_{3DBox} Volume Computation

Require: $S_{3DTargetObject}$
 $S_{3DTargetObject} \leftarrow$ PointCloud matrix of coordinates $[X, Y, Z]_{N,3}$

- Select the maximum and minimum coordinate for each direction
 - $min_x = \min(X), max_x = \max(X)$
 - $min_y = \min(Y), max_y = \max(Y)$
 - $min_z = \min(Z), max_z = \max(Z)$
- Create a regular prism with 8 vertexes $[min_x, min_y, min_z], [min_x, min_y, max_z], etc...$ that encloses the whole point cloud (S_{3DBox})
- compute V_{box} equal to the Volume of S_{3DBox}

Algorithm 4 Cube Explosion Affiliation Criteria

Require: $S_{3DTargetObject}$, N
 $S_{3DTargetObject} \leftarrow$ coordinate matrix of the original point cloud $[X, Y, Z]_{N,3}$
 $th_{intercepted} \leftarrow 3$;
 $N \leftarrow$ the total amount of $P_{generated}$;
 $x_{P_{generated}} \leftarrow x_{min}$;
 $y_{P_{generated}} \leftarrow y_{min}$;
 $z_{P_{generated}} \leftarrow z_{min}$;
 $Points_{Inside} = Points_{Outside} = []$;
 $n_{InsidePoints} = n_{OutsidePoints} = 0$;
while $x_{P_{generated}} < x_{max}$ **do**
 while $y_{P_{generated}} < y_{max}$ **do**
 while $z_{P_{generated}} < z_{max}$ **do**
 Create a cube around $P_{generated}$;
 Initialize $l_{cube} = Q_{0.5} 3.5^{(\frac{Q_{0.85}}{Q_{0.5}} - 1)}$;
 for \forall direction x, y, z **do**
 $J_{\eta Array} =$ Empty Array;
 $\eta \leftarrow$ select the direction;
 for positive and negative direction **do**
 Extrude the Cube face along η ;
 Intercept the $S_{3DTargetObject}$;
 $P_{intercepted} =$ intercepted points;
 $Tot =$ amount of $P_{intercepted}$;
 if $Tot > th_{intercepted}$ **then**
 $l_{cube} = 0.1 * l_{cube}$;
 Repeat the extrusion;
 else
 if $Tot < th_{intercepted}$ **then**
 Subdivide $P_{intercepted}$ in clusters along η ;
 Each cluster is $P_{cluster}$;
 $Tot_{clusters} =$ total amount of clusters;
 if $Tot_{clusters}$ is odd **then**
 $J_{\eta} \leftarrow$ inside ;
 else
 $J_{\eta} \leftarrow$ outside ;
 end if
 Save the J_{η} in $J_{\eta Array}$;
 end if
 end if
 end for
 end for
 $J = \text{mode}(J_{\eta Array})$;
 if $J == 1$ **then**
 Save $P_{generated}$ in $Points_{Inside}$;
 $n_{InsidePoints} += 1$;
 else
 Save $P_{generated}$ in $Points_{Outside}$;
 $n_{OutsidePoints} += 1$;
 end if
 $z_{P_{generated}} += N^{\frac{1}{3}} - 1$
 end while
 $y_{P_{generated}} += N^{\frac{1}{3}} - 1$
 end while
 $x_{P_{generated}} += N^{\frac{1}{3}} - 1$
 end while
 end while
end while

Algorithm 5 Final Volume Computation

Require: InsidePoints, OutsidePoints, V_{box}
 $FinalVolume = \frac{Total_{Inside}}{Total_{Inside} + Total_{Outside}} * V_{Box}$
