UNIVERSITY OF PADOVA

DEPARTMENT OF INFORMATION ENGINEERING

Ph.D. Course in Information Engineering

Information and Communication Science and Technologies Curriculum

XXXVI series

Deep learning-based tools and communications for sensing

Ph.D. Candidate
Silvia Zampato

Ph.D. Supervisor
Professor Michele Rossi

Ph.D. Co-supervisor
Professor Zimi Sawacha

Ph.D. Coordinator
Professor Fabio Vandin

Academic Year
2022–2023

To curious people who inspire others with their passion

Abstract

The advancement of deep-learning technologies and the pervasive deployment of communication networks open new possibilities for sensing applications. Deep-learning tools are changing the paradigm from passive data collection to enhanced possibilities of interpretation and meaningful pattern extraction. On the other hand, communications are becoming widespread and play a crucial role in our society, thus the integration of sensing features appears as a natural consequence of the maximum exploitation of the communication infrastructures and for the enhancement of communications themselves.

In general, sensing is a fundamental and indispensable ability in our daily existence because provides information and allows our interaction with the world, thus the possibilities offered by these relatively new tools have the power to innovate society. In recent years, different domains have been transformed by sensorization as automation in healthcare, industry, agriculture, automotive, and others.

This thesis contributes to the field of sensing from different research perspectives. First, remote sensing is explored for movement monitoring in the clinical gait analysis context. We developed two multidisciplinary approaches involving computer vision and radar sensing coupled with deep learning processing neural networks for data processing. Then, we focused on multimodal acquisition systems. The information derived from different signals collected simultaneously is considered and enabled. Some practical problems related to the creation of adequate set-ups are discussed and we proposed both hardware and software solutions to different scenarios. Finally, we exploited the side information obtained by communications systems for human sensing. This proposed system is part of a more general scenario of integrated communications and sensing that is gaining more relevance due to the features of next-generation communications networks. In the last chapter, we discuss this interesting double role of communication networks and we propose a framework to detect human positioning indoors with no hardware specific for sensing.

Contents

Listing of figures

Listing of tables

Listing of acronyms

Symbols

6G Sixth Generation

A

AoA Angle of Arrival

AP access points

AZ azimuth

B

BLE Bluetooth Low Energy

BN Batch Normalization

BQ Ball Query

C

CA-CFAR Cell-Averaging Constant False Alarm Rate

CM-KF Converted-Measurements Kalman Filter

CMC coefficient of multiple correlation

CSI Channel State Information

CT computed tomography

CWT Continous Wavelet Transform

D

DBSCAN Density-Based Spatial Clustering of Applications with Noise

E

EEG electroencephalogram

EHR electronic health records

EL elevation

EMG Electromyography

F

FFT Fast Fourier Transform

FMWC Frequency Modulated Continuous Wave

FN false negatives

FP false positives

FP Feature Propagation

FPS Farthest Point Sampling

G

GA Gait Analysis

GATT General ATTribute Profile

H

HR Heart Rate

I

IA Initial Access

IF Intermediate Frequency

IMU Inertial Measurement Unit

IoU Intersection over Union

K

KLT Kanade-Lucas-Tomasi

L

LoS Line-of-Sight

M

m-health mobile-health

MIMO Multiple Input Multiple Output

MLP Multi Layer Perceptron

mmWave millimeter Wave

MRG Multi-Resolution Grouping

MRI magnetic resonance imaging

MSG Multi-Scale Grouping

MTI Moving Target Indication

N

NR New Radio

O

OFDM Orthogonal Frequency Division Multiple access

Open-MBIC Multiple Ble for Iot Connections

P

PPG photoplethysmogram

R

RCS Radar Cross Section

RD Range Doppler

REMOCOP REhabilitation MOnitoring of Covid-19 survivors and Chronic Obstructive Pulmonary disease patients

RF Radio Frequency

RMSD root mean squared distance

RPN Region Proposal Network

RSSI Received Signal Strength Indicator

RTT Round Trip Time

S

SA Set Abstraction

sEMG surface electromyography

peripheral oxygen blood saturation

SS Synchronisation Signal

SVD Singular Value Decomposition

T

TN true negatives

TP true positives

U

UUID Universally Unique IDentifier

V

VNA Vector Network Analyzer

1 Introduction

Sensing has a crucial role in our daily lives, in fact, it is the fundamental capability to perceive, understand, and interpret the world in which we live. Through our bodies and the use of our five senses, i.e., sight, hearing, touch, taste, and smell, we collect important information about the environment, enabling us to live everyday situations. Indeed, sensing offers the possibility to gather information regarding both our survival and the more abstract experiences; we can detect dangers and find supplies, but also enjoy the beauty, share experiences, and make informed decisions.

The existence of biological sensing is fundamental to further enhancements and explains our attitude of interaction with the world. The human interest to go beyond the already gained capabilities leads to the continuous technological effort in empowering the sensing possibilities. In fact, due to the possibility of gathering information about the context, sensing technologies have contributed to the development of a great amount of fields opening new possibilities. Among others, we mention some fields that have experienced a particular growth in recent years. In industry, sensors monitor the conditions, enable to make automatic decisions, assist in predictive maintenance, and increase the workers' safety. Smart cities are the concept of the use of sensing technlogies to improve the quality of residents' life, e.g., managing traffic, optimizing services, and reducing consumptions. The smart use of resources in energy management is an example of a challenge that can be faced also through the use of sensing and, in the more general picture, environmental monitoring is an important tool to track and mitigate the effects of pollution, climate change, and natural disasters; e.g., the parameters involved are temperature, humidity, air and water quality. In agriculture, sensors can help to address human intervention and the resources management to obtain increased crop yields and a more sustainable agriculture paradigm. Some established methodologies are related to soil conditions assessment, crop health monitoring, and irrigation optimization. Sensing is practically related to the capability of acquiring important information at the right time to enable time- sensitive decisions based on the gained knowledge.

Moreover, sensing capabilities enables time monitoring and the possibility to evaluate the impact of previous informed decisions.

In this thesis, the sensing solutions are mostly related to healthcare and communication applications. In healthcare, sensors can continuously monitor vital signs, provide tools for diagnosis, and real-time feedback for personalized treatment. Wearable sensors and remote monitoring systems have improved patient outcomes and allowed for early intervention in critical cases.

In telecommunications, network operators exploit the context knowledge to adapt the policies to deliver their services and in the next generation communication the integrated sensing-assisted communications is expected to be a key feature as it enables new applications, such as autonomous driving and augmented/virtual/ mixed reality, that require reliable context information in real-time.

On the other hand, throughout human history, many technologies initially designed for a specific task have been adjusted to address new challenges, often resulting in surprisingly good outcomes. In fact, the aspects that might be seen as a source of errors or disturbance in one context can sometimes be exploited in a new setting to enable novel applications. In the wireless communications domain, the exploitation of the multipath effect is a perfect example of the aforementioned situation. The receiver obtains multiple versions of the transmitted signal that have been delayed, frequency-shifted, and weakened due to the presence and the nature of the objects in the environment which reflect the signal in different spatial directions. Despite being a typical communication hurdle well addressed in the literature, this effect combined with appropriate processing methods can reveal a great amount of information about location, movements, and properties of these reflecting objects. Thus, besides the tools specifically designed for sensing, communications have become an important mean for sensing. Due to the high density of communications infrastructure in our society, there is a practical convenience to exploit communication signals for sensing purposes both indoors and outdoors. Moreover, the integration of deep learning tools in next-generation networks is expected to permit actions based on the network knowledge through data analysis. Deep learning techniques can be incorporated at various network levels, ranging from end devices and access points to the network infrastructure and cloud-based applications, depending on the specific use case[1].

Deep learning integration in communications opens the possibility for frameworks capable of adapting to changing environmental conditions through sensing. In fact, the recent advent of deep learning permits the sensing possibilities exploiting the capability to learn patterns from the data and recognize them in the new inputs; the perspective changes from the data collection to data interpretation, automatic learning, and interpretation. Many different fields are adopting deep learning techniques with success to improve their results or expand their sensing capabilities. For example, in computer vision, neural networks can analyze images and videos, enabling object detection, segmentation, and tracking. Deep learning models can process time series from sensors of any quantity. Referring to this thesis, for example, to monitor the patients' health and output predictions in the healthcare domain.

In conclusion, sensing applications are heavily involved in multiple aspects of our lives, and the evolution of technology continues to create innovation, thus we can expect new sensing solu-

tions to emerge, further enhancing our ability to monitor, interact, and discover the world around us. These applications increase the sensing efficiency and can contribute to an improvement of life quality provided that their use is combined with the awareness of the different tools' risks.

1.1 Thesis content

In the next chapters, different aspects and solutions related to sensing-enabled environment awareness will be deepened. In Chapter 1 we address the problem of human movement monitoring by presenting a deep-learning approach applyied to multiview videos. The proposed solution addresses in particular the lack in literature of a clinical approach that requires a three dimensional joint movement description. A different perspective to the same problem is presented in Chapter 2 in which we explored the radar sensing as a non-invasive wireless technique to predict the human pose during gait and to enable clinical gait analisys. Chapter 3 refers to the practical implementation of sensing platforms for multisignals collection. In particular, the first part of the chapter is devoted to the development of an Android library that manage multiple simultaneous Bluetooth Low Energy connections and its application to a real case for patients health monitoring. In the second part, we describe the design and realization of a synchronization system for the data collection of different biological signals during gait. Then, Chapter 4 describes the work for the realization of a wireless communication-based sensing system for the positioning of a subject in an indoor environment. Beamforming is a key feature of the 5G (and beyond) communication links and it is here exploited for sensing purposes. In Chapter 5 we conclude the thesis with suggestions for future researches and final remarks.

2

Video markerless motion capture

2.1 Motion capture systems

Quantitative movement monitoring is the object of study of motion caption and finds application in different fields such as security, entertainment, and autonomous driving as well as in rehabilitation and biomedical research which is the field of interest of this project. Gait analysis has been known to be an effective method to distinguish between healthy and pathological gaits since the late 1800s [2]. The clinical gait analysis includes different components, i.e., kinematics (joint angles), kinetics (joint forces), muscular activity, foot pressure, and energetics (measurement of energy utilized during locomotion). In this thesis are presented some technical approaches and acquisition systems related to kinematics, thus the measurement of the joint angles in the three-dimensional space. The state-of-the-art most accurate optical systems record the subject's movements with a set of synchronized cameras and usually employ a set of active or passive markers which are attached to the subject's skin [3]. The markers are tracked by feature extraction algorithms either during the acquisition or in post-processing. However, this procedure requires expensive instrumentation and long pre- and post-processing time. In addition, issues related to marker occlusions remain very common [3]. Finally, the subject has to wear only underwear and could be uncomfortable, this is another weakness considering that clinical application usually involves fragile subjects.

Increasing attention has been dedicated to markerless motion capture technologies as a solution to these limitations. These approaches usually use subject-specific anatomical models providing both morphological and kinematic information. The tracking is obtained by iteratively matching these descriptors to the data extracted from the records, via background subtraction and visual hull reconstruction, as in [4]. These methods have achieved remarkable precision, but they generally require controlled experimental conditions and the computational complexity to complete the processing does not permit a real-time application of those systems [5]. With the advent of

deep neural networks, the pose estimation field has started to exploit such architectures [6]. The main limitation is that these models are usually trained to estimate the joints' positions in the 3D space because the main public datasets annotate the joint centers [7]–[9]. However, the joint center itself is not sufficient to capture the complete range of the joint motion. Moreover, the joints' centers are not visible because are inside the body, so they are manually inferred considering the point of view of the picture and obtaining a projection of them on the frame itself. To the best of our knowledge, this has not been proven to be sufficiently precise in the reconstruction of a 3D model with anatomical validity.

The method presented in this chapter addresses the problem in a novel way. The estimation procedure aims to predict a set of superficial landmarks defined according to an established anatomical protocol. The advantages of this approach are summarized as follows:

- to provide a new approach in which the prediction of the superficial keypoints permits the estimation of three-dimensional angles instead of planar angles between limbs and therefore to describe the complete motion of the joints.
- the possibility of estimating the original dataset precision with respect to an established approach. This is not true for the direct annotation of joint centers.

2.2 Related work

The possibility to capture the movement is appealing for a wide range of applications and the number can increase as the technology becomes more affordable and precise. Considering the clinical application, it is of paramount importance to completely describe the motion. The common limitation of a great number of approaches is the direct prediction of the joints' positions. The information of the centers permits associating the movement to planar angles defined by the limb segments, but the range of motion of each joint is described when three-dimensional angles are estimated. Flexion-extension movement with some approximation could be described as the planar angle between the joints, in particular for the knee joint. Besides this approximation, a complete sensing procedure needs to include the abduction-adduction angle and rotations.

Very different technologies and methods have been developed to capture motion: magnetic, inertial, mechanical, and optical systems. In particular, the standard methodologies in the clinical domain are optical because permit the application of a less invasive process and they can reach very high accuracies.

The most common approach is camera-based with infrared cameras that are used to triangulate the location of reflective markers attached to the subject. The markers are tracked from different perspectives and their 2D positions are obtained by means of feature extraction algorithms. The system is calibrated to obtain a global reference system and the relative position of each camera, thus it is possible to retrieve the 3D position of the marker either during the acquisition or in the post-processing phase. However, this procedure requires expensive instrumentation that includes the cameras and the processing system itself it is usually not portable. Moreover, the pre-processing time is often quite long as it includes the subject preparation with markers on the

bones' landmarks and the system calibration. After the data acquisition, there is a data processing step in which the operator has to deal with the occlusions and sometimes they remain a problem. In fact, this is the main problem, even in most advanced systems equipped with markers auto-labeling [3].

Manual video tracking is a method that can be used both with the help of markers or-in some cases-without them if the bones' prominencies are evident. In this case, the markers are required to be visible and adherent to the subject's skin during movement,e.g. tape. In some systems there are implemented some mechanisms that help to speed up the process. Among all we mention the auto-tracking, i.e., the landmarks are manually positioned just in the first frame of a video acquisition and, through the application of computer vision-based algorithms, the marker is tracked across the different frames or at least a tentative position is suggested to the operator. In the software Track on Field (BBSof Srl), the algorithm used for the auto-tracking implementation is the algorithm proposed by Kanade-Lucas-Tomasi (KLT) [10], [11] in which the main idea is to solve the image registration problem through a local search using gradients of the first and second order. This algorithm have shown to be effective to propose the locations of the markers also in challenging environments, reducing the required manual intervention [12].

In this context, markerless methods are a research challenge still to be solved but present several advantages. The first advantage of markerless techniques is related to the time efficiency in the data collection. In general, even if markerless requires some processing time for computation, stereo requires subject preparation and also some post-processing. Moreover, this preparation has to be done by an expert operator to identify the bones' prominences correctly, while a markerless method can be usually used by any operator at least for the data acquisition. Another benefit is the patient comfort that derives from the disuse of markers. The wearing required for stereophotogrammetry is tight or just underwear.

Some markerless approaches employ depth cameras that can estimate depth by computing the delay from light emission projected towards the object to backscattered light detection. In [13] a single depth image is used as it is treated as a point cloud and then, the pose is identified as the best match measuring the similarity within a database of poses. The authors of [14] combine volumetric dynamic reconstruction with data-driven template fitting to simultaneously reconstruct the pose from a single depth camera. These approaches estimate the pose but the computation of the kinematics is difficult or impossible because the output of the system is the joints' centers.

Among the optical markerless approaches based on computer vision techniques, the silhouette-based can obtain good results in terms of completeness of the data as the 3D motion can be retrieved [5]. On the other hand, the main drawbacks of the silhouette-based approach are the great amount of computational resources required and the very low flexibility to new subjects because subject-specific models are necessary. Moreover, the model requires manual intervention in each situation, there is no learning process.

Markerless is gaining even more relevance with the advent of deep-learning techniques that are able to learn from the context. The change of the paradigm done by deep-learning is from passive sensing to active prediction based on learning. This opens the door to a great variety of new tools to empower also gait analysis. Thus, in the last years, a high number of approaches

have been proposed to solve the problem [15], [16].

Still, there is a lack in the literature of approaches that allow the reconstruction of three-dimensional reference frames for each joint to compute the joints' angles completely. Our approach addresses this problem by predicting the superficial keypoints on the subject's skin instead of the joints' centers. This is possible with a vision-based deep-learning model trained on labeled data. Using the joint centers, it is possible to define a complete system reference frame for each limb to estimate the angles while the subject is moving. From this perspective, our project try to fill this gap in the literature and the results we obtained were published in [17]–[21].

2.3 Dataset for training

The dataset used for the training of the pose estimator was manually labeled by us to include the position of the markers defined by the IORgait protocol [22]. This protocol was chosen between the gold standards for the gait analysis because it considers fewer keypoints, i.e., 30, but it permits to build a 3D reference frame for each joint and to estimate the flexion-extension, abduction-adduction, and internal-external rotation angles. The labeling was done by video tracking with Track on Field,i.e., manually annotating the visible keypoints indicated by some markers or looking at the bones' prominences. The auto-tracking implemented suggested the positions, but due to the noise they needed to be checked by the operator in each frame. After the video tracking, there is the triangulation step in which all the annotations on the dataset images are used to create the three-dimensional trajectories for each marker. The trajectories usually present some holes due to occlusions that prevent the annotation in the video tracking step. For this reason, two different strategies are employed to reconstruct the trajectories and we discuss them in Sec. **reconstruction**. Finally, the 3D markers trajectories are reprojected frame by frame onto the cameras image planes. With these steps, we gain the positioning of occluded keypoints.

The design of these steps aims to estimate the position of the points that are occluded from a certain observation angle. The projection of the reconstructed trajectories permits to increase the number of markers that could be detectable only by tracking. Moreover, in some cases, the available data had markers. To create a markerless system is important to decouple the annotation from the visible markers in order to permit the model to learn an association between markers and bones' prominences, not with markers.

The steps for the dataset creation are summarized in the following:

1. **Acquisition:** the subjects are recorded while walking.
2. **Video tracking:** the keypoints are identified in the video frames.
3. **Triangulation:** the two-dimensional information from different points of view is triangulated to obtain 3D keypoints.
4. **Reconstruction:** the trajectories of the 3D keypoints were reconstructed where needed.
5. **Projection:** 3D points were projected onto the image planes of the cameras.

2.3.1 Dataset acquisition

The subjects involved in the experiment are 5 children (Tab. 2.1). Data were collected from the subjects with stereophotogrammetry, with tape for video tracking, and just walking without markers for markerless video tracking and testing the proposed approach. All data from each subject was collected subsequently on the same day. They were asked to walk at a self-paced speed while recorded by 4 GoPro cameras positioned at the room angles. The walking records had to contain at least 6 gait cycles, 3 for each side.

Subject	Age [years]	Height [cm]	Weigth [Kg]	BMI [kg/m^2]	Shoe size	Gender
Sub1	7	130	33	19.53	35	F
Sub2	6	117	27	19.72	30	M
Sub3	10	151	40	17.54	37	F
Sub4	7	121	27	18.44	35	M
Sub5	9	137	29	15.45	35	M

Table 2.1: Subjects demographic and biometric information.

2.3.2 Video tracking

For each video frame, the keypoints were annotated with the help of Track on Field (BBSof Srl). The synchronized acquisition videos were trimmed to obtain separated gait cycles. The operator annotated the markers in the first frame and then the tracking of the markers position was proposed by the software. The markers IORgait protocol is described in Fig.2.3

The lighting condition and the occlusion due to moving objects with different depths were challenging for the embedded algorithm, thus all frames were checked manually. Moreover, for the annotation of the trials in which the subjects did not wear the markers, the algorithm was not able to identify the bones' prominences. An example of tracking session is shown in Fig.2.1-2.2.

In Track on Field also the camera calibration is performed and the image matrices are retrieved. The standard procedure is to perform both intrinsic and extrinsic calibration with the use of a checkerboard and the application of the Zhang algorithm [23].

2.3.3 Reconstruction

The three-dimensional predictions are often incomplete due to occlusions and the 3D trajectories result with holes. The small holes, i.e., less than 5 frames are filled with spline interpolation [24]. While a wider loss of information is retrieved making some assumptions about the body shape. The first is related to the protocol definition; the human body in the IOR gait model is composed of 8 segments and for each there are 4 estimated keypoints, three of them are used to compute the system reference frame and the fourth is a backup point. The other assumption is the rigidity of these segments: the lengths are considered fixed between them with no modifications due to the movement.

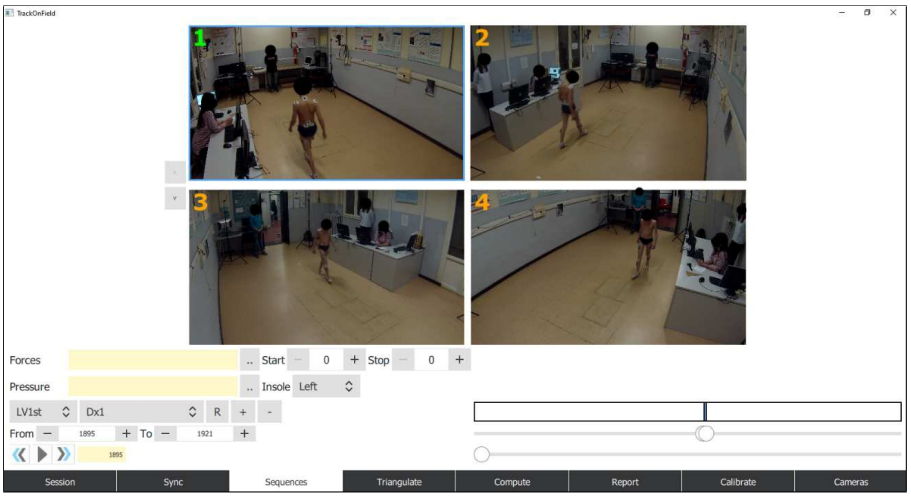


Figure 2.1: Screen of the software used for video tracking.

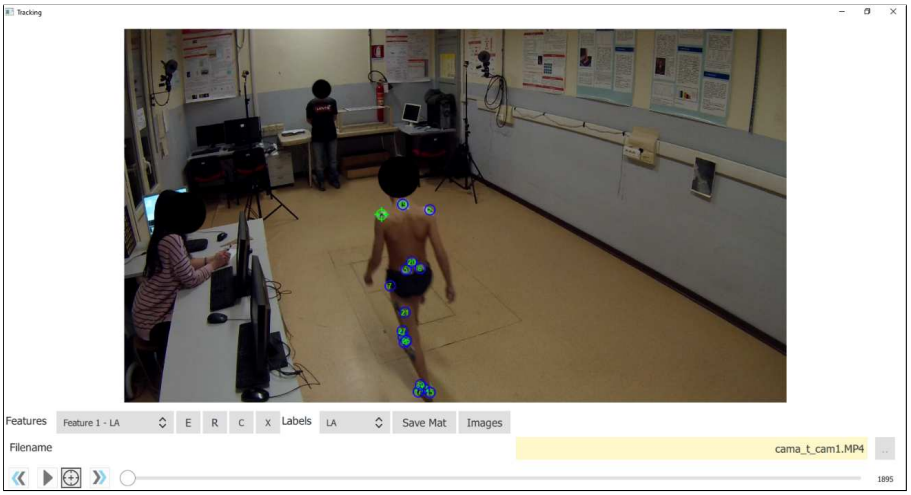


Figure 2.2: Screen of the software used for video tracking - window for tracking.

IORgait protocol description According to [22], the considered landmarks are 4 for each considered anatomical segment, except the thighs. In thighs, there are three markers and the fourth point is the center of the femoral head (FH) which is assumed to coincide with the center of the acetabulum. It is reconstructed using the geometrical prediction method proposed by [25] that is based on the location of the four anatomical landmarks of the pelvis (RASIS, LASIS, RPSIS, LPSIS). The segments are defined as follows:

- trunk: the 7^{th} cervical vertebra (C7), acromions of scapulas (RA, LA), the 5^{th} lumbar vertebra (L5)

- pelvis: the anterior (A) and posterior (P) iliac spines (RASIS, LASIS, RPSIS, LPSIS)

- thighs: the most lateral prominence of the great trochanter (RGT/LGT), of the lateral and medial epicondyle (RLE/LLE, RME/LME) and the hip joint center (RHJC/LHJC)
- shanks: the proximal tip of the head of the fibula (RHF/LHF), the most anterior border of the tibial tuberosity (RTT/LTT), the lateral prominence of the lateral and medial malleolus (RLM/LLM, RMM/LMM)
- feet: the achilles tendon insertion on the calcaneous (RCA/LCA), and the dorsal margins of the first (RIM/LIM), second (RIIT/LIIT) and fifth (RVM/LVM) metatarsal heads, the ordinal numbers are expressed as roman.

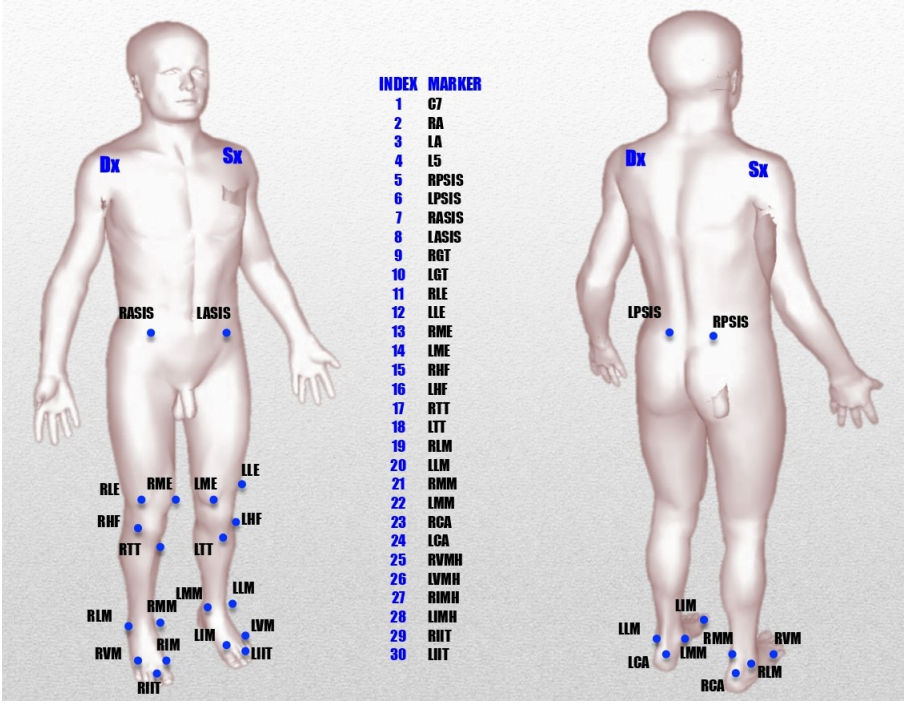


Figure 2.3: IOR-gait protocol description (Sec. 2.3.3).

Considering both the body shape and rigidity of segments, for each segment it is possible to estimate its position provided that the other three marker positions of that segment were already estimated. This happens with the help of a static acquisition from which to extract the relative distances between markers belonging to the same segment. The rigid segment reconstruction is not exclusively but prominently applied to medial landmarks which are often not visible due to anatomical occlusions during gait.

2.3.4 Projection

Both intrinsic and extrinsic calibration used for undistortion and triangulation were saved and used to project the 3D points on the image planes of the cameras. This permits to have the informa-

tion of the hidden keypoints at the cost of projection error. Each pre-processing step introduces error and the projection error is measured as root mean squared error (RMSD) and coefficient of multiple correlation (CMC). The results are in Sec.2.6. For the scope of the estimation of the projection error, the original manual annotations are assumed to be correct when availables and the projections are compared to them. The error is measured in pixel units and it is compared to the average width of the tape area (7-8 pixels) which is assumed to be the maximum imprecision that the operator can introduce when annotating the subjects with tape. In this assumption, we assume the operator annotate markers correctly where the tape is visible, excluding human errors as confusion between different keypoints, forgetting markers, and others. Moreover, the imprecision is related to the operator when annotating the subjects that are not wearing tape. Indeed, the operator should be well instructed to recognize the bones' landmarks. Finally, it exists inter-operator error in the positioning of the markers. The manual annotation is here assumed to be the ground truth but the real scenario shows a greater complexity.

Figure 2.4: Keypoints after processing.

After the projection step, another interpolation step is done to fill the remained holes considering the 2D trajectories with spline interpolation.

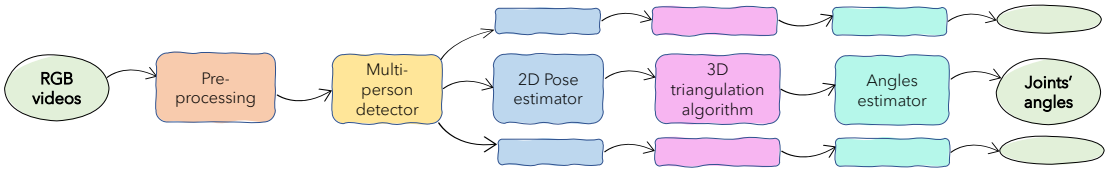


Figure 2.5: Processing pipeline - The processing scales with the number of subjects found by the person detector.

2.4 Deep IOR

2.4.1 Data preprocessing

DeepIOR is a highly modular pipeline that involves an undistortion step, a neural network for person detection, another deep-learning model to estimate the pose, a triangulation step, and the angles estimation. The steps are sequential but independent in order to permit further improvements to each step and to access intermediate results, if required. In this section, the pipeline steps are discussed in the order of the data flow as indicated in Fig. 2.5.

2.4.2 Person detection

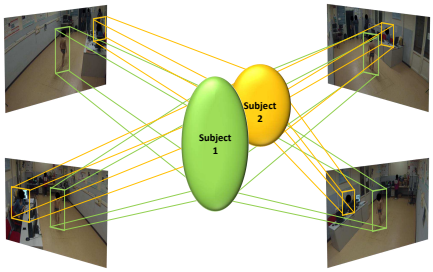
For each image, a person detector extracts the boxes. The model used is the Faster R-CNN developed by [26] in which the main idea is to merge a Region Proposal Network (RPN) and Fast R-CNN into a single network by sharing their convolutional features in a sort of attention mechanism; the RPN component tells the unified network where to look. The authors describe the RPN as a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are then used by Fast R-CNN for detection [26].

Multiperson

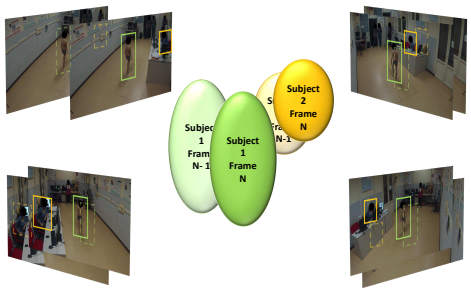
Multiple subjects could appear in the scene and the person detector could fail to retrieve the desired box if the selection is based only on thresholds. Indeed, the desired output of the person detector is the region of interest associated to the subject. The proposed procedure to select the boxes is based on the assumption that for the number of people in the scene from each point of view is fixed and known, for each gait cycle. The procedure starts with the selection of a maximum number of boxes, i.e., the number of expected subjects in the scene. Then, for each instant, the retrieved boxes from different points of view are matched to the same subject with the algorithm described in the following row. Finally, frame by frame the subject is tracked matching the set of boxes with the previous one considering the distance of the box with respect to the ones in the previous frame.

(a) For each view, the person detector return the bounding boxes that are recognized as people (score > 0.5).

(b) The boxes are matched: the bounding box represented with the same color are referring to the same subject in the real world.



(c) The boxes that are bounding a subject seen from all the points of view are kept for the next processing steps.



(d) The selected boxes are tracked across the different time instants. The spatial movement of the subject in the scene is retrieved.

Figure 2.6: Matching and tracking of the boxes across the different frames.

Algorithm for selecting multiperson:

- set how many boxes (n_1,n_2,n_3,n_4) to detect from each point of view. For each box, you obtain the four corner positions.
- the combinations of different boxes are tested and ordered according to the Learned Perceptual Image Patch Similarity (LPIPS) of the set of boxes.
- in the first frame, the best S sets of boxes are kept, where S corresponds to the number of subjects of interest in the scene. They represent the number of subjects to process and are assumed to be at the center of the scene or at least their body is entirely visible from all the cameras. This assumption is not limiting the system possibilities as the need to have the subject in the scene is necessary for the keypoints estimation.
- in the subsequent frames, the sets of boxes that correspond to the same subjects are matched to the first sets choosing the best Intersection over Union (IoU) score between the combinations of past (saved) and current (detected by person detector) boxes.

2.4.3 Pose estimation

Train-val-test dataset

The dataset used for training was obtained after the processing described in Sec. 2.3.

Subject	Gait cycles	Markers presence	Num frames	Dataset	Percentage of total dataset
Sub1	6	yes	756	Training	73.1%
Sub2	6	yes	676		
	6	no	636		
Sub3	6	yes	668		
Sub4	6	yes	784		
	6	no	648	Validation	13.5%
Sub5	6	no	644	Test	13.4%

Table 2.2: Description of dataset for pose estimator training, validation and test.

Data referred to subject 5 were kept to test the model generalization capability on a completely new subject without markers. Nevertheless, the prediction without markers is the target result.

Architecture

The architecture chosen for the pose estimation is a modified version of HRNet which was originally designed by [27]. The authors focused on the creation of multi-scale features fusion through the combination of parallel high-to-low resolution subnetworks that exchange information across them.

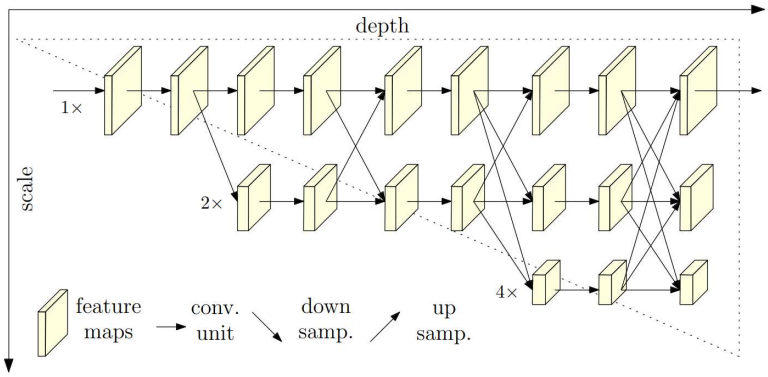


Figure 2.7: Image from [27] illustrating HRNet architecture. The horizontal direction correspond to the depth of the network and the vertical is the scale of the feature map.

This supervised model requires as input images, a bounding box for each person considered by the model - if any - and annotations about the keypoints positions- during training. Inside this region of interest, the joints are computed as a heatmap for each joint which indicates the

probability map for the position of the joint. The final prediction of the joint position is computed as the location of the maximum in the heatmap summing a quarter of a pixel in the direction of the highest adjacent one. The estimation approach rely on the use of multi-resolution subnetworks in parallel and in [27] the authors claim that the model demonstrated a good capability in predicting the human pose even in challenging environments. Armed with these results, the choice was to keep this architecture and to train it with our dataset.

2.4.4 Triangulation

The pose estimation step outputs the 2D positions of the keypoints from the camera's point of view. The joints' angles computation requires the markers' positions in the 3D space, thus the information from the different points of view is combined in the triangulation.

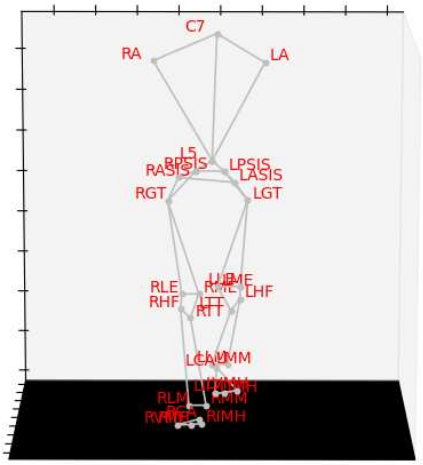


Figure 2.8: Estimated landmarks corresponding to the IOR-gait markers.

The algorithm chosen is the linear least squares triangulation, described for the case of two points and then generalized to more perspectives. We consider the pinhole model in which the projection p on the image plane of the point P is defined as:

$$p' = MP = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} P = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix} = K \begin{bmatrix} I & 0 \end{bmatrix} P \tag{2.1}$$

where α and β are the coefficient to obtain the coordinates on the axes x and y. In the classical pinhole model formulation they are obtained multiplying the focal distance f by a scale factor specific for the axis. While c_x and c_y are the translation value to be considered in the model when the principal point is not centered in the origin. More details about the pinhole model

formulation can be found in [28]. Considering the triangulation problem with two views, there are two cameras with known camera intrinsic parameters described by the camera matrices K and K' respectively. The relative orientations R and offsets T of these cameras with respect to each other are also known. Ideally, there should exist a point P in the three-dimensional space which corresponds in the images of the two cameras to p and p' respectively. Although the location of P is unknown, the exact locations of p and p' in the images could permit to find the exact location of P because K, K', R, T are known, thus the two lines of sight l and l', which are defined by the camera centers O, O' and the image locations p, p'. Therefore, P can be computed as the intersection of l and l'.

In the practical real scenario, this formulation has to be adapted because the observations $p*$ and $p'*$ are noisy and the camera calibration parameters always have an error, thus finding the intersection point of l and l' is not straightforward as in most cases, it does not exist because the two lines may never intersect (Fig. 2.9)

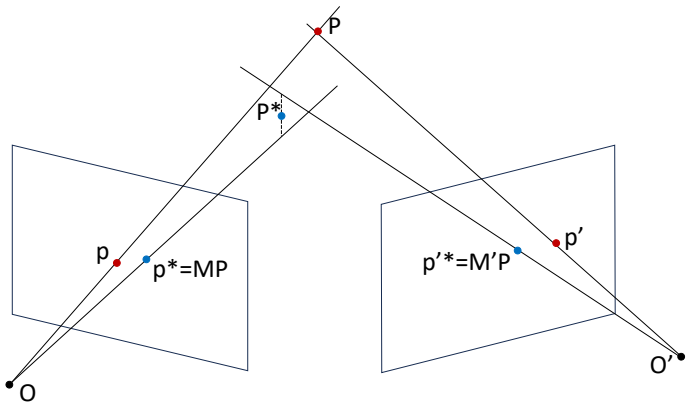


Figure 2.9: Least squares triangulation. P is the triangulated point in the ideal case. $P*$ is the least square estimate of P.

Considering the two points $p*$ and $p'*$ in the images that correspond to each other $p* = MP = (x, y, 1)$ and $p'* = M'P = (x', y', 1)$. By the definition of the cross product, $p * \times (MP) = 0$ thus, considering both images, we can create some constraints in the form of an equation:

$$AP = \begin{bmatrix} xM_3 - M_1 \\ yM_3 - M_2 \\ x'M'_3 - M'_1 \\ yM'_3 - M'_2 \end{bmatrix} P = 0 \tag{2.2}$$

where M_1, M_2, and M_3 are the rows of the matrix M. This equation can be solved with Singular Value Decomposition (SVD) and the homogeneous coordinates of the best estimate $P*$ of the point P is the eigenvector corresponding to the smallest eigenvalue. Considering that a greater number

of noisy lines is even more unlikely to intersect in a single point in the real case, the reasoning done for two images can be extended and the equations of the images appended to the matrix before SVD.

2.4.5 Angles estimation

The angles were estimated following the conventions adopted in [22]: the centers of the hip, knee and ankle joints are taken respectively as the head of the fibula FH, the midpoint between the epicondyles (LE, ME) and the mid point between the malleolus (LM, MM). Then, the anatomical reference frames for the body segments are defined as in [29].

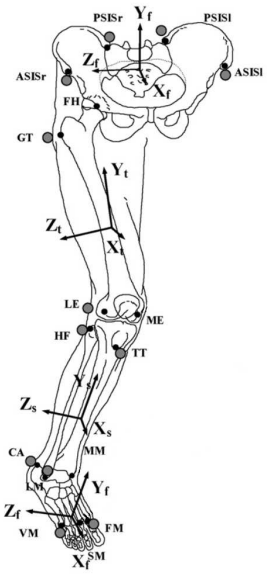


Figure 2.10: Reference frames definition. Image modified from [22].

For each joint the movement is measured considering the following three angles:

- flexion/extension: the relative rotation about the medio-lateral axis (Z) of the proximal segment
- internal/external rotation: the relative rotation about the vertical axis (Y) of the distal segment
- abduction/adduction: the relative rotation about a floating axis orthogonal to these two at each collected sample.

This terminology is adopted for the hip and knee joints, but for the special ankle joint these three rotations are referred to respectively as dorsiflexion-plantarflexion, inversion/eversion, and abduction/adduction. Spatial rotations of the pelvis, respectively tilt, rotation, and obliquity, are

calculated considering the virtual joint between the laboratory global frame as **proximal** and the pelvis as **distal** segments.

Each joint angle is computed for the three gait cycles of the corresponding body side, i.e., the right hip, knee and ankle angles are computed for the right gait cycles and viceversa for left side. Trunk and pelvis are computed using the GCs of the right side of the body, by convention.

2.5 Validation

To check the performance of the system and the errors induced by the various processing steps, we performed various analyses. First, we computed the error introduced by the projection step in data processing. To estimate this error, the projected points are compared to the manually annotated, where available.

Then, the prediction of the markers by our pose estimator is evaluated by comparing the estimated markers on the test subject to the manually annotated labels. The comparison was done for the different perspectives.

Finally, the joints' angles are compared with stereophotogrammetric ground truth obtained for the same subject during a sequential acquisition. Despite they are not exactly the same gait cycles, the walking patterns for healthy subjects are assumed to be similar in the acquisition happens near in time. Thus, each angle profile estimated over three GCs by our system is compared with the three GCs obtained from the stereophotogrammetric system.

2.5.1 Metrics description

The metrics used for the evaluation of DeepIOR results are the root mean squared distance (RMSD) and coefficient of multiple correlation (CMC)[30].

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \tag{2.3}$$

where N is the number of ordered samples in the sets x and y.

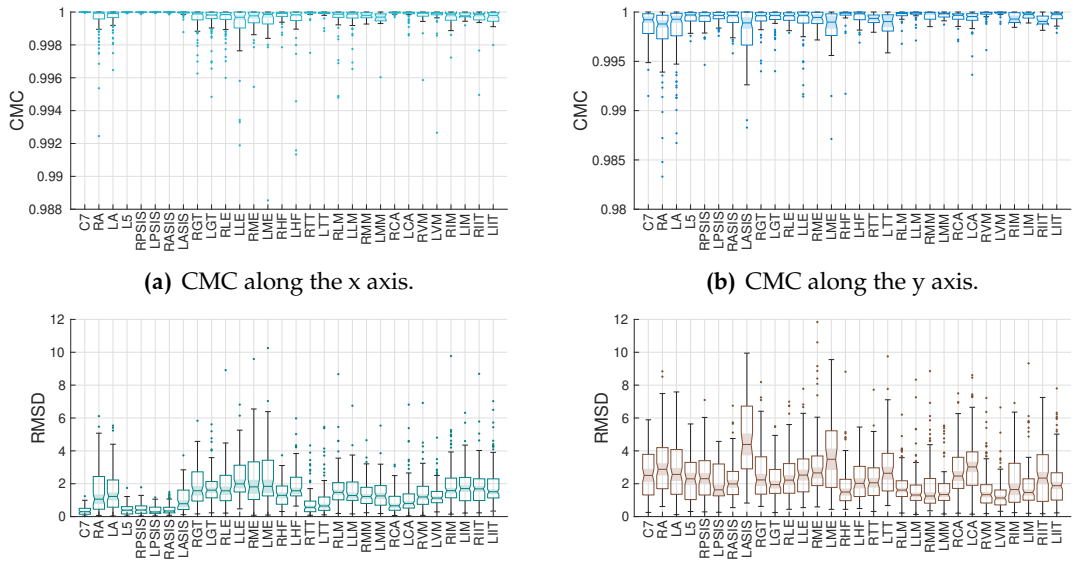
While CMC is an index that expresses the correlation between waveforms considering possible grouping. It is specifically designed to compare different protocols applied to the same gait cycles, for the definition refer to the original paper [30]. The threshold we set to consider two waveforms as correlated is 0.7.

2.6 Results

2.6.1 Projection error

In Fig. 2.12 are shown the results related to the projection step in the dataset preparation. The error is distinguished by considering the axes separately. For the whole dataset, the (x,y) coordinates

of the projected markers in the image are compared to the labels that were possibile to annotate manually. For each trajectory related to a gait cycle, it is obtained a value of RMSD and CMC that we reported in Fig. 2.11.



(a) CMC along the x axis.

(b) CMC along the y axis.

(c) RMSD along the x axis.

(d) RMSD along the y axis.

Figure 2.11: Estimation of the error due to projection step in data processing.

The metric CMC expresses a high correlation between all the compared trajectories. The average width of the taped markers is about 7-8 pixels and this can be assumed as the threshold of the imprecision that the operator can do while annotating manually, excluding errors due to operator distraction. Thus, we can notice that on average, the RMSD is not higher than this threshold, only few outliers overcome this value and the maximum is lower than 12 pixels. In general, we observe that the y coordinate presents an higher error with respect to x. A reason to explain this difference could be that the camera calibration obtained from Track on Field is more precise for the area around the feet thus the markers on the upper part of the body suffer of a greater triangulation error. Indeed, the main difference between x and y coordinates is observed for the first markers.

2.6.2 Markers prediction

The reconstruction of the three-dimensional model based on the markers' predictions is a crucial step. In Fig. 2.12, the test subject is shown with the estimated markers and also the ones obtained by video tracking and the other data processing steps. In Tab. 2.3, the average RMSD and its standard deviation over the test subject trials. It can be observed that the maximum average RMSD is approximately double of the error due to annotation, considering the average tape width of 7-8 pixels.

(a) **(b)**

Figure 2.12: Keypoints projected after triangulation compared with original labels obtained by video tracking and the others data processing steps.

2.6.3 Angles estimation

The main goal of DeepIOR is angles estimation.

In Fig. 2.14 the joints' angles estimated for the test subject are shown. Some patterns are captured by the system, e.g., the knee and the hip flexion-extension angle, but, in general, the angles profiles are not perfectly adherent to the stereophotogrammetric curves. However, the normality bands themselves show a great variability indicating that this type of data present an intrisic difficulty in its precise measurment and modeling.

In Tab. 2.4 are reported the metrics to evaluate the system performance. The RMSD column indicates the average root mean distance and its standard deviation considering the comparisons between the angles profiles obtained from the three gait cycles measured with stereophotogrammetry and the three analyzed with DeepIOR. The resulting number of comparison is 9. While for CMC there are reported the number of those comparisons that result in a score greater than 0.7.

RMSD [pixels]	avg	std
C7	6.68	0.93
RA	8.64	2.14
LA	10.17	2.05
L5	4.61	0.41
RPSIS	8.03	0.31
LPSIS	7.89	1.68
RASIS	10.83	3.81
LASIS	11.33	3.81
RGT	14.42	6.98
LGT	13.06	6.32
RLE	9.74	3.37
LLE	9.93	3.65
RME	7.75	2.19
LME	8.37	3.22
RHF	10.62	3.63
LHF	10.48	4.15
RTT	9.18	3.52
LTT	9.75	3.63
RLM	12.38	4.93
LLM	10.74	4.27
RMM	10.17	3.76
LMM	9.35	3.76
RCA	13.86	2.87
LCA	9.50	3.71
RVM	14.94	5.45
LVM	13.18	5.51
RIM	9.89	3.02
LIM	11.35	4.58
RIIT	11.49	4.20
LIIT	12.53	4.72

Table 2.3: RMSD between the dataset labels and the estimated keypoints averaged over the gait cycles of the test subject.

2.6.4 Comparison with baseline

In this section, we compare the results from DeepIOR to a baseline markerless method in literature [5]. The method belongs to the visual hull reconstruction category. The results in [5] are expressed as average and standard deviation of the RMSD between the angles computed on the same trials with both markerless and marker-based techniques. This is a difference with respect to our method that necessary has to compare the angles obtained by our system with different trials stereophotogrammetry. Moreover, the results are averaged for the two sides considering that we are operating with healthy subjects.

It is interesting to notice that besides the value of knee flexion-extension angle, the average RMSD values are not very different for the two methods. Considering that the stereophotogram-

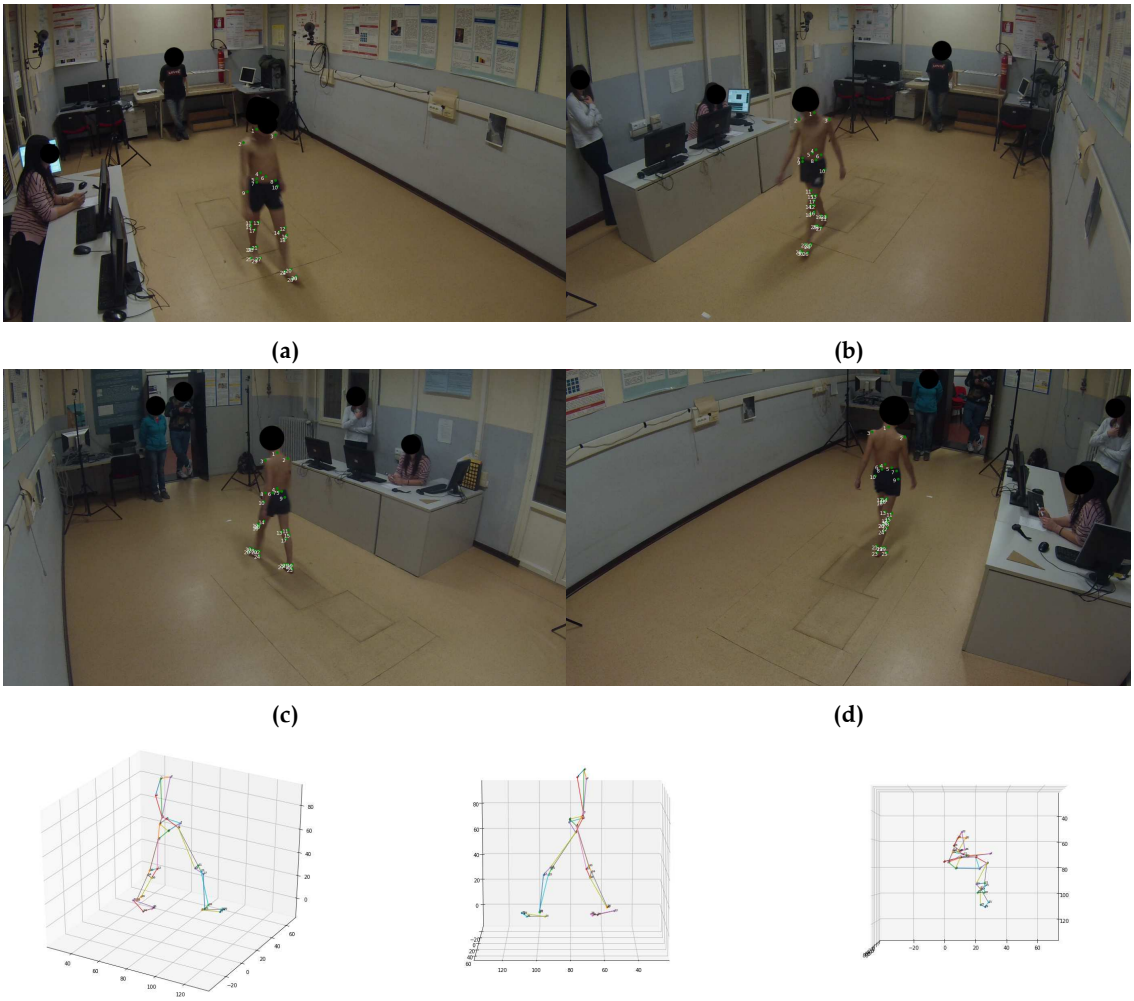
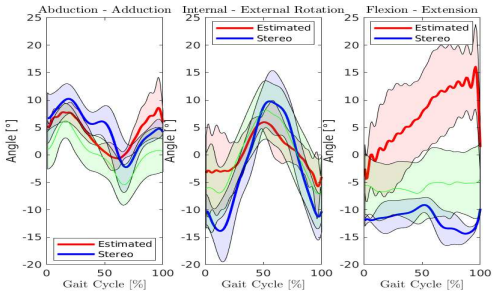
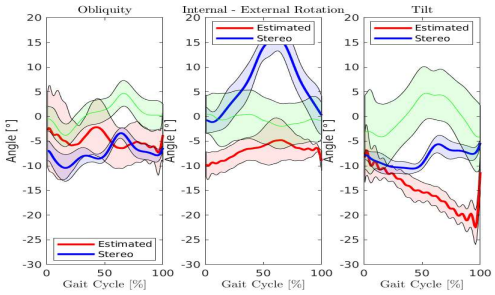


Figure 2.13: The markers estimated for each camera and the three-dimensional reconstruction at the same instant. The numbering of the markers in the images is according with the IOR gait protocol.

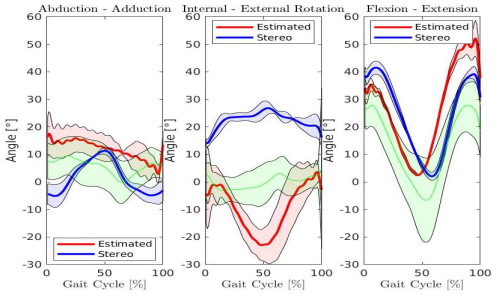
metry is the method for the gait analysis in the clinical domain, the results suggest that our method move a step in the correct direction for the establishment of a robust markerless method. However, the RMSD error is higher than the marker-based method, thus our approach still requires more precise dataset for the models training to obtain more reliable results.



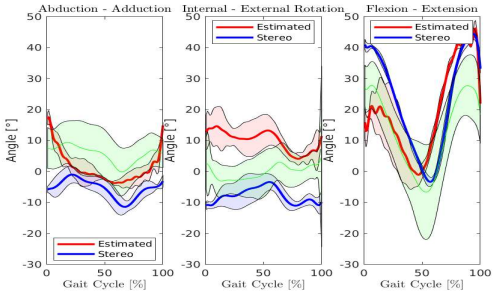
(a) Trunk



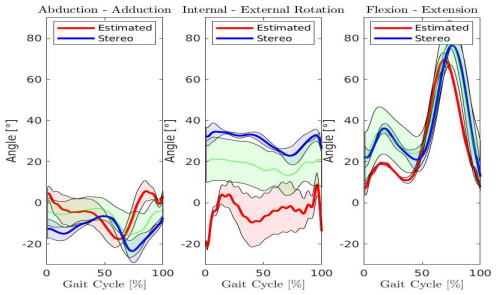
(b) Pelvis



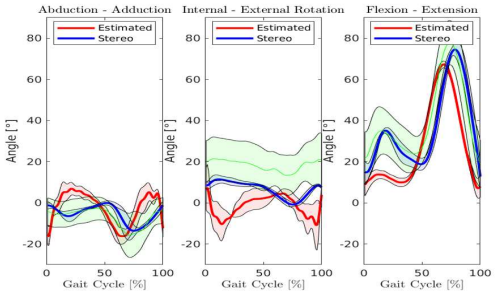
(c) Right hip



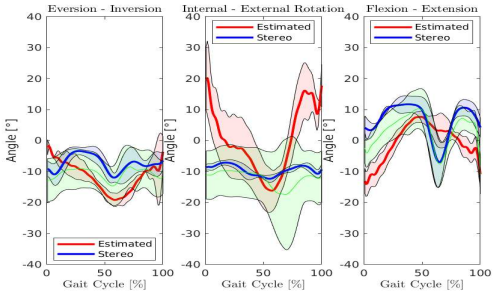
(d) Left hip



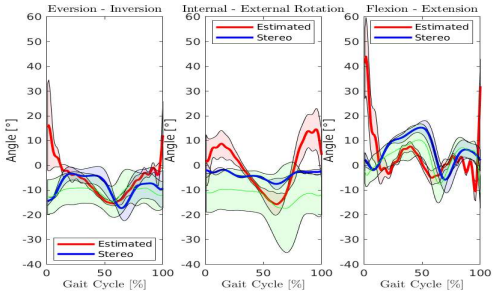
(e) Right knee



(f) Left knee



(g) Right ankle



(h) Left ankle

Figure 2.14: The estimated joint angles profiles (red) are compared to the ones obtained with stereophotogrammetry (blue). In green are shown the normality bands for children.

Metrics	RMSD in degree(°) - avg ± std			CMC > 0.7 (number of comparisons)		
Angle types	Abduction Adduction	Internal external rotation	Flexion Extension	Abduction Adduction	Internal external rotation	Flexion Extension
Trunk	5.43 ± 2.07	6.73 ± 2.44	11.01 ± 2.03	2	1	1
Right hip	9.51 ± 2.46	18.36 ± 3.83	26.64 ± 13.37	1	1	1
Left hip	5.47 ± 2.46	9.59 ± 1.39	23.19 ± 13.31	0	1	1
Right knee	11.31 ± 3.34	22.49 ± 2.57	34.88 ± 18.25	2	1	1
Left knee	6.78 ± 3.60	6.67 ± 2.57	33.39 ± 17.70	0	1	1
Right ankle	10.11 ± 2.99	10.57 ± 2.43	7.56 ± 1.63	0	1	1
Left ankle	8.72 ± 3.88	7.05 ± 2.74	8.03 ± 5.27	1	1	1
Pelvis	7.00 ± 1.08	8.08 ± 2.13	12.38 ± 2.05	1	1	1

Table 2.4: The metrics to evaluate the system performance are reported. RMSD is indicated as average and standard deviation, while for CMC are reported the number of comparisons that result in a score greater than 0.7.

Metrics	RMSD (°) - avg ± std					
	DeepIOR			Visual hull baseline		
Angle types	Abduction Adduction	Internal external rotation	Flexion Extension	Abduction Adduction	Internal external rotation	Flexion Extension
Trunk	5.43 ± 2.07	6.73 ± 2.44	11.01 ± 2.03	NA	NA	NA
Right hip	9.51 ± 2.46	18.36 ± 3.83	26.64 ± 13.37	14.1 ± 2.3	21.6 ± 9.3	17.6 ± 3.5
Left hip	5.47 ± 2.46	9.59 ± 1.39	23.19 ± 13.31			
Right knee	11.31 ± 3.34	22.49 ± 2.57	34.88 ± 18.25	NA	NA	11.8 ± 2.5
Left knee	6.78 ± 3.60	6.67 ± 2.57	33.39 ± 17.70			
Right ankle	10.11 ± 2.99	10.57 ± 2.43	7.56 ± 1.63	7.0 ± 3.6	12.9 ± 7.0	7.2 ± 1.8
Left ankle	8.72 ± 3.88	7.05 ± 2.74	8.03 ± 5.27			
Pelvis	7.00 ± 1.08	8.08 ± 2.13	12.38 ± 2.05	NA	NA	NA

Table 2.5: Comparison with visual hull markerless [5] results.

2.7 Conclusions and future directions

The possibility of quantifying the complete range of joints' motion in a precise way with markerless techniques is still a challenge. Despite the results we obtained show a good reconstruction of the subject 3D model, the angles estimation is still really noisy and not adherent to the stereo ground truth.

The main limitations we faced during this research were related to the creation of the input dataset because of the burden of the manual annotation which is time consuming. Thus, this results in a low amount of data which means low precision in the context of the deep-learning. Another implication of manual annotation is the great amount of imprecision and human errors that make the initial dataset not completely reliable for this task.

Our work proved that a deep neural network can be trained to estimate superficial landmarks instead of the more typical approach in literature in which are estimated the joints' center. This is more clinically relevant because permits to build the joints' 3D reference frames and to compute the anatomical angles.

Moreover, this approach can be adapted to any different clinical protocol training the pose estimator with a different dataset. Indeed, the high degree of modularity is a strong point of the proposed solution. On the other hand, the results suggest that there is still some work to be done, in particular, we observed that the limitations were mainly due to the original input data and its labeling. Thus, future research can focus on the production of a precise dataset using an automatic labeling strategy, e.g., using a marker-based system equipped with autolabeling. Another approach can consider different types of input data as radar data, as in our more recent research project discussed in Chapter 2.

3

Radar-based markerless motion capture

3.1 Introduction

The fundamental concept of radar sensing is the exploitation of communication disturbances to extract information about the environment used for the communication exploiting multiple copies of the signal received due to the multipath effect. The recent advancements in telecommunications are in the direction of the use of high frequency transmissions to increase communication speed and reliability. This feature also permits us to sense the environment with much more precision. The fields for the application of such capabilities are multiple and very different. In this project, we applied radar sensing to the clinical and biomedical field for the development of a markerless motion capture system. As widely discussed in Chapter 1, the markerless approach presents various advantages but it is still a challenge due to the technical difficulties of predicting the pose in a complete way for a clinical gait analysis. Different technological solutions have been proposed including cameras, accelerometers, and also some recent mmWave radar solutions which can offer numerous advantages with respect to the other cited technologies. A markerless radar-based is portable, so not specifically related to an environment and calibration procedures, also because radar data are not affected by illumination conditions. Moreover, privacy is preserved because the extraction of personal information is less direct than, for instance, from an RGB video. On the other hand, there are still some challenges that require some research effort from both the hardware and especially the data processing point-of-view. In fact, the reflected signals are heavily influenced by noise and unwanted reflections by external sources which can degrade the results. As the required accuracy on the localization of the body parts is very high, i.e., in the order of a few millimeters, the content of this chapter is devoted to the development of a solution to the markerless motion capture problem using a combination of radar sensing and deep-learning techniques. In particular, we contributed to the development of a complete pipeline to enable an end-to-end training of deep-learning models from raw mmWave radar data.

The data collection itself is a process that is not well-established in the literature and the set-up definition is of paramount importance for any further research. A first version of this project was presented in [31].

3.1.1 Frequency Modulated Continous Wave (FMCW) radars

Radars are electronic devices built to emit known electromagnetic signals which are reflected by the environmental obstacles and received delayed, shifted, attenuated, and distorted due to reflections, scattering, and, in general, noise. From the comparison between the transmitted and the received signals it is possible to compute properties of interest of the objects encountered by the signals that are the radar targets, usually the interesting quantities are position and velocity.

The classification of radars includes differences in positioning of TX and RX antennas. Monostatic is the term to define radars with both types of antennas on the same device, while if the antennas are physically placed in a different place the term is bistatic. Another criterium for radar classification is related to the transmitted signal type. For the following description we consider a monostatic FMCW radar. The Frequency Modulated Continuous Wave (FMWC) radars transmit sinusoidal waves whose frequency increases linearly during the pulse. Those pulses are called chirps and are defined by the starting frequency f_c, the bandwidth B computed as $B = f_m - f_c$ with f_m the maximum frequency and the duration T_c. The slope of the chirp S is computed as:

$$S = \frac{B}{T_c} \tag{3.1}$$

Thus, the instantaneous frequency can be computed as:

$$f(t) = f_c + St \tag{3.2}$$

with $0 \leq t \leq T$

Being the phase of the transmitted signal related to the instantaneous frequency by the following relation

$$\varphi(t) = \int_0^t f(t')\, dt' = 2\pi \left(f_c t + \frac{S}{2} t^2 \right) \tag{3.3}$$

it is possible to express the transmitted signal as

$$s(t) = \exp(j\varphi(t)) = \exp\left[j2\pi \left(f_c + \frac{S}{2} t \right) t \right] \tag{3.4}$$

The transmission of a sequence of N equally spaced chirps in time with duration is NT_c is denoted as frame. At the receiver, a mixer combines the received signal (RX) with the one transmitted, generating the Intermediate Frequency (IF) signal which is a sinusoid whose instantaneous frequency corresponds to the difference between those of the TX and RX signals. Each chirp is sampled with sampling period T_f (referred to as fast time sampling) obtaining P points, while N samples, one per chirp from adjacent chirps, are taken with period T_c (slow time sampling).

The slow time index p simply corresponds to the chirp number. On the other hand, the fast time index n assumes that for each pulse, the corresponding continuous beat signal is sampled with frequency f_s to collect N samples within the time duration T. The slow time is usually still faster than the movement of the targets thus each frame is associated to a single static capture of the environment at time t.

Range and velocity estimation

Considering a reference antenna, it is possible to extract the target range and velocity with respect to the radar. Infact, the signal reflected by a target is an attenuated version of the transmitted waveform with a delay τ that depends both on the distance between the target and the radar and on their relative radial velocity.

The time it takes a transmitted signal to reach a target object at distance r and, in general, in relative movement with respect to the radar, the signal reflected back is:

$$\tau = \frac{2(r + vt)}{c} \tag{3.5}$$

which is the Round Trip Time (RTT) for a given distance r, assuming that the velocity of propagation is the speed of light c and the target velocity is v.

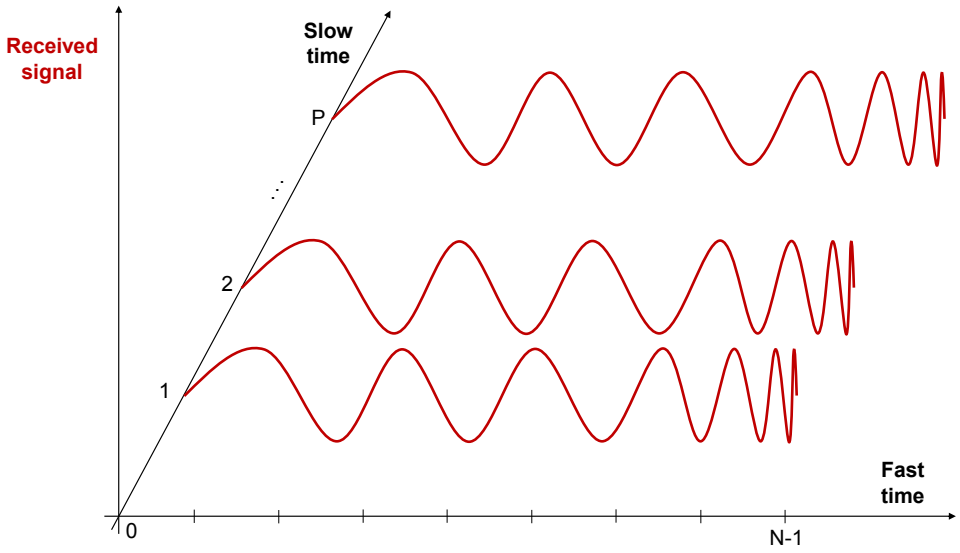


Figure 3.1: Received signal amplitude in FMCW radar.

A the receiver the signal is mixed and sampled, then assuming single target and neglecting reflected signal distortions, the FMCW radar receiver output is a function of p and n time indices [32]

$$y(n,p) = \alpha \text{exp}[j\varphi_{IF}(n,p)] + w(n,p) \tag{3.6}$$

where p and n are the sampling indices along the slow and fast time, respectively, α is a coefficient accounting for the attenuation effects due to the antenna gains, path loss and Radar Cross Section (RCS) of the target and $w(n,p)$ is additive white gaussian noise term with zero mean.

Introducing the quantities Doppler frequency $f_d = \frac{2f_c v}{c}$ and beat frequency $f_b = \frac{2Kr}{c}$, we obtain

$$y(n,p) = \alpha \text{exp } j2\pi \left[(f_b + f_d) \frac{n}{f_s} + f_d p T_c + \frac{2f_c R}{c} \right] + w(n,p) \tag{3.7}$$

From this sample, it is possible to obtain the radar Range Doppler (RD) map, i.e. the matrix containing all the information provided by a single antenna for a given time frame. This can be obtained creating the matrix N x P with the samples y(n,p) and, by applying a bi-dimensional along the fast time and slow time dimensions, the frequency shifts of interest can be extracted and they contain the range and velocity of each reflector. After the DFT step, the RD is obtained by taking the square magnitude of each obtained complex value.

The reflections are still noisy due to the background contribution, source of interference and the fact that each target give multiple reflections. To select the most significant reflections is applyied the Cell-Averaging Constant False Alarm Rate (CA-CFAR) algorithm. CA-CFAR is described in [33] and the main idea is the decision of a dynamic threshold on the power spectrum of the output signal retaining only the points that exhibit a significantly higher reflected power compared to their local background considering each point of the RD map, called bin. Furthermore, to filter out reflections from static objects, a high-pass filter known as Moving Target Indication (MTI), as detailed in [33], is employed. MTI eliminates reflections with Doppler frequencies close to zero, ensuring that the points likely to be reflected by the static background are not considered, while the human subject is characterized by rapid Doppler Frequency variations during motion [34]. The threshold we set to defined a reflection close to zero is $\pm$ 0.05 m/s as it is a low velocity probably not related to a walking subject.

The RD map obtained after these steps is sparser and along the fast time it is possible to retrieve the frequency of the IF signal $f_d + f_b \sim f_b$, while the peak along the slow time reveals the Doppler frequency f_d. So for each reflector preserved after the filtering steps, the range r and velocity v can be expressed as

$$r = \frac{f_b c}{2S} \tag{3.8}$$

$$v = \frac{f_d c}{2f_c} \tag{3.9}$$

Estimation of azimuth and elevation

In a FMCW radar, the single antenna can be used to to determine the distance and the velocity of a target object (Sec. 3.1.1). The spatial diversity due to multiple antennas allows to compute the Angle of Arrival (AoA) of the received signals. The design of the antennas arrays is crucial to

determine the capability of computing both elevation (EL) and azimuth (AZ) angles, infact, the quantity of interest for the computation is the phase shift due to the different distance to the target. If we assume the distance between two subsequent antennas along both azimuth and elevation dimensions is the same and we denote it as d, the phase shifts along the two dimensions ψ_{AZ} and ψ_{EL} can be expressed as:

$$\psi_{EL} \approx \frac{2\pi f_c}{c} dsin\phi \tag{3.10}$$

$$\psi_{AZ} \approx \frac{2\pi f_c}{c} dcos\theta \tag{3.11}$$

where θ and ϕ are the AZ and EL angles of a reflecting point. And to compute the phase shift values, an approach similar to the one used for range and velocity can be used. The peak positions aterf the DFT across the samples taken at the azimuth and elevation antennas are the desired values.

Point cloud cartesian coordinates

Using the derived quantities for range, velocity and AoA, the cartesian coordinates for each reflector can be expressed as

$$x = rcos\theta = r\frac{\psi_{AZ}c}{2\pi d f_c} \tag{3.12}$$

$$y = \sqrt{r^2 - x^2 - z^2} \tag{3.13}$$

$$z = rsin\phi = r\frac{\psi_{EL}c}{2\pi d f_c} \tag{3.14}$$

3.2 Related works

The authors of [35] explore different methodologies that can be exploited to predict the pose using a point cloud as input, including templates, features, and machine learning approaches. They consider depth point clouds, but some similar conditions and features can be found also in radar point clouds. In [36] instead, the authors focus mainly on the advancements in neural networks for point clouds datasets. On the other hand, the potential of the mmWave FMCW radar has started to be understood for the pose estimation task. In [37], a CNN is used to perform skeleton estimation with a mmWave FMCW radar by converting the 3D spatial coordinates of the point clouds to 2D heatmaps. This method does not fully exploit the 3D positions of the points with the projection step and this could be acceptable for some applications, but not clinically. For the same reason, the prediction of the joints' centers only is not sufficient. Moreover, the time correlations obtained from a sequence of point clouds are not considered as a source of further information. In [38] the time correlation of the sequences is used and it helps in the prediction, however, it is not considering a biomedical perspective that requires a 3D reference frame for each joint as well

as others recently published works [39], [40]. However, the results in [38] inspired our use of the recurrent layers to exploit the temporal information.

3.2.1 Pointnet++ Architecture

Pointnet++, also referred to as pointnet2, is the object of the the work done in [41] and it is an extension of the older Pointnet network [42] to include a hierarchical structure. The general idea of PointNet++ is to partition the set of points into overlapping local regions. Then, the model extracts local features that are grouped into larger units and processed to produce higher level features. The processing is similar to the CNN functioning but it is tailored for non-uniform spaces. This process is repeated iteratively to learn a set of features that describe the spatial structure of a point cloud at different resolutions. In [43] the authors present PointRNN which consists of a RNN that includes the feature extraction technique of Pointnet++ with the aim to predict future point clouds.

In the following, some more details about the Pointnet++ special layers are presented.

Set Abstraction layer

The Set Abstraction (SA) layer is used to extract the features from the input point set at a given scale. Given a point cloud C with optional features f, N centroids, and a radius R, the SA layer computes the set S of N centroids. The algorithm used for the choice of centroids choice is the Farthest Point Sampling (FPS). FPS is an iterative algorithm used to extract a subset S of points from a point cloud C. In the space exists a distance function d exists such that the sampled points are as far as possible among each other, according to d, so that C is a good approximation for S with a fixed number of points. At the beginning, the first point is selected from C and added to S, then it is iteratively computed the distance between each point in the cloud C and its nearest point in the subset S. Based on this the point in C with the maximum distance is added it to S and the procedure is repeated until the subset reaches the desired number of points.

A certain degree of randomness is implicit in the choice of the first point to add to S, thus the final outcome depends on the criterium used to choose it, e.g. randomly sampled from a uniform distribution. Then, the partitioning at the SA layers is obtained with a modified version of the Ball Query (BQ) algorithm, i.e., given a point cloud, a set of centroids, and a radius, returns all the points from the cloud that are contained in a sphere with a given radius centered in each centroid. The modification by the authors is to limit the maximum number of points to associate with each centroid. Each partition is passed to a convolutional layer with a kernel of unitary size, followed by an optional Batch Normalization (BN) and, then, they can be used as input to other Multi Layer Perceptron (MLP) layers with a different number of filters and the final result is obtained after a max pooling. The output of the abstraction steps is a set of features. From the learning point of view, our problem is the reconstruction of a scaled version of the original dataset; the reconstruction role is played by the Feature Propagation (FP) layers.

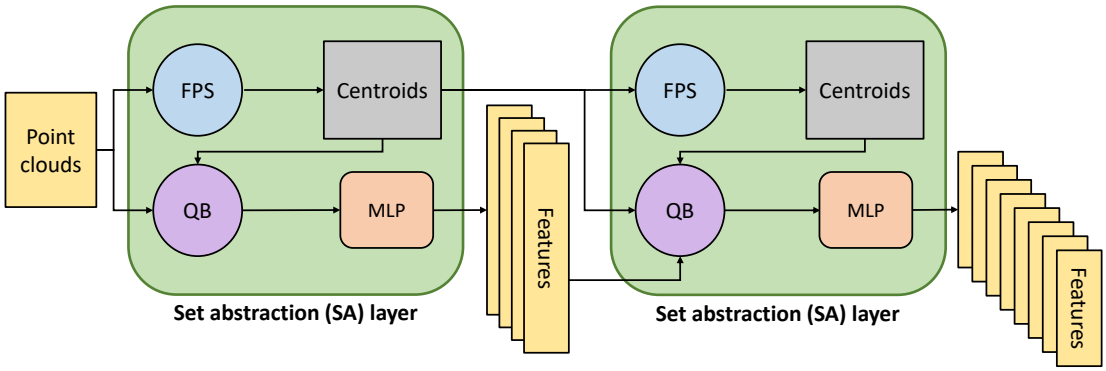


Figure 3.2: Schematic of the Set Abstraction layer architecture.

Grouping Strategy

In general, point clouds present different density distributions in the diverse areas. To effectively capture the cloud structure it is important to extract features that capture both the details and the wider structures. One of the main contributions of Pointnet++ is its hierarchical architecture that allows the network to adapt the learning resolution depending on the density of the points. Point clouds, especially if obtained with mmWave radars, contain spaces where the concentration of points is high and others where it is low, which means that the point clouds have usually different densities in different areas. The authors in [41] propose two different approaches that we briefly describe in the following lines. The Multi-Scale Grouping (MSG) approach is based on the concatenation of features at different scales to create a multi-scale feature. To extract them, the input set C and the number of centroids N are used as input to multiple SA layers with different radius $r = [r_0, ..., r_s]$ and the final features are concatenated. This approach runs local Pointnet at large scale neighborhoods for every centroid thus it has a high computational burden. The other approach is the Multi-Resolution Grouping (MRG) in which the features vector at the level L_i is obtained by concatenating the features at each subregion from the lower level L_{i-1} using the SA layer and the ones computed processing all raw points in the local region using a single pointnet.

Feature Propagation Layer

The FP layer allows propagating features from the subsampled layer to an upper scaled version of the features, until the reconstruction of the original set. The new interpolated features are ob-

tained through the application of an inverse distance weighted average interpolation of k nearest neighbors and then they are passed to a unit pointnet. In Fig. 3.3 is shown the structure of the layer.

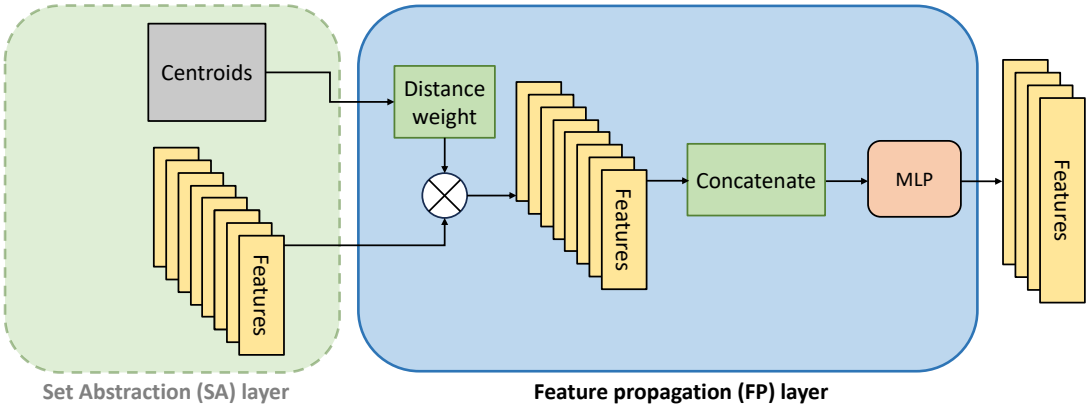


Figure 3.3: Schematic of the Feature Propagation layer architecture.

3.3 mmIOR

The proposed solution mmIOR includes a processing framework that exploits radar sensing and deep-learning techniques for the prediction of the superficial landmarks required by the IOR-gait biomedical protocol [44]. The first version of our project was presented in [31].

The goal is to estimate the set of superficial points defined by the IOR-gait protocol [44] receiving as input a sequence of point clouds derived from radar data. The ground truth to train our model was obtained through the use of a marker-based motion capture system. Contrary to the majority of the current research on this topic that employs public datasets and tries to identify the joints' centers, we aim to predict the position of the protocol markers. This permits to build a 3D reference frame for each joint, as described in Chapter 1. The prediction of the marker's coordinates is obtained by adapting the neural network architecture PointNet++ [41] to receive as input the point clouds obtained from the processing of radar raw data. After data acquisition, raw data are processed to extract point clouds. Then, the set of points belonging to the subject is identified by removing the noise introduced by other objects through clustering techniques. After that, the point clouds sequences are synchronized with the ground truth data from the motion capture system. The resulting dataset is used to train our model. In the following sections, these

steps are detailed.

3.3.1 Experimental set-up

The acquisition sessions of the datasets for this project were performed in the Biomov research laboratory in the Department of Information Engineering of the University of Padua. The main goal for the experimental set-up is the simultaneous acquisition of motion data and the corresponding mmWave radar reflections where the radar data are the input for our model while the motion capture trajectories are the ground truth for the supervised training of the system. Therefore, the acquisition system includes the AWRx Cascaded Radar RF mmWave (MMWCAS-RF-EVM) radar from Texas Instruments and a Vicon MX T-Series motion capture system.

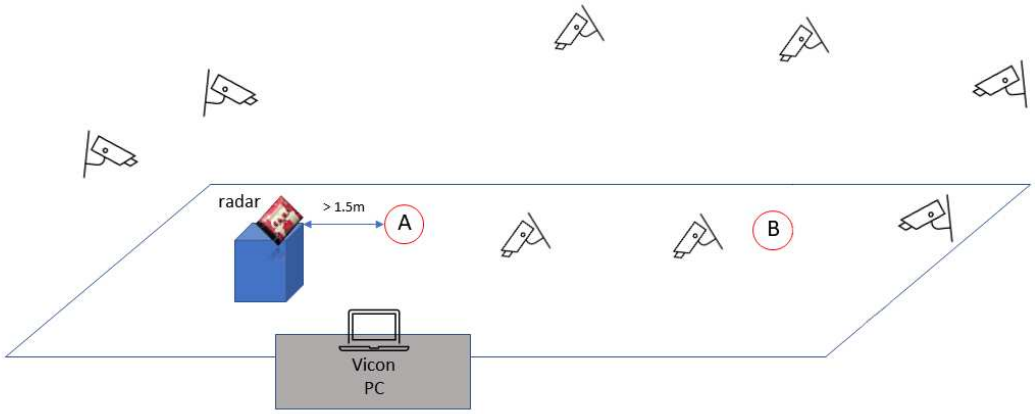


Figure 3.4: Set-up for data acquisition with 8 stereophotogrametric cameras.

Motion capture system

The motion capture system records the ground truth of the markers' trajectories while the subject is moving. The laboratory we used was equipped with six cameras placed on the two long walls at the first session of data acquisition, while the other datasets were collected with eight cameras. The markers' positioning was chosen accordingly with the IOR-gait protocol [44] designed specifically for the computation of the joints' angles and to describe the subject movement with the minimum number of markers.

Before the acquisition phase, the software used to control the motion capture system requires a subject model that represents the relative positions of the markers with respect to each other. The version of the motion capture system we used is equipped with auto labeling, thus after the acquisition is performed, the software fits the subject model on the visible markers and provides a label for each marker it was possible to associate with the model.

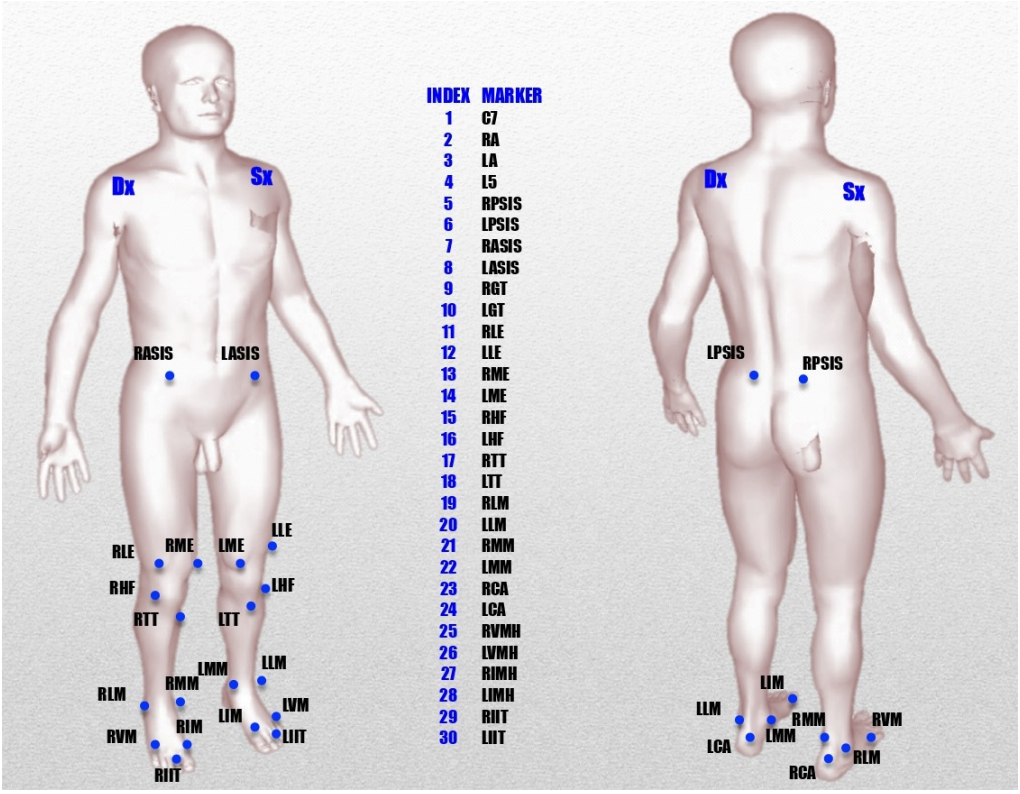


Figure 3.5: IOR-gait protocol.

Due to the positioning on the subject body, some markers are more difficult to track, in particular those in the inner side of the legs (Fig. 3.5). As expected, we noticed that a higher number of cameras usually permits a better reconstruction.

Radar Set-up

The radar used for our project is the AWRx Cascaded Radar RF Evaluation Module (MMWCAS-RF-EVM) from Texas Instruments. It is equipped with four AWR2243P radar chips, and each chip includes 4 RX and 3 TX antennas for a total of 16 RX and 12 TX antennas per chip. The range for the radar operating frequency is from 76 GHz to 81 GHz and for this project, we chose one configuration among the possible ones, thus we used all antennas with a MIMO setup resulting a total of 86 non-overlapping virtual antennas.

The radar placement in the laboratory set-up was along the short wall, thus recording the front or the back of the subject while walking. For the walking area definition, it has to be considered the field of view of the radar both in azimuth and elevation limiting the minimum distance between the subject and the radar to prevent the reflection to represent only part of the subject's body. We set this limit at 1.5 m and the radar was positioned on a table at about 80 cm of height.

We acquired gait trials of different subjects. The first session we acquired a total of 100 se-

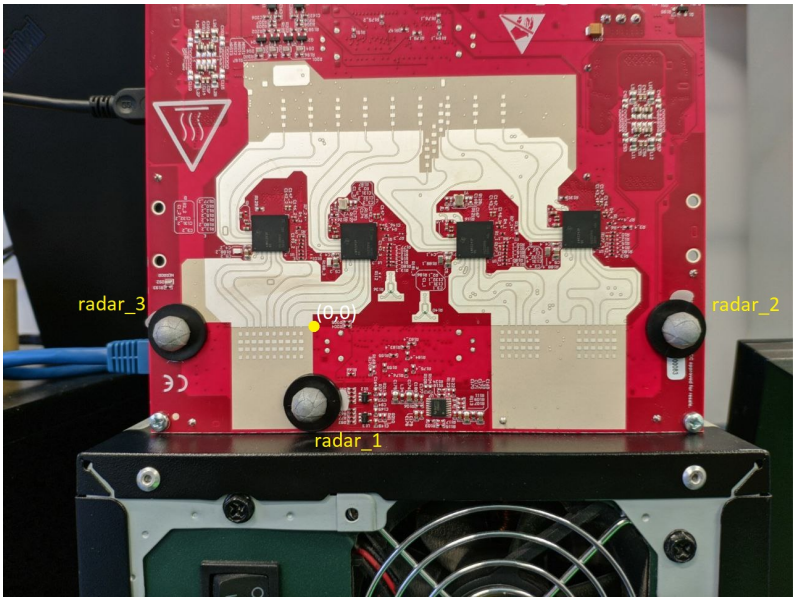


Figure 3.6: MMWCAS-RF-EVM radar during data acquisition. The markers are used to create the radar reference system with respect to the stereophotogrametric cameras.

quences of 10 seconds. However, in this first data collection session, the actual usable sequences in which at least one marker is visible amount to around 8 minutes of acquisitions due to limited number of available Vicon cameras. Moreover, some data were discarded due to the limited field-of-view of the radar that we did not consider in the first data acquisition.

Then, with the set-up with 8 cameras, we performed about 160 sequences of 10 seconds in which the markers reconstruction is visibly more accurate but we do not include them in the training and the results are not considering this parts of data because there are some problems to be solved in the synchronization between radar and stereo data (Sec. 3.3.2).

Parameter	**Value**
Chirp duration (T_c)	81 μs
Chirps per frame (N_c)	64
Samples per chirp (N_s)	512
Sampling frequency (F_s)	8 MHz
Starting frequency (f_c)	76 GHz
Bandwidth	4.16 GHz
FOV Azimuth	$\pm\frac{\pi}{3}$
FOV Elevation	$\pm\frac{\pi}{6}$

Table 3.1: Parameters for radar set-up.

The parameters were chosen to allow a correct acquisition of the distance and velocity ranges typical of a subject walking in the available laboratory area. Both the range of values and the

resolutions were considered. With these parameters, the maximum range is

$$r_{max} = \frac{F_s c}{2S} = 23.36 \text{ m} \tag{3.15}$$

with a resolution in range of

$$\Delta r = \frac{c}{2B} = 0.036 \text{ m} \tag{3.16}$$

While the maximum detectable velocity for the frequencies considered is

$$v_{max} = \frac{c}{4 f_c T_c} = 12.18 \text{ m/s} \tag{3.17}$$

The velocity resolution is

$$\Delta v = \frac{c}{2 f_c N_c T_c} = 0.38 \text{ m/s} \tag{3.18}$$

3.3.2 Data pre-processing

Points cloud creation

The use of an FMCW radar in a MIMO configuration enables the estimation of three-dimensional spatial coordinates for all the reflection points of the transmitted signal. For each reflector, the radar's data include the range, velocity, azimuth, and elevation angles of arrival, and the power of the reflections. Although the CFAR and MTI application is useful for static objects removal, a single target can result in a great amount of reflectors and the target's reflections are mixed with noise.

Clustering and tracking

The method applied here is taken from [45] and is briefly described in the following. The steps described in [45] include the application of a clustering algorithm to identify the points reflected by the different objects and exploiting the clustering stability along time the target can be distinguished with respect to the noise and background, then tracked in the sequence of frames.

Thus, first, the clusters detection happens frame by frame using the density-based detection approach named Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [46]. The main idea is to identify as clusters the subset of points which are sufficiently dense in a localized portion of the considered dimensions. The parameter that defines the density threshold is m_{pts}, i.e. the number of points contained in a sphere of radius ϵ. For each point, it is considered dense and part of a cluster if the number of points contained in the sphere of radius $epsilon$ centered in it contains at least m_{pts} points. In this project, the clustering is performed on x and y, the two dimensions that define the walking plane for the subject. DBSCAN has been successfully applied to points clouds in literature [47]–[49]. Moreover, this algorithm choice is justified by the capability of DBSCAN to correctly identify the clusters as long as the subjects are sufficiently separated [48]. In this case, the environment is controlled, there is only one subject, thus the clustering is

not a difficult problem, but the algorithm is unsupervised, thus easily applicable to a multiperson scenario in which the number of subjects is unknown or changes during the data acquisition.

The subject is detected as a cluster containing multiple points due to the extension of the subject's body with respect to the range resolution of the mmWave radar. The spatial extension can be ignored, i.e. the body is considered as a point reflector with no dimensions, or estimated as in the approach proposed in [45]. Given a cluster of points n at the time instant k, the true position of the subject is estimated computing the centroid. It is the weighted mean of the positions of the points using the normalized received power as weight. The dimensions considered are x and y, thus the points are defined as $\mathbf{p}_k^n = [x_k^n, y_k^n]^T$ and the centroid is

$$\mu_k^n = \sum P_k^{RX} \mathbf{p}_k^n \tag{3.19}$$

where P_k^{RX} is the received power normalized [0,1]. From the covariance matrix, Σ_k^n, it is possible to extract the axes lengths of the ellipse that estimate the extension of the target. We denote them as $\hat{l}_k^n$ and $\hat{w}_k^n$ and they can be computed as the norms of the eigenvectors of the covariance matrix. While the orientation of the ellipse $\hat{\xi}_k^n$ has the same direction of the eigenvector corresponding to the largest eigenvalue of Σ_k^n.

The tracking step employs a Converted-Measurements Kalman Filter (CM-KF) that estimates both the position of the targets in Cartesian coordinates and the extension of the subject in the clustering plane x-y [50], [51]. The state of the track t at the instant k contains different quantities of interest. It is defined as $s_k^t = [x_k^t, y_k^t, \dot{x}_k^t, \dot{y}_k^t, l_k^t, w_k^t, \xi_k^t]^T$ where (x_k^t, y_k^t) is the target position, $(\dot{x}_k^t, \dot{y}_k^t)$ are the velocity components, (l_k^t, w_k^t) represent the extension and ξ_k^t is the orientation angle. Each track t is then defined as a tuple, $T_k^t = [\hat{s}_k^t, P_k^t, Z_{k-K+1}^t : k, I_k^t]^T$ where $\hat{s}_k^t$ is the estimate of the current state with P_k^t the associated error covariance matrix as computed by the model, the collection of the last K clusters associated with the track, $Z_{k-K+1:k}^t$, and the integer I_k^t representing the estimated identity of the subject. While the observation vector contains the information for a detected cluster n at time k is defined as $z_n^k = [\mu_x^n, \mu_y^n, \hat{l}_k^n, \hat{w}_k^n, \hat{\xi}_k^n]$.

Thus, given the sequence of the collected measurements until the instant k, $z_{1:k}$, the CM-KF provide an estimate of the state of the track t at the instant k. The motion model is defined by the matrices **F** and **H**, where **F** is the transition matrix that gives the update of the state s_k starting from the previous one s_{k-1}, while **H** is the observation matrix, which relates the observation vector z_k to the state s_k. The system dynamic model can be defined as

$$s_k = \mathbf{F} s_{k-1} + u_k \tag{3.20}$$

$$z_k = \mathbf{H} s_k + r_k \tag{3.21}$$

where $u_k \sim \mathcal{N}(0, \mathbf{Q})$ and $r_k \sim \mathcal{N}(0, \mathbf{R}_k)$.

The matrices **F** and **H** are

$$\mathbf{F} = blkdiag \left[\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2, \mathbf{I}_3 \right] \tag{3.22}$$

where $blkdiag$ is the block diagonal matrix and $\mathbf{I}_n$ is an $n \times n$ identity matrix.

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times3} \\ \mathbf{0}_{3\times2} & \mathbf{0}_{3\times2} & \mathbf{I}_3 \end{bmatrix} \tag{3.23}$$

where $\mathbf{0}_{n\times m}$ is an $n \times m$ all-zero matrix and $\otimes$ refers to the Kronecker product between matrices.

With some assumptions about the noises, it is possible to make some considerations about the estimation uncertainty and to compute $\mathbf{Q}$ and $\mathbf{R}_k$, further details can be found in the original paper [45].

For the computation of the best association of clusters to the corresponding tracks in subsequent frames it is applied the Hungarian algorithm [52] over a matrix in which are arranged all the scores for each possible association between the new clusters and the tracks. The scores are the probabilities computed using an approximate version of Joint Probabilistic Data Association (JPDA), called Cheap Joint Probabilistic Data Association (CJPDA) [53] applied to the kinematic state, i.e., the components of the KF vector related to the Cartesian position and velocity of the targets. More details about the scores computation can be found in [45], [53].

The method adopted for the first steps of data processing has been studied to be robust to blockage of the radar signal and to subjects that move inside or outside the radar sensing area. Even in a single-subject scenario, the tracking of the cluster is not a straightforward problem because of noise and in a general situation, a multiperson radar sensing system should deal with a higher error probability while still keeping a low complexity. The strategy is to consider as a new trajectory the detections that are detected for at least m out of the last n frames. Then, the algorithm maintains the tracks that received a match with any of the clusters detected by DBSCAN for at least m out of the last n frames. This is the so-called m/n logic.

Data synchronization

The synchronization between radar and motion capture system is not obtainable by hardware. Thus, the point clouds sequence Sr = [r_0, r_1, ..., r_n] and the markers positions S_m = [m_0, m_1, ..., m_z] have to be aligned to match them frame by frame. The main idea of the synchronization algorithm is to shift one of the two sets until the best shift is found according to some distance measure, taking into account the different sampling frequencies of the acquisition systems. As this is true for our acquisition devices, we assume in the algorithm development that the motion capture system has a higher sampling frequency with respect to the radar, thus $z \geq n$.

The distance function used for the implementation of this synchronization algorithm is the so-called BQ which is a simple algorithm that, given a set of points S, a set of centroids C, and a radius R, returns all the points in S that are distant at most R from any point in C. In the synchronization algorithm, the set of points S is the radar point cloud and the markers have the role of the centroids. The choice of this distance leads to have the minimum distances when the markers are surrounded by radar points.

Due to a technical limit in the set-up, only radar provides the timestamp for each frame, while the motion capture system records only the beginning instant. So the algorithm we developed so far is based on the assumption that the two independent clocks are not experiencing strong

drifts, thus that it is still possible to obtain an approximate time difference for the whole sequence. Moreover, considering that the target action is repetitive there could be more shifts that give a good match according to the distance-based score, thus the number of overlapping frames is tracked and is used as a complementary score. Finally, some prior knowledge about the dataset construction can help to speed-up the computation, e.g., negative shifts can be excluded iif it is known which system first started acquiring and it is used as a reference.

Algorithm 3.1 Point clouds and markers synchronization.

function look_around(set, $index_{start}$, $index_{end}$)
    $subset \leftarrow []$
    for s in $set[index_{start}, ..., index_{end}]$
        $subset$.insert(size(s))
    $interval \leftarrow (index_{end} - index_{start})/2$
    $penalty \leftarrow [, ..., interval)$
    $penalty \leftarrow$concatenate(flip($penalty, [0], penalty$))
    $subset \leftarrow subset - penalty$
    $index \leftarrow$argmax($subset$)
return $set[index_{start} + index]$
function compute_shift(S_m, S_r, fps_{marker}, fps_{radar}, d)
    $n \leftarrow$size(S_r)
    $z \leftarrow$size(S_m)
    $shifts \leftarrow [-z, -z+1, ..., -1, 0, 1, ..., z]$
    $distances \leftarrow []$
    $fps_{diff} \leftarrow$floor($(fps_{marker}/fps_{radar})/2$)
    for $shift$ in $shifts$
        $d_{shift} \leftarrow 0$
        $n_{non\_overlaps} \leftarrow 0$
        for i in $[0, ..., n)$
            if $i + shift$ in $[0, ...z)$
                $r_i \leftarrow S_r[i]$
                $z_i \leftarrow$floor($(fps_{marker}/fps_{radar})$)
                $m_i \leftarrow$look_around($S_m, z_i + shift - fps_{diff}, z_i + shift + fps_{diff}$)
                $d_i \leftarrow$d(r_i, m_i)
            else
                $n_{non\_overlaps} \leftarrow n_{non\_overlaps} + 1$
            $d_{shift} \leftarrow d_{shift} + d_i$
        $n_{non\_overlaps} \leftarrow n_{non\_overlaps}/n$
        $d_{shift} \leftarrow d_{shift} * (1 - n_{non\_overlaps})$
        $distances$.insert(d_{shift})
    $index \leftarrow$argmin($distances$)
return $shifts[index]$

Dataset preparation

Data normalization is a typical step in deep-learning projects to map the input data to the interval [0,1] which is better handled by the neural network [54]. In fact, the values not normalized can bring to an explosion or a vanishing of the gradients involved in the weights computation. Moreover, the dimensions with higher absolute values could be weighted with a higher importance leading to poor learning.

Various normalization strategies were tested on these data. Some aspects to be considered are the differences in the data dimensions and monitored quantities. The body shape has different measures on the three axes, thus a normalization considering them separately would lead to loose the body shape and create a sort of sphere with all the coordinated values in the interval [0,1]. Thus the axes have to be considered together. Moreover, the main variations in values are in range because the subject is moving back and forth with respect to the radar so the azimuth is constant, and also the elevation does not have a great excursion during gait.

In this context, we define the frame centroid the point whose coordinates are the average of the cloud points' coordinates for each dimension. For each strategy, the translation matrix is computed using the point cloud and then the markers are shifted by the same translation matrix.

- each frame is centered on its centroid. This normalize each dimension with the same weight but loose the spatial information due to the motion.
- each frame of the whole sequence is shifted accordingly to the centroid of the first (or the last) frame to globally maintain the spatial movement.
- each frame is shifted accordingly to the centroid of the previous frame. It is a trade-off between the previous methods to partially maintain the spatial movement at a local level but still removing the bias on the direction of walking.

Moreover, due to the network structural requirement to be fed with data with a constant shape, the point clouds are required to be with the same number of points that was fixed to 256 points. This value is close to the maximum number of points reached in the whole dataset. For the frames with a lower number of points, we repeat some of them, sampled uniformly at random. Finally, as in [38], the dataset was split into sequences of 30 consecutive frames that is equivalent to 3 seconds of data acquisition, with an overlap of 25 frames. Considering that the frame rate of the radar is 10 frames per second, those values were tuned to mantain an acceptable complexity of the model but still provide to the network enough temporal information.

Data Augmentation The use of data augmentation techniques is well established in the deep-learning domain with the aim to extend the input dataset. The techniques are related to the type of data but the general idea is to transform the available data in a new set of samples changing some aspects and thus, for the model it is not obvious that are coming from the same source. In this project, we considered shifts, permutations, and rotations of the points.

With the considerations reported in Sec. 5.1.1 we obtain in total 1116 sequences of 30 frames with around 80% overlapping frames. We divide the dataset into 1000 sequences for training and

116 for test. The test sequences are chosen uniformly from the original set of acquisitions and correspond to complete captures, which implies that they do not share frames with sequences from the training set.

3.3.3 mmIOR architecture

The proposed architecture is inspired by Pointnet++ [41] which is considered in literature a good choice for the extraction of features from point clouds, even when they present a high degree of sparsity [36], such as the ones produced by mmWave radars.

The architecture of PointNet++ has interesting features for the elaboration of point clouds, but still, some modifications are needed being the input data and the task different in our case with respect to the original paper. In the authors' experiments, the tasks were classification or semantic segmentation while in this project the problem is related to the prediction of new points. Then, the datasets used in the original work [55], [56] are about 10 times more dense than the mmWave radar point clouds in our dataset. Moreover, each instance in our dataset is a frame of a temporal sequence and the correlations between the subsequent data can be exploited.

Some modifications are required and the models are here distinguished in Baseline and Recurrent to investigate the impact of the exploitation of temporal correlations.

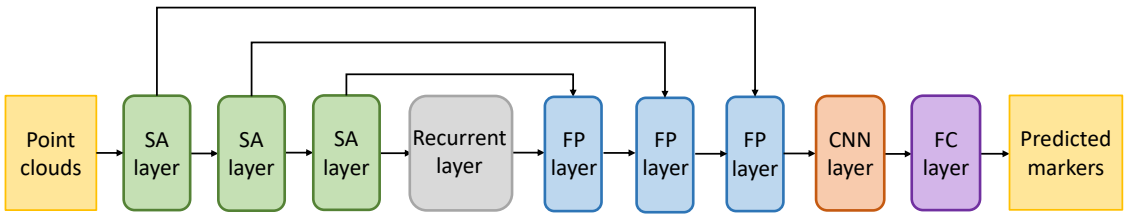


Figure 3.7: Model schematics. The RNN layer is not present in the baseline model.

Baseline Model

The input data are first fed to the SA layers. Their parameters need to be adapted to a lower density with respect to the original architecture, so reduced number of centroid and SA layers itself. MRG strategy was adopted. Then, the features have to be rescaled through the use of FP layers that were also reduced in their size. To obtain the final position some extra layers were appended. After the FP layers, we included a convolutional layer with 32 filters and kernel size [1, 5] followed by max pooling a final FC layer with 30×3 units to predict the Cartesian coordinates of the markers. To prevent overfitting, dropout is applied with a probability equal to 0.2. In Tab. 3.2 are reported the parameters after the tuning step.

SA layer 1	
Centroids	32
Radius	0.15 m
Points in sphere	8
MLP filters	[4,4,8]
SA layer 2	
Centroids	16
Radius	0.25 m
Points in sphere	8
MLP filters	[8,8,16]
SA layer 3	
Centroids	8
Radius	0.4 m
Points in sphere	8
MLP filters	[16,16,32]
FP layer 1	
MLP filters	[32,32]
FP layer 2	
MLP filters	[32,16]
FP layer 3	
MLP filters	[16,16,16]

Table 3.2: Layers' parameters.

Recurrent Models

The design choices are related to the placement, the type, the number of recurrent layers, their parameters, and how the features are passed in input to the recurrent layers. Regarding the placement with respect to the other layers of the baseline model, the choice was to add them between SA and FP layers where the features are more compact. Then, both GRU and LSTM layers were considered to explore the differences between different recurrence types, while in terms of number of layers the architecture we designed involves one layer as a trade-off between the exploitation of temporal information and complexity. Finally, the structure of the features in input at

the recurrent layer was investigated. In the first considered scenario, the features from the SA layers were all concatenated in a single input vector, we called this architecture ConcatGRU. We explored also the use of the features from the different SA layers passed to the recurrent layer sequentially, thus we named this model SequentialGRU.

For ConcatGRU we used a GRU layer with 384 units, while for the SequentialGRU model with 256 units. In both recurrent models, to prevent overfitting dropout was introduced with probability 0.2 and L2-regularization with parameter $\lambda = 0.01$.

Selective Per-Point Distance (SPPD)

The choice of the loss function is important in general in deep-learning projects, in particular, this dataset presents some unbalanced representations of the different markers due to their positioning on the body. Since the goal of our model is to predict the positions of the markers with an ordered association across all the outputs, we require the loss to consider each point-by-point pair between predicted points and ground truth markers, i.e., per-point distance. In particular, for the dataset acquired with a reduced number of cameras, the ground truth is not available for each marker in every frame because of the occlusions and reconstruction errors. Only the visible markers were considered for the loss computation denoted as Selective Per-Point Distance (SPPD) and defined as

$$L_{SPPD} = \frac{1}{\sum_{i=1}^{N} v_i} \sum_{i=1}^{N} d(M_i, T_i) v_i \tag{3.24}$$

where N is the number of protocol landmarks, T_i is the position of the i-th target predicted point, M_i is the position of the i-th marker and v_i can assume value 0 if the i-th marker is not visible in that frame and 1 otherwise.

Learning strategy

The learning strategy for model training was not to reduce the learning rate throughout the whole training process, unlike the common practice in neural networks. Instead, we reset the optimizer to its original state and learning rate twice, accelerating convergence after 300 epochs. Subsequently, we allowed the optimizer to dynamically adapt the learning rate for the remaining 900 epochs. The total number of training epochs to achieve convergence is 1500. We used Adadelta optimizer with an initial learning rate value of 0.01.

3.4 Results

In this chapter, we describe the results of the proposed system for markers' prediction. The dataset used to obtain these results is limited to the first data acquisition because of the already described technical problems in the synchronization of the sequences.

In Fig. 3.8 the ground-truth is shown with the mmIOR reconstruction obtained for the same frame. The pose is well-reconstructed but it presents a shift. The observed shift is not a systematic

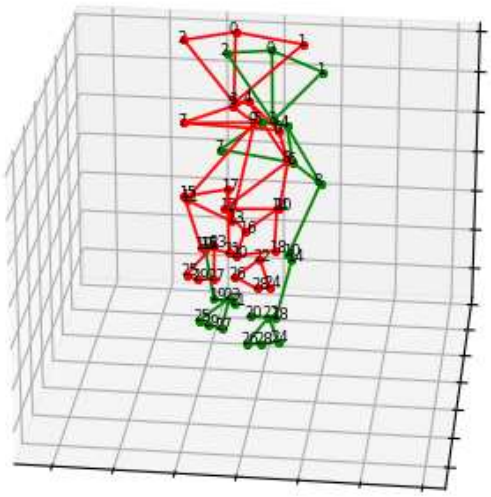


Figure 3.8: Pose reconstruction. The ground-truth (green) is shown with the mmIOR reconstruction (red) obtained for the same frame.

error of the model that could be corrected. We suppose that it is the effect of a still uncompleted learning, thus this could be improved training the model using a more complete and increased dataset.

Model	x (m)	y (m)	z (m)
Baseline	0.11	0.21	0.17
ConcatGRU	0.08	0.13	0.09
SequentialGRU	0.08	0.14	0.10

Table 3.3: Absolute mean distance between predicted points and ground truth for the tested models.

Tab. 3.3 reports the absolute mean distance between the predicted points and the ground truth for the tested models. There is a clear improvement in the overall results with the addition of the RNN layers with respect to the Baseline model. A slight error decrease is obtained concatenating the features from the SA layers instead of sequentially passing the SA layers.

From Tab. 3.3, there is a prevalent error for all models in the depth dimension, i.e., along the y-axis. A source of error can be the sequence synchronization which operates mostly along that di-

mension. Another reason is the implicit difficulty of the task along that dimension which includes a change of direction. In Tab. 3.4 are reported the mean distances obtained across the dataset for the ConcatGRU model for each dimension. It can be noticed that the bias is comparable for symmetric markers and for similar positions on the body, e.g., the last 12 markers that are all on the feet. Even if the precision required for clinical gait analysis is higher, the errors we obtained have the order of centimeters and this is similar to the results in literature [38]–[40].

Marker	x (m)	y (m)	z (m)
C7	0.05	0.12	0.08
LA	0.12	0.11	0.08
RA	0.11	0.12	0.09
L5	0.05	0.10	0.06
RPSIS	0.05	0.11	0.06
LPSIS	0.07	0.11	0.06
RASIS	0.09	0.10	0.07
LASIS	0.10	0.09	0.07
RGT	0.14	0.09	0.07
LGT	0.10	0.10	0.08
RLE	0.10	0.11	0.07
LLE	0.10	0.11	0.08
RME	0.06	0.11	0.07
LME	0.04	0.12	0.08
RHF	0.10	0.11	0.09
LHF	0.10	0.12	0.09
RTT	0.09	0.14	0.10
LTT	0.07	0.12	0.11
RLM	0.08	0.14	0.12
LLM	0.08	0.14	0.12
RMM	0.06	0.14	0.12
LMM	0.06	0.14	0.11
RCA	0.07	0.15	0.11
LCA	0.07	0.17	0.12
RVMH	0.09	0.15	0.11
LVMH	0.10	0.15	0.12
RIMH	0.06	0.16	0.13
LIMH	0.07	0.16	0.11
RIIT	0.08	0.16	0.11
LIIT	0.08	0.16	0.11

Table 3.4: Mean distance from the ground truth for each markers for ConcatGRU.

3.5 Conclusions and future directions

In this chapter, the capabilities of radar sensing for motion capture application were explored. In particular, the advantages of radar sensing are the adaptability to any illumination conditions both outdoor and inside a laboratory, then the subject privacy is preserved, and the combination

with deep-learning tools can lead to markerless system, i.e, no need to an expert for positioning and long subject preparation time. We proposed and built a pipeline for the creation of a FMCW radar-based markerless motion capture system, even if the applicability to a real gait analysis scenario is still a challenge. Our current results show that the learning process requires a greater amount of data and some technical problems have to be solved to increase it, but still the reconstruction is coherent with the expected body structure. Moreover, in the first dataset, there was a strong noise due to the positioning of the radar too close to the subject's walking area.

Furthermore, the results comparison between baseline and the models with recurrence show that the recurrent layers can improve the prediction by exploiting the temporal correlation in the sequences of data from radar. This also suggests that in the presence of more data and at the expense of slightly greater complexity, the markers prediction could further improve. Future developments could tackle different aspects to improve the applicability of the radar-based markerless solution. First of all, the implementation of a more robust synchronization algorithm could help to exploit a greater amount of training data which seems to be the main limitation. Then, the radar can be combined in a sensing network and the information coming from different points of view with respect to the walking direction could improve the marker reconstruction.

4

Multimodal sensing platforms

he possibility to sense simultaneously multiple quantities enables the exploitation of more knowledge coming from the quantities themselves and the correlation between them. However, to obtain some reliable results it is of paramount importance to properly design the set-up for the experimental data collection. In this chapter, two practical solutions for multimodal sensing are discussed. Both the systems presented are designed for research purposes. In the first case, Open-MBIC is more oriented to the general case of multiple simultaneous Bluetooth connections and the contribution is in the creation of a framework that can be adapted to different research needs while offering a practical use case. On the other hand, the second contribution is specifically oriented to solve a practical scenario of multi-signals data acquisition during gait and the main aspect is the synchronization requirements and their impact on this kind of data.

4.1 Motivation to develop Open-MBIC

Bluetooth is a consolidated and widely used communication technology to connect wearable devices [57]. Bluetooth Low Energy (BLE) is a version of the Bluetooth technology which has been conceived and is mostly used for applications. First, since the BLE is designed for the communications with wireless sensors, e.g., microphone, and transducers, e.g. headphone, any BLE learner needs to know the application coding paradigm. This being granted, learners find they have access to a vast amount of information about this technology, but also that the available resources are not well organized, and that it is often difficult to organically access them. In particular, the BLE documentation provides explanation for typical client-server configurations, where a single server and a single client are involved in the data exchange. However, the management of multiple and simultaneous connections is not duly described.

Our contribution is the development of the Open-MBIC library which has been implemented to ease the setup of BLE links across multiple devices concurrently, providing a multi-client communication framework towards a single server that receives and manages different information flows by avoiding inter-flow interference and related packet losses. The library is easy to understand, as only official BLE resources are used. Also, libraries provided by third-parties are avoided to facilitate the understanding of the basic concepts involved in the protocol workflow and to allow connecting new sensors with minimal extra effort.

To demonstrate the Open-MBIC functionalities, in this chapter we showcase the use of the library for the acquisition of vital signals from Covid-19 patients (Sec. 4.4). In this use case, a finger tip pulse oxymeter, a chestband heart monitor and an Inertial Measurement Unit (IMU) are connected to the patient or the caregiver smartphone through a custom Android application built on top of Open-MBIC, allowing the collection of vital signs during physical exercise for the assessment of the disease progression and recovery.

4.2 Related works

4.2.1 Android libraries for BLE

The use of multiple connections exploiting the Bluetooth communication technology has been extensively investigated so far.

During the development of Open-MBIC many projects were analyzed, but none of them provided a clear code to be inspected to verify the working phases of the protocol, and none was easy to adapt and reshape to other means, such as connecting different sensors or controlling the communication and connection procedures. An example is provided by the Redfang library from [58] where a Bluetooth communication framework is developed, but the code is only partially shown. In this way, the user may understand the general working flow, but finds it rather hard to manipulate the code for a specific use case of interest. Also, the connection is established using the classic Bluetooth protocol: this is a common problem in the literature, where only the creation of standard Bluetooth connections is explained/exemplified in depth, while Bluetooth Low energy communication is just mentioned, but never treated in detail. On the other hand, BLE communicating devices are usually sold with a working application that can be used to collect data from their internal sensors. The collected data is usually aggregated or processed depending on the seller interests, and they way in which this is done is hardly documented. In addition, raw data is rarely accessible: some manufacturers provide the code along with the compiled application and this permits to adapt the code and retrieve it. If the project is not open-source, the extra work also includes a lot of reverse engineering and this increases the burden on developers. The Open-MBIC library is here proposed to fill these gaps: it exploits the general properties of the BLE communication system, providing a simple to use and developer-friendly environment that can be useful for everyone who would like to use BLE in IoT scenarios. The extra work is reduced and guided through some steps in the library code, so that the communication procedures can be promptly adapted to different needs.

4.2.2 BLE and E-health

E-health and its branch, telehealth, are emerging fields in both the clinical and technological domains. They are becoming increasingly important giving the emergence of new diseases, an ageing population in most developed countries and the pervasiveness of modern digital and communication technologies.

We advocate that most E-health services could be offered, at a low cost, using standard smartphones, acting as digital hubs to retrieve health data from the users and send it via secure channels to the clinic. An added benefit of such technology is that end users are already acquainted with them, which makes their adoption natural and fast. For these reasons, many E-health applications rely on smartphone technology, and this is gaining momentum, to the point that the term mobile-health (m-health) was recently coined. Two examples of smartphone-sensors systems for heart monitoring are proposed in [59] for IOS devices and in [60] for Android ones.
The environment in which the monitoring has to be accomplished is a main aspect to consider in the design of m-health applications. One example is provided by the SMS-based monitoring system of [61], where mobile technology is utilized to provide communication in geographical regions that are not reached by the standard Internet. If standard cabled Internet communication is available, more sophisticated methods become possible: for example, in [62] and [63] ZigBee based systems are exploited. A recent and more suitable protocol to perform remote sensing is BLE. The authors of [64] perform an analysis of medical data streaming using BLE, concluding that it is very efficient to send small amounts of data during short term connections. Moreover, in [65] a comparative study between BLE and ZigBee shows that, for intensive monitoring applications, BLE is the least energy-expensive solution. For these reasons, BLE is becoming quite popular for the development of monitoring platforms. For instance, in [66], the authors demonstrate the robust acquisition and transmission of several health signals using BLE and different sensors. This is very useful in a clinical scenario, where diverse data types have to be recorded, usually from wearable sensing devices.

4.3 Open-MBIC

Multiple Ble for Iot Connections (Open-MBIC) is a Java library that can be promptly integrated into modern Android applications: it can be a valuable tool towards developing IoT solutions involving BLE communications. For a practical example of its use, please refer to Section 4.4.

4.3.1 Main functionality

For a description of the working principles of BLE, the communication protocol and the related terminology, the reader is referred to [57]. The library is developed around the `MainActivity` class, which performs the following activities:

- Require the permission to access the device location.

- Scan the environment using the `ScanResultAdapter` class, to populate a list on the User Interface (UI).

- Activating the connection at the General ATTribute Profile (GATT) level in case a connection is selected by the user: the connection state is returned and, if the attempt was successful, the device services are discovered.

- Handling notifications and indications. This means that the client (the Open-MBIC app) receives a notification when some server (device) characteristics change. Only the characteristics that have the "notify" or "indicate" property can be enabled and, in turn, generate/send notifications.

- Handling the reception of notifications, and the processing of the received data packets, which are decoded according to the producer's encoding/decoding rules, to retrieve data and temporarily store it into local buffers.

- Data from different sensors can be saved locally in the Android device storage. Moreover, upon receiving the data on the smartphone, it can be forwarded to a server: Open-MBIC supports the relay of data via TCP/IP.

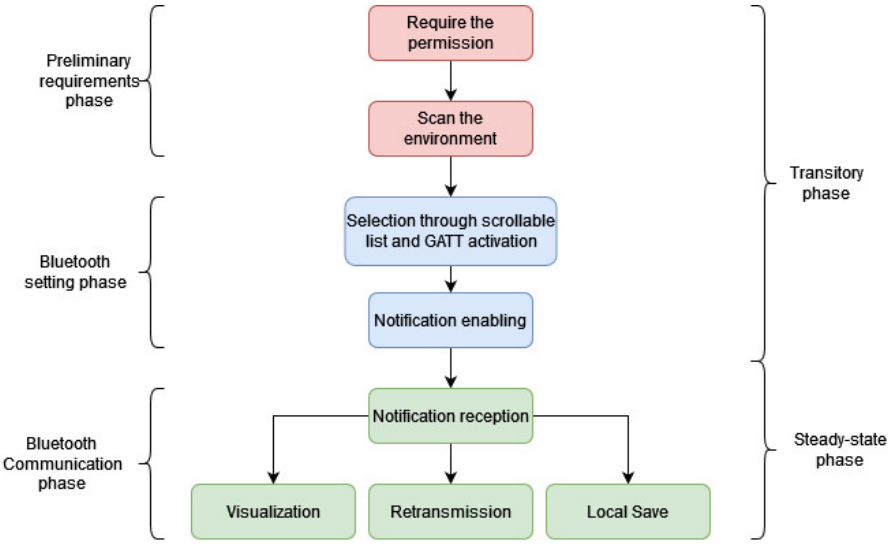


Figure 4.1: Open-MBIC main functionalities.

4.3.2 Software and hardware requirements

Open-MBIC was designed to work on commercial Android devices. Its functionalities were tested on different smartphones and Android versions, as reported in the following Table 4.1.

Device model	Android version
Samsung s8	Android 9
Redmi Note 8T	Android 10
Galaxy A71	Android 11
OnePlus Nord2 5G	Android 11

Table 4.1: Tested smartphones and OS versions.

4.3.3 Using Open-MBIC

The code can be easily adapted to different problems, setups and devices. There are some properties of the devices that should be known or found out by the user to adapt the library to the specific use case. First of all, the Universally Unique IDentifier (UUID) associated with the device needs to be selected and defined in the `GattAttributes.java` file. Some profiles are standard, e.g., the heart rate monitor, others are custom-defined by manufacturers. For custom-defined profiles, modifications to the library are to be implemented through the `enableNotifications` and the `onCharacteristicChanged` functions, which respectively define which characteristics - data source -are enabled for each notification and how the received data is to be collected and accessed. Some typical mutually exclusive configurations are:

- one service for each data type.
- a control service to manage the communication and a data service to handle the transmission of the sensor data. As the latter service can group more characteristics, different data types can be transmitted within the same packet.
- a unique service grouping more characteristics specifying data structures and control rules.

Finally, the data could be encoded prior to its transmission and so the packet decoding could change according to the connected device. Usual conversions are supported and specified in the `Utils.java` file: it includes the translation from hexadecimal strings to byte arrays and viceversa.

4.4 Use case: REMOCOP

The use case we now discuss is a home telemedicine system for REhabilitation MOnitoring of Covid-19 survivors and Chronic Obstructive Pulmonary disease patients (REMOCOP) that was object of a collaboration with a hospital. Open-MBIC and the use case described in the following sections resulted in publications to international conferences in the field [67]–[69]. Moreover, REMOCOP platform was used with didactic purpose in different courses for master degree in "ICT for Internet and Multimedia" at the University of Padua.

The typical REMOCOP connectivity architecture for data sensing and acquisition is shown in Fig. 4.2. First of all, it includes the REMOCOP Android application based on Open-MBIC, managing all the connections. It has to be installed on an Android smartphone which can be used

Figure 4.2: The REMOCOP system.

by the patient or the caregiver. Then, multiple sensors are connected via BLE to the smartphone that receives data from them. In the considered setup, the first sensor is a chestband heart monitor to collect Heart Rate (HR) and single-lead ECG. Then, there is a pulse oxymeter measuring the peripheral oxygen blood saturation (SpO_2) and the photoplethysmogram (PPG). The last sensor is an IMU used to measure both three-axis acceleration and angular velocity. The application allows the user to either store the data locally or transmit it to a server and, in turn, save it into the patient electronic health records (EHR). In our study case, the EHR was emulated using a personal computer in the lab. Some Python scripts complete the library: providing functions to process the data offline in case they were stored on the smartphone, or at runtime, if received in streaming mode. The TCP/IP connection is also managed by the server by the means of a dedicated Python script.

We now discuss some key functionalities of Open-MBIC. As a first step, in Fig. 4.5 the smartphone is scanning the environment to find devices to communicate with. After the user selects the sensors, the communication is set up and the information exchange can start. The last necessary step to receive data is to perform the reading operation. The acquired data is stored into a memory buffer and, depending on the user's choice, Open-MBIC can be utilized to save data locally into a file or to stream it through a TCP/IP channel to a server. In Fig. 4.7, the procedure to locally

Figure 4.3: A subject equipped with sensors.

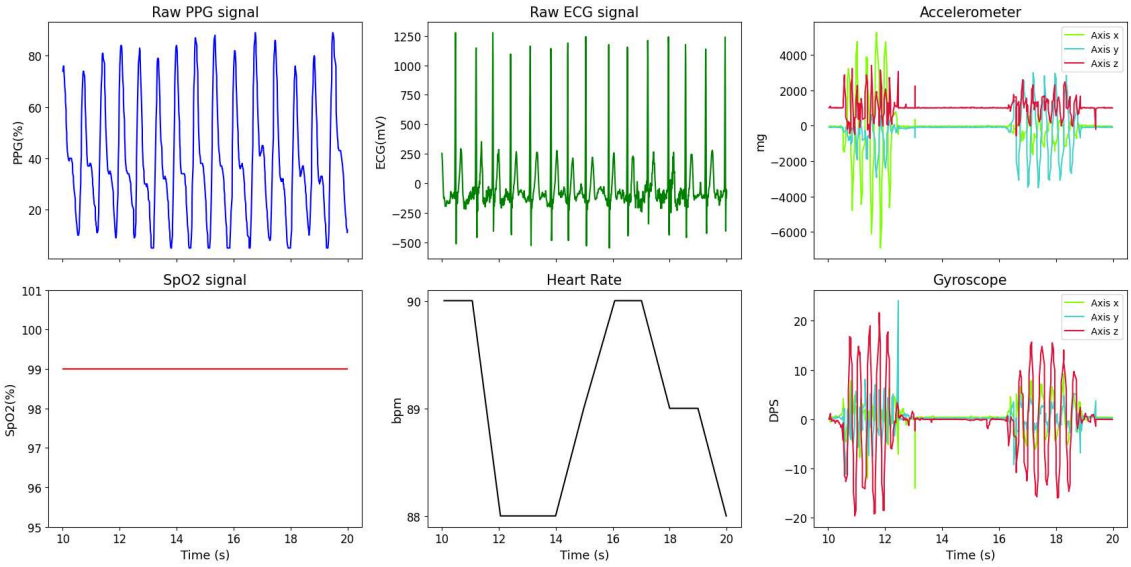


Figure 4.4: An example of 10 seconds of data collection.

save data is shown. For the data forwarding to a server, knowledge of the server IP address in the local network is required. The data received by the server is managed by an adapted version of the Open-MBIC Python script. In our case, only the real-time visualization was implemented, but any other operation to store or process data can be implemented on the server side, depending on the use case.

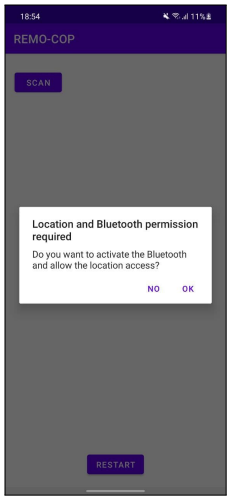
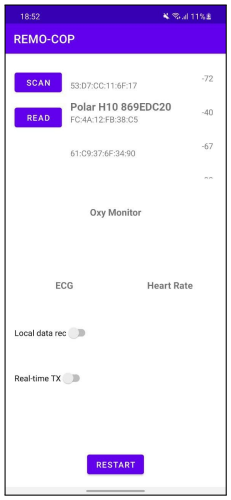
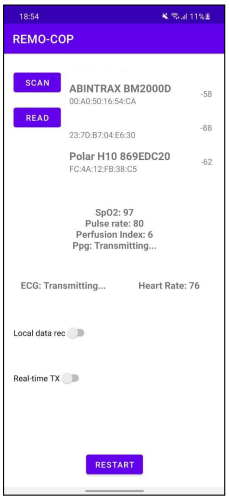


Figure 4.5: The permission to access the device location is required for the scanning.

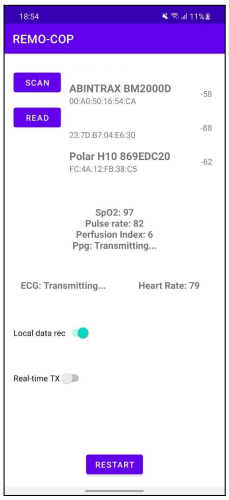


(a) The user can select the correct devices from a list.



(b) The notifications from sensors are captured and data is received.

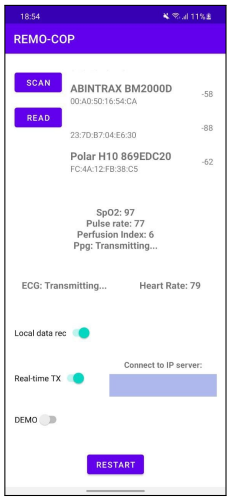
Figure 4.6: Receiving data from the field sensors on the smartphone.



(a) Data are saved on the smartphone since the user decides to save them.

(b) A snapshot of the file manager.

Figure 4.7: Saving data locally.



(a) Saving and trasmitting can be performed simultaneously.

(b) Input the IP address to reach the remote server.

Figure 4.8: Transmitting data to a server.

4.5 Conclusions and future directions

Open-MBIC is an easy to use framework to fill the still-open gap towards developing Android-BLE applications. The proposed use case shows how it can be used to build applications to interact with and receive data from multiple BLE sensor devices. Research teams can use the library as a starting point to tackle their use cases and accelerate their research work. Moreover, the simplicity of the Open-MBIC code makes it suitable for educational purposes, to teach how to set up BLE connections and how to collect data from sensors. Indeed, to be adapted to a new sensing application it requires only few guided steps. Present and future endeavours include adaptations to automate some of the operations that are now required from the user and the addition of support for the interaction with desktop applications. Open-MBIC is an open source project and the code is shared with the community, via GitHub repository at `https://github.com/silvzamp/Open-MBIC`.

4.6 Motivation to develop our trigger box

In the following sections, we discuss the importance of the simultaneous collection of biosignals with the help of a practical scenario and the solution we developed. The work described was the set-up used for data collection in [70].

4.6.1 Gait analysis

Gait Analysis (GA) is a process of evaluation of walking ability through a instrumented measurement [71]. In the clinical domain, the subjects are patients with locomotion impairments and the aim is to provide information to support clinical decisions, rehabilitation monitoring, and other related situations. The measurement of joint angles (kinematics), joint forces (kinetics), muscular activity, foot pressure, and energetics (measurement of energy utilized during locomotion) allows the physician to design procedures adapted to the individual needs of patients as long as two conditions are satisfied. The first is the presence of an adequate instrumentation for measurements and the second is the knowledge about healthy and pathological walking [72].

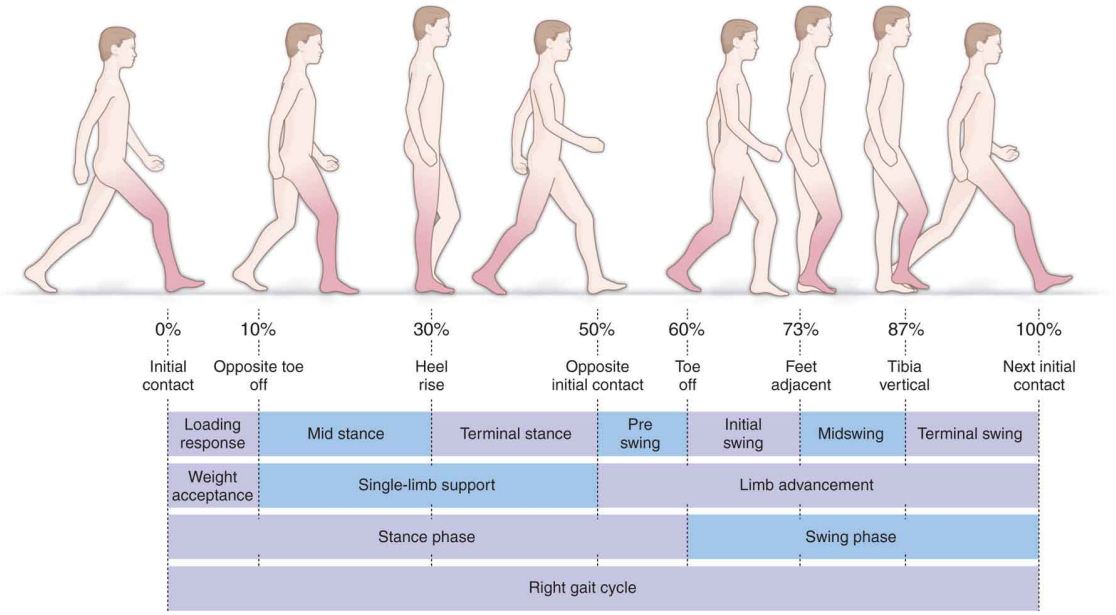


Figure 4.9: Gait cycle phases.

Since late 1800 it is known that the individual joint angles, the displacements of segments, and of the whole body are essential measurements to distinguish between healthy and pathological gaits [2]. The pioneers techniques included interrupted light, reflective strips, manual goniometers, and photography [73]. With the evolution of vision technologies, the development of visual markers-based techniques has reached great level of precision becoming the gold-standard technique for gait analysis.

4.6.2 EEG signal

electroencephalogram (EEG) is the signal obtained recording the electrical activity of the brain measuring the currents that flow during synaptic excitation in the cerebral cortex. It is a non-invasive technique in which the EEG traces are obtained by placing electrodes on the scalp. The frequency range of the EEG of an healthy subject is 1-30 Hz with maximum amplitudes about $100\mu V$. While the spatial resolution is very limited, the temporal resolution is of the order of milliseconds and makes the EEG a source of valuable information. Indeed, this signal contains information about the whole status of the body and it can be used for research purposes and to inform medical diagnosis [74]. Contrary to other high-resolution anatomical imaging techniques, as magnetic resonance imaging (MRI) and computed tomography (CT), EEG can be recorded during movement. Despite the gait introduces some noises in the signal, this makes the EEG a good choice among neural signals for gait decoding and analysis at the present time.

4.6.3 EMG signal

Electromyography (EMG) is the signal that records the muscular electrical activity, in particular, surface electromyography (sEMG) the EMG signal recorded on the surface of the skin, in the rest of the chapter we refer to that simply as EMG. The frequencies of interest are usually lower than 150 Hz and the maximum typical amplitude is about 10 mV. When acquired during gait, EMG study permits to retrieve a great amount of information about the subject's motor health. In particular, the force exerted by muscles is related to the EMG amplitude, so the correct envelope estimation and derived quantities are important sources of information for the movement analysis[75]. This information have been proved to be sufficient to decode gait cycle phases[76] that can be useful for the gait analysis itself, but also to control some devices based on the gait phases. Moreover, the pattern of muscular activations are indicators of the subject health in terms of activation frequency, involved muscles, and timing with respect to the movement[77]. Thus the EMG study is a tool for both research, diagnosis, and assistive devices control.

4.7 Related works

In this section, we explore the literature related to the study of EEG and EMG signals during gait.

The information contained in EEG and EMG signals have been shown to be sufficient to classify the gait phases with the use of different classification tools. First, considering the EMG signal, it has proven to be informative for the gait decoding. This finds application for a deeper understanding of the gait biomechanics, in the clinical domain, and to control assistive devices. For example, in [76] hidden Markov models are used to decode the gait cycle phases considering five different states. More recently, in [78] the combination of features and the neural networks have been investigated to find the more suitable for the multiclass discrimination of seven gait phases. More in general, the literature report a growing interest in the research of more precise

tools for gait decoding [79], [80]. EMG gait-related information have been used also to classify pathological subjects demonstrating that this signal is informative for the purpose [81].

On the other hand, the EEG signal encodes the information of the whole state of the body, gait included. The authors of [82] were able to train an LSTM model to predict the stance and swign gait phases. While in [83] the authors investigated the gait decoding through the estimation of joints' angles using EEG signal combined with different learning algorithms and obtained the reconstruction of the angles' profiles during gait.

In [82], the authors propose a human-machine interface that interact using the motion information coming both from the brain signals and from the muscular activity to decode gait. From this perspective the information fusion enhance the reliability of the control system. The work published in [84] explore the use of both EEG and EMG features for the monitoring of the drug treatment in Parkinson disease patients showing that the information they carry is sufficient to detect anomalies due to bad drug dosages. The great amount of information contained in these signals is exploited also in [85] in which different set of features were analyzed for gesture recognition in IoT for healthcare applications. In general, the combined use of the two signals enhance the estimation performance in the studies that consider both EEG and EMG [86].

The general picture depicted by the recent literature suggests the importance of studying the EEG and EMG for different applications. Moreover, it seems that the joint investigation of the two signals during gait can offer new perspectives for a better comprehension of the motor control mechanisms. However there is still a lack of standards for joint EEG and EMG study during gait and the definition of an effective set-up is a first important step. According to our literature review, the set-up aspect can be further investigated to create a standard that assures reliability.

4.8 Proposed sensing platform

Due to the importance of the EEG and EMG study during gait, we propose a sensing platform that practically create the conditions for the collection of these data. It is of paramount importance to study and prepare the correct set-up before starting any further research. Our contribution is in the creation a practical, low cost and effective set-up and its validation against a commercial one. The common feature we can control for all the considered acquiring systems it is the beginning of the data collection. The external input that causes the beginning of the data acquisition process is defined as "trigger".

4.8.1 EMG acquisition system

The system we considered in the sensing platform integration is "Wave Plus" by Cometa which is a multi-channel wireless surface electromyographic system (Fig. 4.10). The sensors are attached to the subject skin with certified double side tape and the sensors' transmission modules communicate wirlessly with the base unit in order to avoid to create obstacles in the subject movement.

The trigger mechanism for the acquiring device is controlled by a 1-bit signal with logic high level at 5 V and low level corresponding to 0 V. The data acquisition starts and continues while

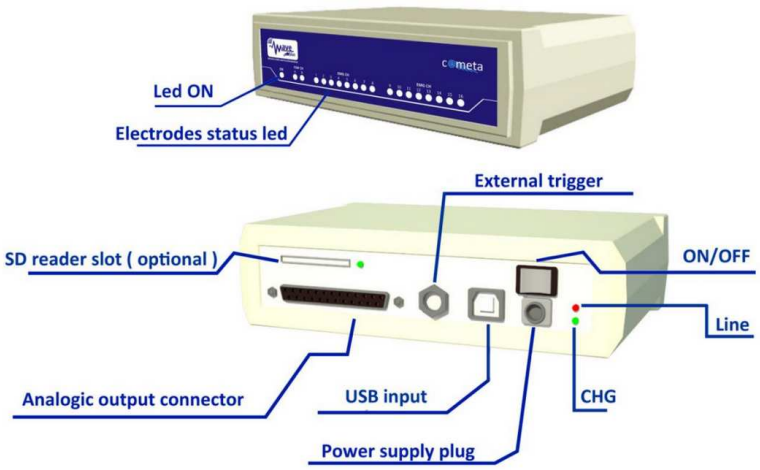


Figure 4.10: EMG acquisition system (Cometa Wave Plus) - figure from the device user manual

the level is high and, viceversa, it stops when it is low.

4.8.2 EEG acquisition system

The system we used is the AntNeuro kit, including cap with electrodes and the amplifier. The processing and recording unit is integrated in the amplifier (Fig. 4.11). The communication between the sensors in contact with the subject's head and the amplified is wired. Thus, the system can be used with a backpack to allow the subject to move while recording.

Figure 4.11: EEG acquisition system (AntNeuro) - figure from the device user manual

The system can receive start and stop triggers from external signal but it can also record other events. The amplifier can receive a 8 bit signal as external trigger, thus 256 values can be encoded

and associate to different events in the data collection with an experimental protocol.

4.8.3 Motion capture system

The system we connected for the gait anaylsis is a Vicon stereophotogrammetric system of MX-T series equipped with infrared cameras controlled by the hub called Giganet MX. The recording system accepts triggers for both the start and the stop events. Differently from EMG acquisition system that operates with the level change, the hub to control these cameras operates a state transition when the signals in the start and stop triggers correspond to one of the combinations in Tab. 4.2.

Start	High	Low	High	Low
Stop	High	High	Low	Low
Trigger action	Signal held high until the pin in pulled to GND. This is also the default system setting when no device is connected.	Remote Start pin is pulled to GND and capture starts at the next video frame.	Remote Stop pin is pulled to GND and capture stops at the next video frame.	Undefined (do not use).

Table 4.2: Signals triggering combinations for Giganet.

In Fig. 4.12 is shown the electronic circuit to be implemented in an external device to trigger the stereophotogrammetric system.

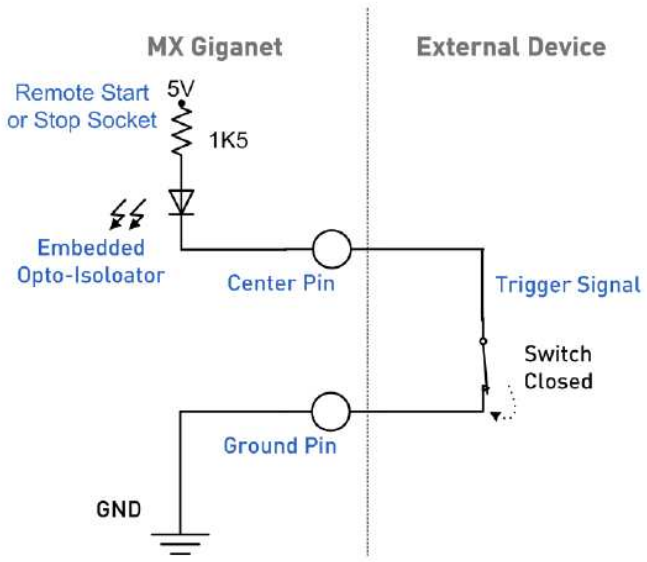


Figure 4.12: External trigger mechanism (Vicon) - figure from the device user manual

4.8.4 Trigger box

Given the information about requirements for the triggering of the three devices, we developed an Arduino-based trigger box.

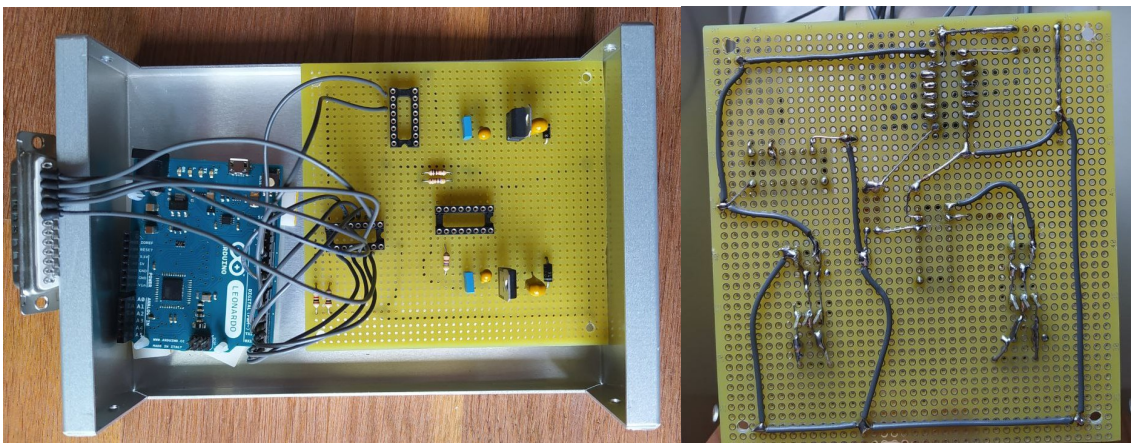


Figure 4.13: Trigger box hardware.

In the microcontroller it is possible to commute simultaneously a set of 8 registers representing the digits of a byte. Thus, we connected two of these registers to the Giganet, one to the Cometa device and the remainings to the AntNeuro. The connection for the Cometa is the simplest and just change the TTL level between 0 V and 5 V according to the logic of the user inputs. For the Giganet, more electronic components were necessary to map the start and stop events; one of the signals usually held high has to be pulled to GND to activate the trigger. The schematic we follow is the one in Fig. 4.12. Finally, the five bits remaining are used to record events in the EEG data series. The software we implemented maps the numbers [0-31] in the five bits, keeping the combination of five zeros as the one dedicated for starting trigger. The sofware was implemented with a combination of a C++ code for the low level logic to be loaded in the microcontroller and a Python script for the user interaction.

4.9 Validation

4.9.1 Methods

The validation was performed through the use of waveform generator. Both the EEG and the EMG acquisition system were connected through their sensors to the generator and we injected square waves with period $T = 1$s, 25% duty cycle, voltage level [0-5]V. For both systems we chose two channels to further check the synchornization. We used channel 4 and 5 from EMG and channel Oz and Bip from the EEG acquistion system. Due to the non-electric nature of the acquired signals, we could not test with the waveform generator input the synchornization of stereo system. The sampling frequency of EMG acquisition device is 2000 Hz, while the EEG amplifier was

set to 1024 Hz. The values are typically used in the real data aquisition. Before extracting the wavefronts instants, a resampling was operated to have the same sampling frequency for both signals.

After the data collection, an edge detector was applied to extract the wavefronts in an automatic way. We used the open source MATLAB implementation of the Canny algorithm [87] available at `https://cismm.web.unc.edu/resources/tutorials/edge-detector-1d-tutorial/`.

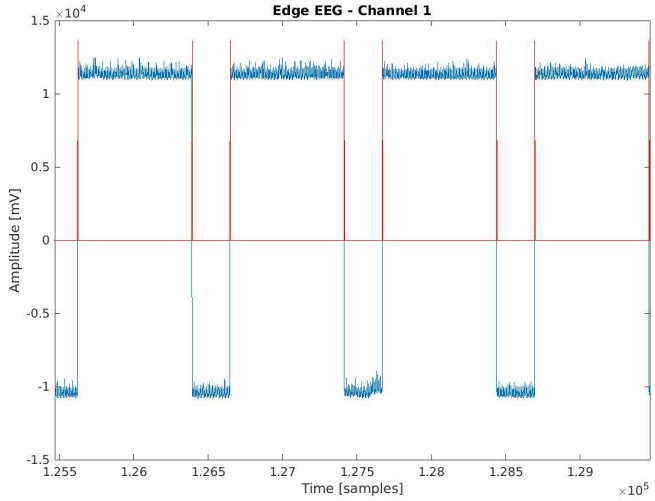


Figure 4.14: Wavefronts indentified by the edge detector.

Then, the samples corrisponding to the same wavefronts were compared to find the time shifts and we computed both intrasignal and intersignal delays.

Finally, the same tests we repeated triggering the systems with a commercial trigger box sold by the same company that produces the EEG acquisition system. The control software for this box is given by manifacturers and is implemented in MATLAB.

4.9.2 Error correction

From the first experimental results we noticed that the delay is incremental (Sec. 5.4). Thus, a simple solution is to keep the data acquisition shorter than a maximum duration. This maximum duration can be defined as the corresponding duration of the maximum tolerated error in terms of samples delay. A more accurate solution is to add a correction term to the samples approximating the error to a staircase function. The solution we propose is not computationally expensive but it could permit to increase the precision of the data collection. The staircase pattern can be fitted with a staircase function in which the parameters to be learned in the fitting are the number d of discontinuities and the lenght l of each step. Moreover, it has to be considered during the function fitting is that the samples shifting is a discrete process, thus the fitting function need to assume only integer values. We define this function as follows:

Algorithm 4.1 Error correction.

function steps($instants$, d, l)
  $result \leftarrow data$
  $result(instants < a) \leftarrow 1$
  $result(instants > a) \leftarrow 0$
  for $index$ in $1 : l$
    $result(instants > a * index) \leftarrow -index$
return $result$

Then, with the non-linear least squares error it is possibile to find d and l and to obtain the correction function for the set-up based on a previous acquisition for error calibration.

4.9.3 Results

From the analysis, both systems compensate the delays between their own channels and the intrasignal delay is limited to ± 1 sample of delay. However, the different clocks start to diverge and the intersignal delay is accumulated during the data acquisition session; in the analysis of the two synchronized systems there is an incremental delay that appears to follow a staircase trend.

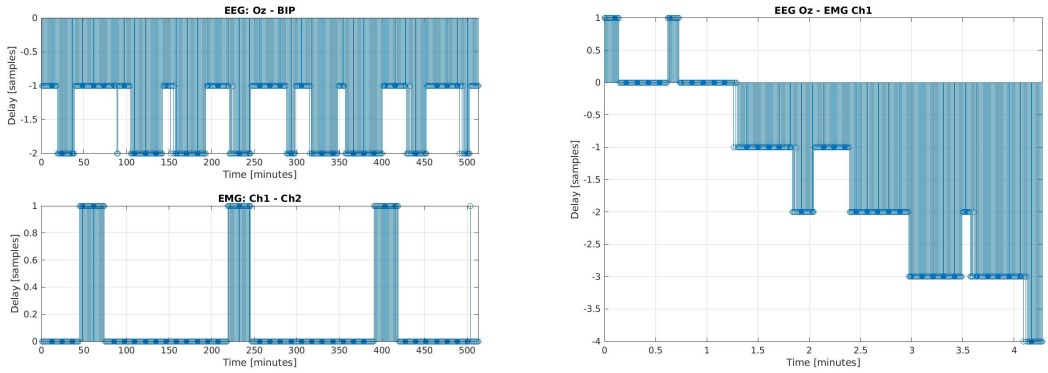


Figure 4.15: Intrasignal delays - Different channels of the same device are compared.

Fig. 4.15 shows both the intrasignal and intersignal delays, while in table 4.3, the average delay obtained comparing the detected wavefronts' instants. After the simple error correction procedure we implemented, the error decreased for each test. In Tab. 4.4 are reported the total number of samples of delay for each test and sampling frequency. While in Fig. 4.16 is shown the intersignal delay pattern considering the correction, it is clear that the correction impacts positively.

Finally, the comparison of the tests performed with our box and th commercial one shows that the errors have the same order of magnitude and are comparable.

Test	AVG error ± STD [ms]	
	F_s = 1024 Hz	F_s = 2000 Hz
Test 1	-1.25 ± 1.35	-1.07 ± 1.31
Test 2	-1.62 ± 1.35	-1.64 ± 1.29
Test 3*	-3.81 ± 1.17	-3.76 ± 1.29
Test 4*	-2.99 ± 1.34	-3.01 ± 1.31

Table 4.3: The average delays between acquisition systems, recording duration of 4 min and 27 sec, according to the shortest test. The symbol * indicates the tests with the commercial trigger box.

Test	Before [samples]	After [samples]	F_s [Hz]	Duration [min]
Test 1	809	209	1024	4.27
	1471	667	2000	4.27
Test 2	4784	745	1024	5.89
	9197	4313	2000	5.89
Test 3*	3400	2241	1024	5.85
	6508	4682	2000	5.85
Test 4*	6165	1250	1024	5.39
	11895	4919	2000	5.39

Table 4.4: Error results before and after the application of the correction algorithm.

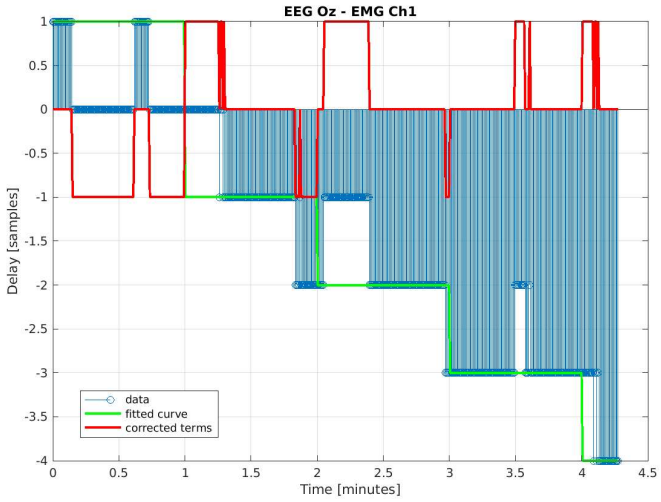


Figure 4.16: Intersignal delay and error correction.

4.10 Conclusions and future directions

We developed a platform to collect simoultanously the EEG and EMG signal from a subject during gait proposing a solution based on commercially available devices. The enabling of simultaneous data collection is of paramount importance to the advancement of the research for EEG and EMG during gait. We propose also a simple and effective methodology for the validation of the set-

up induced error. The results we obtained suggest that a technical limitation is the absence of a feedback signal from the sensing devices. Infact, the systems are designed with the possibility to trigger the beginning and, in some cases, also other events, but to support future research that will require extreme temporal precision, the design could include a feedback signal for the clocks synchronization. In absence of this kind of mechanism in their devices, the researchers should be aware of this problem and they could apply different strategies depending of the precision level required.

5 Positioning at 28 GHz

5.1 Introduction

Considering an indoor context, environment sensing and human position estimation are useful for different emerging use cases of the Sixth Generation (6G) of communication systems; e-Health, smart buildings and the optimization of energy consumption, augmented and virtual reality, industry 4.0, and pervasive connectivity[88]–[90]. In this holistic context, sensing, and communication are strictly connected [88]. Due to the increasing ubiquity of communication devices, the requirements for listed smart applications and processes can be satisfied using wireless communication to sense the environment with the advantage of exploiting existing devices. Moreover, wireless sensing is also device-free, i.e. the subject is not required to wear any sensor [89].

Despite opening possibilities for sensing, some key features of 5G and 6G, i.e. millimeter Wave (mmWave) and Multiple Input Multiple Output (MIMO), imply the usage of narrow beams in mmWave communications. This poses some challenges due to the increased sensitivity to blockages. In particular, the human blockages are disruptive at mmWave frequencies and they are a frequent event in the scenario of indoor communications [91], [92]. Thus, it is of paramount importance to identify the human blockages in order to adapt the beamforming schemes and to preserve the link quality [93]. In other words, sensing can improve the communications itself exploiting the knowledge of the blockage directions due to human presence.

The novelty in the presented work is in the exploitation of the RSSI information to predict human positioning in the context of narrow and directional communication beams in 5G and 6G scenarios. The rest of the chapter is organized as follows: first, in Section 5.2 we describe the context of wireless sensing coupled with communication purposes, exploring both full Channel State Information (CSI) and Received Signal Strength Indicator (RSSI)-based solutions. Then, Section 5.3 discusses in detail the proposed system with details about the hardware, experimental set-up, Radio Frequency (RF) data collection, and processing. After that, Section 5.4 presents the

results of our predictions considering different subject positions. Finally, in Section 5.5 the future work is proposed.

5.2 Related works

Considering the indoor localization task, RSSI has been extensively studied due to its availability. The error obtained in previous RSSI-based approaches in literature was high; ranging from about one to a few meters in the indoor case [94]. Due to the omnidirectional radiation patters of user equipment in previous telecommunication generations, better results were, in general, achievable with combinations of multiple access points (AP), and some evidences have been presented by the community to support the intrinsic limits of RSSI fingerprinting [95]. However, this was before the introduction of beamforming which is a key feature of our used 5G NR mmWave validation system **5gchampion** and a previous work showed that the methodology of the proposed framework is promising [96], thus we decided to extend it considering a more complex scenario.

The full CSI has been extensively used for sensing with good accuracy in tasks like indoor positioning [97], activity recognition [98], and even finer approaches for gesture recognition [99]. Despite its reliability, the use of full CSI information is still difficult because it is vendor-specific.

Other approaches in the literature consider modeling the environment [100], [101]. Despite they can achieve good results, those are obtained at the price of increased complexity because the factors that influence the communication must be included in the model. Then, they usually lack of generalizability due to the specific conditions of the modeled environment. Our approach is instead model-free.

Finally, some multi-modal solutions employes additional hardware to improve the beam alignment such as radars, lidars, or even cameras [102]–[104]. Despite multi-modality being an opportunity for future systems, there are still some open problems related to data fusion, such as hardware synchronization and the need to maintain multiple devices over time. Moreover, in the case of cameras, privacy issues have to be taken into account. In this work, the use of an already-existing 5g mmWave communication link simplifies the hardware need and the related possible issues, providing a good solution, while multi-modal approaches can be necessary for solving more complex tasks.

5.3 Proposed system

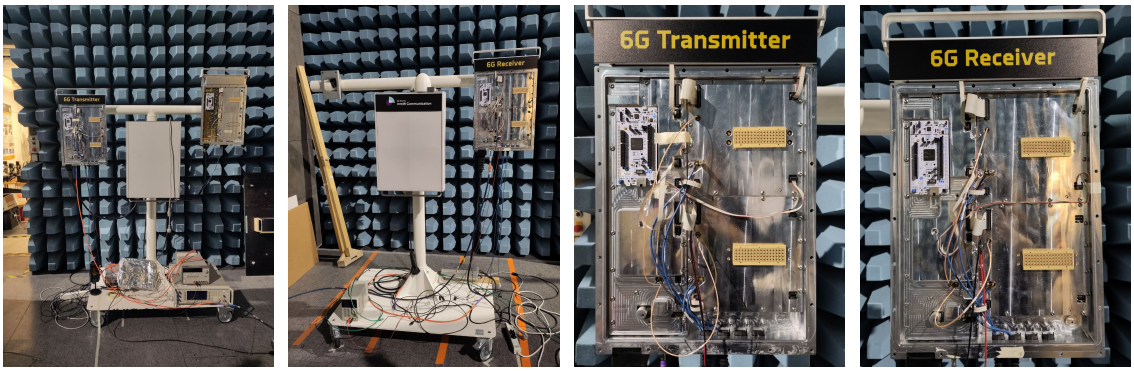
The 5G New Radio (NR) beamforming protocol starts with a four-stage Synchronisation Signal (SS) block-based Initial Access (IA) followed by different random beam selection procedures such as exhaustive, hierarchical and other fast beam alignment algorithms [105]. Under this protocol, each SS burst followed by beam selection appears periodically after 20ms [106]. The beam-selection process during communication essentially provides spatial information of the environment as the side information. Such information is helpful to observe dynamic changes in the

radio environment over multiples of these periodic bursts during communication and to sense human static positions, blockages, activities, etc. over time.

In this work, we employ the traditional exhaustive beam-search methodology to scan human static positions in mmWave RF environment and capture the RSSI characteristics at IF around 4 GHz using a Vector Network Analyzer (VNA). We defined the RSSI characteristic information captured during each beam-selection procedure as RSSI fingerprint and it is used to estimate the indoor human static positions near the radio link during communication.

5.3.1 Experimental hardware

We consider an indoor environment consisting of mmWave 5G radios, VNA, and a camera. The used 5G mmWave radio system [107] consists of TX and RX radio units located at ≈ 4m distance from one another at two corners of the room (Figure 5.1). Both RX and TX units are connected to the VNA which can measure both the amplitude and phase responses seen by the RX radio unit. The VNA is an RF measurement device used to measure amplitude and phase responses. In this work, we used the Keysight VNA N5247 device to measure the signal responses at RX and captured their s-parameters. A Microsoft LifeCam camera is mounted on the rooftop for the aerial view to capture the human static locations in the indoor environment as shown in Figure 5.2. The captured camera information is used to verify the subjects' adherence to the experimental procedure. We note here that the antennas at both TX and RX are at a fixed height of 162 cm from the ground.



(a) **(b)** **(c)** **(d)**

Figure 5.1: (a)-(b) TX and RX with their supports in the room. (c)-(d) TX and RX boards.

5.3.2 Experiments design

In Figure 5.2, the set-up is represented from the top-view. The distances d_1 and d_2 measure respectively 45 cm and 69.5 cm. The angle that the beam forms with respect to the Line-of-Sight (LoS) follows the sign convention indicated in Figure 5.2, i.e. assumes a negative sign if pointing

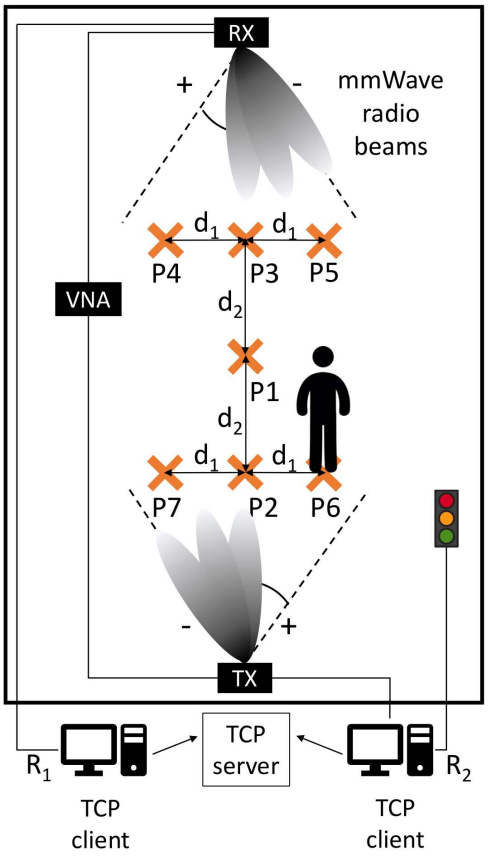


Figure 5.2: Schematic of the set-up from the top-view. The numbered positions represent the different classes, i.e. positions.

to the left of the device and viceversa. In this work, we refer to LoS positions to indicate P1, P2, and P3 (Figure 5.2) that table are along the LoS, thus where the angle is $0°$.

A MATLAB interface is used to synchronize and select beam steering directions of TX and RX using computing resources R_1 and R_2, respectively (Figure 5.2). For both RX and TX, the considered steering angles are $[-25°, +25°]$ with a resolution of $5°$, resulting in 11 directions for each link side. In this work, we chose to implement the exhaustive beam-search scanning protocol to distinguish the human static positions in mmWave RF environment capturing the RSSI characteristics at RX using VNA. Thus, the RX angle is fixed and the TX scans all the considered angles and then it is repeated for a new RX direction until the end of all the 121 combinations of beam directions. The entire procedure is repeated for each of the frequencies of interest by the VNA. A set of orange position markers is arranged between the two radio transceivers to indicate to the subjects the positions to occupy during the measurements (Figure 5.2).

In Table 5.1 are reported the beam angles for both TX and RX for each position marker. We note that we set the resolution to 5 degrees but the human body is wider than the marker size; d_1 is comparable with the average distance between the shoulders of the volunteers. So we expect

Angle	P1	P2	P3	P4	P5	P6	P7
TX angle (°)	0	0	0	-10	10	22	-22
RX angle (°)	0	0	0	22	- 22	-10	10

Table 5.1: TX and RX beam angles for each position.

to observe a human-induced impact even in some intermediate angular positions, i.e. 22 . In this work, all markers are symmetrically placed (Figure 5.2) by measuring their angular positions with respect to TX and RX radio units. The experiments are designed to verify the applicability of the method in order to distinguish symmetric positions with respect to the LoS and to the perpendicular line (Table 5.2). However, the proposed method is not limited to this assumption and can be extended to different indoor human movements in the future. Moreover, our interest is also in the impact of the blockage along the LoS and the capability of our system to distinguish the empty environment.

Experiment	Classes (positions)	Description
E1a **E1b**	4-5 6-7	Symmetric w.r.t. LoS
E2a **E2b**	3-4-5 2-6-7	Symmetric w.r.t. LoS considering LoS positions
E3a **E3b**	4-5-ER 6-7-ER	Symmetric w.r.t. LoS considering empty room
E4a **E4b**	5-6 4-7	Same side w.r.t. LoS
E5a **E5b**	1-5-6 1-4-7	Same side w.r.t. LoS considering LoS positions
E6a **E6b**	5-6-ER 4-7-ER	Same side w.r.t. LoS considering empty room
E7	1-4-5-6-7-ER	LoS, empty room and all nLoS positions

Table 5.2: Experiments description. ER is the empty room.

5.3.3 Data collection

The dataset consists of radio signals collected from 28 volunteers with each subject standing in the pre-defined set of positions between the TX and RX (Figure 5.2). RGB videos while the subject is performing the task were also collected to further verify the adherence of the subject to the procedure and retrieve the standing orientation. The volunteers were 19 males and 9 females and the following features are expressed as average $\pm$ one standard deviation unit: age 34 ± 11 years, height 172.8 ± 9.4 cm, and distance between shoulders 45.3 ± 3.6 cm. The distance between shoulders is a measure of the width of the volunteer's body. Each volunteer signed an informed consent form before data acquisition.

The measuring procedure consists in controlling the TX and RX radio units to perform an ex-

haustive beam search (Section 5.3.1) for each standing position and the whole procedure is repeated three times per each volunteer. During the acquisition, the measurement environment was isolated from external factors by closing the door, thus to communicate to the volunteer about the position to occupy, we used a traffic light and an associated color code. The volunteer could freely decide their standing orientation, thus leading to higher variability in the dataset, in fact, the subject standing in the same position can have a different impact on the RX signal due to the body orientation. The only condition we required was to be able to check the traffic light color.

Furthermore, the RX signal during the exhaustive beam search was collected for the empty room. We considered a trial as incorrect if the subject did not perform according to the procedure (i.e. standing for the whole exhaustive beam search procedure in the same position). Moreover, the data corresponding to two of the subjects were discarded because the TX beamforming was not correctly set due to an error in the control system. After discarding the incorrect trials, the dataset is composed of 69 examples for each class, except for the empty room which has 66 examples.

5.3.4 Pre-processing

For each pre-processing approach, the first step is the extraction of the RSSI signal as follows.

$$\mathrm{RSSI} = 10\log_{10}\|V\|^2 = 10\log_{10}\left(I^2 + Q^2\right) \tag{5.1}$$

where V denote the received signal amplitude, I and Q are the real and imaginary coefficient of it, $\|.\|$ is the L_2 norm, and $.^2$ is the square power. Each RX signal has 121 samples corresponding to the 121 combinations of TX and RX beam angles scanned in time, so each sample corresponds to a signal measurement.

Time Domain Fingerprints

The RSSI in the time domain carries the spatial information due to the design of the experiment. We refer to this signal as "direct fingerprints". Figure 5.3 shows the RSSI signal for the different classes and it can be noticed that for LoS positions (1,2,3) the fingerprints present more variability, while the cleanest signal is in the case of the empty room. For the other classes, the most noticeable perturbation of the signal corresponds to the subject presence, e.g. for position 4 (RX angle equal to 22°) it is in correspondence of the last two main peaks, i.e. RX angle equal to 20° and 25°. Due to the symmetric role played by TX and RX, we considered also the signal artificially reordered to invert the scanning roles between RX and TX, thus resulting in the TX fixed with RX scanning all the considered directions and then repeated for each beamforming angle at the TX. We refer to this signal as "inverse fingerprints".

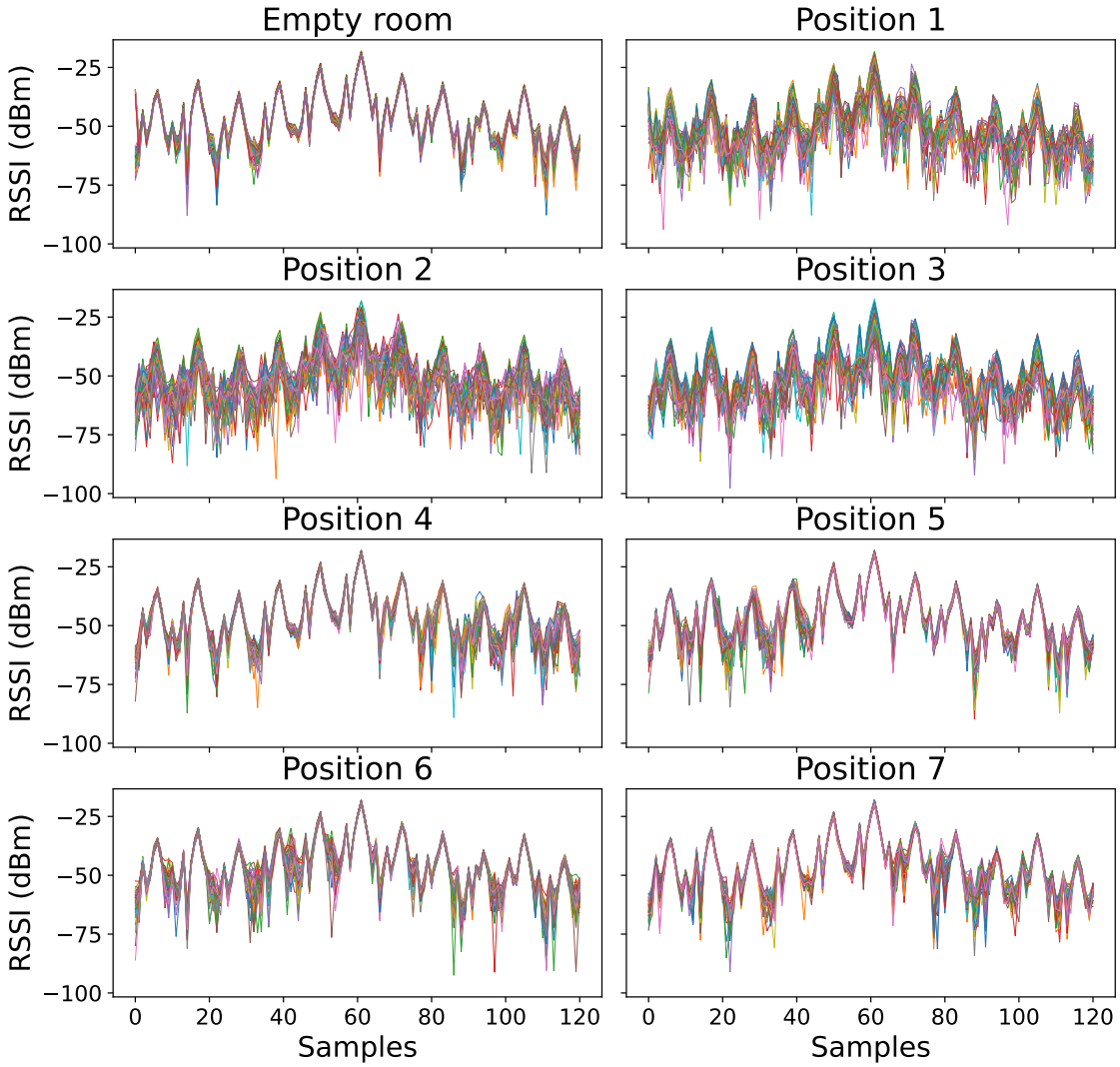


Figure 5.3: One exhaustive search for each position in the time domain. For better visualization, the plot represents only the instances considered for the training in the first fold.

FFT-based Fingerprints

Following this approach, we considered as fingerprints the coefficients obtained through Fast Fourier Transform (FFT) applied to the signal in the time domain. The FFT was computed after applying Hanning windowing with window size corresponding to the lenght of the signal to reduce the effect of side oscillations. To build the fingerprints, real and imaginary coefficients were concatenated to create a single-channel vector or a double-channels vector, adjusting the learning framework in accordance with this choice (Section 5.3.5). We tested a number of coefficients vary-

ing between $[10, 60]$ with a step of 10 coefficients to search for an appropriate trade-off between the amount of information carried by coefficients and the complexity of the classification model.

RSSI Difference

the fingerprints were built with coefficients obtained after windowing and FFT, as in the previous method with FFT-based fingerprints, over the changes in the RSSI signal between subsequent samples. The real and imaginary coefficients were considered either concatenated to create a single-channel vector or a double-channel vector and the learning model was then adapted to the shape of the fingerprint (Section 5.3.5).

Wavelet Transform

for each instance, the fingerprint is the RSSI signal converted to the scale domain through the application of the Continous Wavelet Transform (CWT)[108] resulting in a set of coefficient for each considered scale. Different types of wavelets were considered ("mexh", "moral", "gaus1", "shan") and the relevant scales were found to belong to [0, 50]. The coefficients obtained at the different scales were considered as separate channels or as pixels of an image and the learning framework was adjusted accordingly (Section 5.3.5).

5.3.5 Learning framework

In this work, we adopted a supervised approach with the classes' labels obtained by the acquisition procedure and verified through the inspection of the top-view videos. The fingerprints obtained after pre-processing are used to feed a neural network, the 10% is reserved for final testing, while the remaining 90% is used to train the model with a 9-fold cross-validation strategy. The folds are obtained through StratifiedKfold from scikit-learn Python package, a tool designed to preserve the percentages of the classes in each fold with respect to the complete dataset. In our case the dataset is balanced, thus it prevents the selection of unbalanced folds. The number of folds is chosen to obtain the validation fold with a similar size to the final test set. While the other hyperparameter choices were tuned accordingly to the results obtained on the validation set.

We considered feed-forward neural networks and combinations with 1D and 2D convolutional layers. Depending on the pre-processing approaches, the fingerprints result in different shapes, thus the learning model was adapted to them. In particular, with 1D features the learning layers implied were feed-forward or 1D convolutions. For 2D input data, the convolutional layers were 1D with multiple channels or 2D convolutional layers with a single channel. If convolutional layers were included, they were followed by max pooling layers. We tested also the inclusion of Dropout and Random Gaussian Noise layers. In all the models we tested, the activation function was rectified linear unit (ReLu) for all the layers but for the predictions, where Softmax was used. Then, the loss was the categorical cross-entropy and the model was updated using Adam optimizer with a learning rate (lr) equal to 0.0001, with the exception of models implying 2-D convolutions where we adopted lr = 0.001. For the implementation, we used Tensorflow. Among

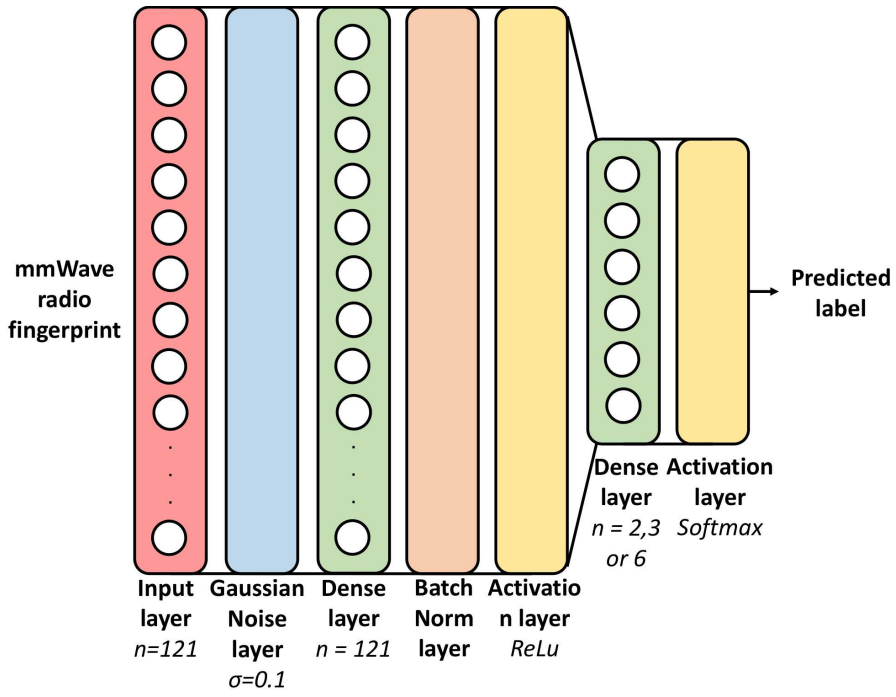


Figure 5.4: Proposed feed-forward neural network (FFNN).

the considered, we found the best combination of pre-processing approach and learning model to be the use of time-domain direct fingerprints with a feed-forward neural network built with the following layers (Figure 5.4): Input layer, GaussianNoise layer (standard deviation equal to 0.1), Dense layer (same size of the input), BatchNormalization layer, Activation layer (ReLu), Dense layer (size equal to the number of considered classes).

5.3.6 Metrics

In the evaluation of supervised classification problems, for each considered class, each instance can belong to the true positives (TP) group if the actual label and the predicted label are both belonging to the considered class. Instead, the instance belongs to false positives (FP) if the model predicts it to be an instance of the considered class but the actual label is different. Similarly, true negatives (TN) and false negatives (FN) are defined. The metrics used for the model evaluation are accuracy, recall, precision and F1-score, defined using the amounts of positive and negative predictions.

Accuracy (ACC) expresses the fraction of overall correctly classified instances in a specific class and is defined as follows:

$$\mathrm{ACC_c} = \frac{\mathrm{TP_c} + \mathrm{TN_c}}{\mathrm{TP_c} + \mathrm{FP_c} + \mathrm{FN_c} + \mathrm{TN_c}} \tag{5.2}$$

where c is the class index.

Recall (REC) measures the ratio of samples correctly predicted as positive with respect to the to-

tal number of actual positives, while precision (PRE) is the ratio of the correct positive predictions to the total number of positive predictions:

$$\mathrm{REC_c} = \frac{\mathrm{TP_c}}{\mathrm{TP_c} + \mathrm{FN_c}} \quad \text{and} \quad \mathrm{PRE_c} = \frac{\mathrm{TP_c}}{\mathrm{TP_c} + \mathrm{FP_c}} \tag{5.3}$$

Finally, F1-score is the harmonic mean of precision and recall

$$\mathrm{F1_c} = \frac{2\mathrm{PRE_c}\mathrm{REC_c}}{\mathrm{PRE_c} + \mathrm{REC_c}} \tag{5.4}$$

In this multi-class classification context, the overall metrics are computed as the average among the scores obtained for each predicted classes (Table 5.3). The F1-score computed as the average across the F1-scores of each classes is denoted in literature as macro F1-score [109].

5.4 Results

In this Section, we note the results referred to the best configuration of pre-processing and learning framework among the considered and described in Sections 5.3.4 and 5.3.5. The results we obtained suggest that the information retrieved from the RSSI signal during the exhaustive beam search protocol is sufficient for position prediction.

Considering the results we obtained over the validation set in the preliminar analysis for the classification of the positions, we found the frequency 27.502 GHz to be the most informative for this classification task, so the following analysis considers results obtained at that frequency.

We observed that the classification of the LoS position (P1, P2 or P3) is usually the sharpest with respect to other classes (Figure 5.5d, 5.5i, 5.5j, 5.5m) with the exception of experiment E2a where P3 is the only label with one misclassification (Figure 5.5c). The empty room shows an opposite behavior with respect to the LoS positions, in fact, it is, in general, one of the most misclassified classes (Figure 5.5e, 5.5f, 5.5l, 5.5m) with the exception of E6a (Figure 5.5k). This suggests that the impact of the blockage along the LoS is easier to be recognized by the model with respect to the empty room or the other position patterns. This is something we could expect from the different appearances of the fingerprints (Figure 5.3).

Some experiments are highly symmetric, we call them "paired experiments". Then, we observed that, in general, the paired experiments (a) - (b) results show less than 10% difference for each metric which indicates in this work that one of the paired experiments (a) - (b) misclassifies, in the worst case, one more instance with respect to the other (Figure 5.5). Experiments E6 are the only exception with more misclassifications in version E6a (Table 5.3, Figure 5.5k-5.5l). However, experiment E7 is the most complete and shows good performances in recognizing all classes with the highest number of errors due to some false positives for the empty room (Figure 5.5m, Table 5.3).

It is interesting to remember that the subjects' orientation was left as a personal choice to the volunteers, with the only condition they could see the traffic light. Position P4 was observed

Experiment	Accuracy	Recall	Precision	F1-score
Experiment 1a	1.00	1.00	1.00	1.00
Experiment 1b	0.93	0.93	0.94	0.93
Experiment 2a	0.97	0.95	0.96	0.95
Experiment 2b	0.94	0.90	0.93	0.90
Experiment 3a	0.94	0.90	0.91	0.90
Experiment 3b	0.97	0.95	0.96	0.95
Experiment 4a	0.93	0.93	0.94	0.93
Experiment 4b	0.93	0.93	0.94	0.93
Experiment 5a	0.94	0.90	0.90	0.90
Experiment 5b	0.97	0.95	0.96	0.95
Experiment 6a	0.81	0.71	0.70	0.70
Experiment 6b	0.94	0.90	0.91	0.90
Experiment 7	0.95	0.86	0.88	0.86

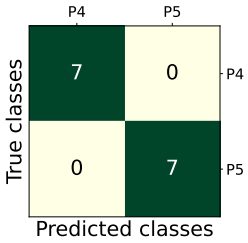
Table 5.3: Overall classification metrics for test set computed as average with respect to the scores obtained for each class.

from the videos to be the one in which subjects chose more different standing orientations but the classification of position P4 seems to be not penalized with respect to the other classes and, in general, the subject position was retrieved correctly with good performances (Table 5.3) also allowing free orientation. Thus, our proposed framework can capture information related to the position excluding the orientation factor.

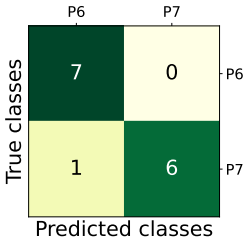
5.5 Conclusions and future directions

In this work, we proposed a novel learning-based solution for human position estimation with the purpose of improving also the communication itself in the context of mmWave communications. The design of this solution exploits a feature already implemented in the IA of the communication procedure and the information provided by the RSSI signal. We build the experimental dataset by designing and performing the data collection described in the previous Sections and we analyzed them considering multiple pre-processing approaches and learning procedures for the classification.

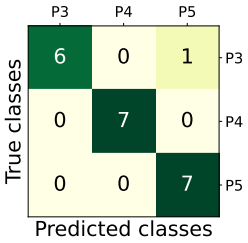
Our results suggested that the RSSI in directional communications contains enough information to retrieve indoor positioning with a resolution of positions comparable to subject dimensions which was obtained measuring the distance between the shoulders. The future work will include more realistic scenarios from the communication side, for example using universal software radio peripheral reproducing Orthogonal Frequency Division Multiple access (OFDM) communication schemes. Then, the exhaustive beam-searching protocol is an experimental choice, thus other protocols, and their correspondent fingerprints could be explored. Moreover, a more realistic environment can include multiple subjects and non-static activities.



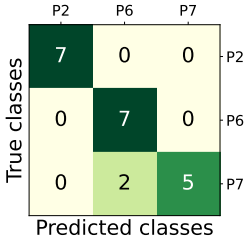
(a) Experiment E1a



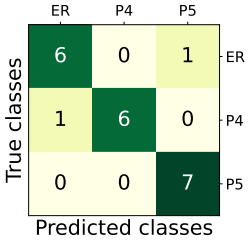
(b) Experiment E1b



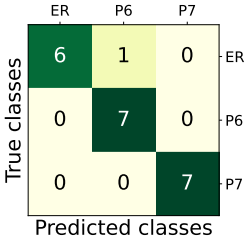
(c) Experiment E2a



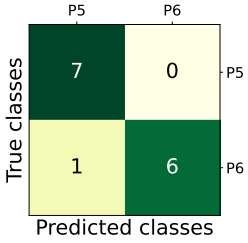
(d) Experiment E2b



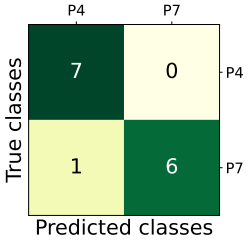
(e) Experiment E3a



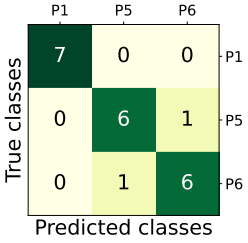
(f) Experiment E3b



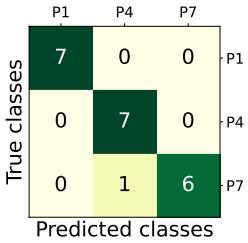
(g) Experiment E4a



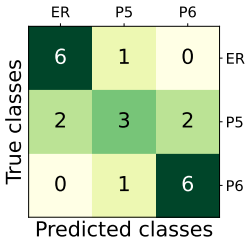
(h) Experiment E4b



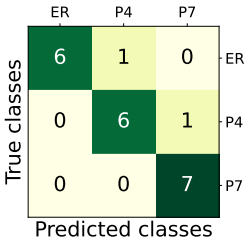
(i) Experiment E5a



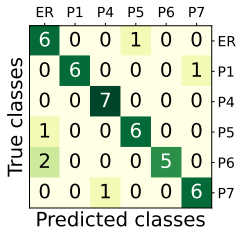
(j) Experiment E5b



(k) Experiment E6a



(l) Experiment E6b



(m) Experiment E7

Figure 5.5: Test set confusion matrices for each experiment.

6

Conclusions

Sensing is expected to evolve and continuously change different aspects of our society. Different sensing perspectives and applications have been analyzed and explored in this thesis.

First, we analyzed the application of deep learning tools to images and video analysis for biomechanical sensing. The opportunity and limitations related to this approach have been presented and we underlined the importance of the collected dataset to create a measurement tool based on deep-learning. Indeed, the quality and quantity of data is of paramount importance for the prediction of the subject's keypoints.

Then, the same keypoints estimation problem was tackled combining radar sensing with neural networks. Despite radar sensing being a promising approach some technical limitations are still challenging for the realization of a clinically oriented tool. However, the radar sensing approach opens the doors to a completely new type of sensing for gait analysis. This project is a step towards a profitable contamination between different research areas. The perspective is to exploit the contamination to develop innovative solutions to complex problems.

In Chapter 4, we explored the data acquisition platforms including wearable sensors. The multimodal acquisitions enable the possibility to tackle challenging problems due to the combinations of different perspectives. In both the proposed systems this aspect is crucial for the design and the time synchronization problem is accounted and discussed. The temporal correlations between signals are an important aspect for the extraction of information about context and the synchronization error could influence the evaluation of the situation.

Finally, a communications-based framework for human sensing is proposed, experimented, and discussed. Next-generation communications networks are expected to be more integrated with sensing features and the possibility to detect human positioning indoors through the communication hardware creates the possibility for new applications.

6.1 Future directions

In this thesis, we explored the synergies between different research areas with the aim to address complex challenges. In fact, the efficacy of sensing can benefit from a multidisciplinary approach, as it brings together expertise from diverse fields and, besides the progress within the individual fields, a multidisciplinary approach can contribute to cross-disciplinary breakthroughs that have the potential to change entire research and industrial sectors. In the future, sensing research will definitely have to consider the great potential of combining different perspectives to address real problems that usually present a high degree of complexity.

Moreover, the data collected by the great amount of sensors has to be analyzed and interpreted to create knowledge about the context. To this aim, the advancement of deep learning techniques is closely linked to the technical ability to understand increasingly deeper and more complex mechanisms in a more precise and useful way for our society.

References

[1] U. Challita, H. Ryden, and H. Tullberg, "When machine learning meets wireless cellular networks: Deployment, challenges, and applications," IEEE Communications Magazine, vol. 58, no. 6, pp. 12–18, 2020.

[2] D. Sutherland, "The evolution of clinical gait analysis: Part ii kinematics," Gait Posture, vol. 16, no. 2, pp. 159–179, 2002.

[3] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," Computer vision and image understanding, vol. 81, no. 3, pp. 231–268, 2001.

[4] S. Corazza, E. Gambaretto, L. Mündermann, and T. P. Andriacchi, "Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications," IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 806–812, 2010.

[5] E. Ceseracciu, Z. Sawacha, and C. Cobelli, "Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: Proof of concept," 2014.

[6] M. Ben Gamra and M. A. Akhloufi, "A review of deep learning techniques for 2d and 3d human pose estimation," Image and Vision Computing, vol. 114, p. 104 282, 2021.

[7] T. Lin, M. Maire, S. J. Belongie, et al., "Microsoft COCO: common objects in context," CoRR, vol. abs/1405.0312, 2014. arXiv: 1405.0312.

[8] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2014.

[9] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," International journal of computer vision, vol. 87, no. 1, pp. 4–27, 2010.

[10] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in IJCAI'81: 7th international joint conference on Artificial intelligence, vol. 2, 1981, pp. 674–679.

[11] C. Tomasi and T. Kanade, "Detection and tracking of point," Int J Comput Vis, vol. 9, no. 137-154, p. 3, 1991.

[12] F. A. Magalhaes, Z. Sawacha, R. Di Michele, M. Cortesi, G. Gatta, and S. Fantozzi, "Effectiveness of an automatic tracking software in underwater motion analysis," Journal of sports science & medicine, vol. 12, no. 4, p. 660, 2013.

[13] M. Ye, X. Wang, R. Yang, L. Ren, and M. Pollefeys, "Accurate 3d pose estimation from a single depth image," in 2011 International Conference on Computer Vision, IEEE, 2011, pp. 731–738.

[14] T. Yu, Z. Zheng, K. Guo, et al., "Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7287–7296.

[15] C. Zheng, W. Wu, C. Chen, et al., "Deep learning-based human pose estimation: A survey," ACM Computing Surveys, vol. 56, no. 1, pp. 1–37, 2023.

[16] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep learning based 2d human pose estimation: A survey," Tsinghua Science and Technology, vol. 24, no. 6, pp. 663–676, 2019.

[17] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Proof-of-concept exploratory study markerless motion capture methodology based on deep learning," in Congress of European Society of Biomechanics (ESB), 2021.

[18] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Cnn-based markerless motion capture approach: A pilot study," in Congress of International Society of Biomechanics (ISB), 2021.

[19] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Smart motion capture based on deep learning: A proof-of-concept," in Congress of Società Italiana di Analisi del MOvimento in Clinica (SIAMOC), 2021.

[20] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Deep learning-based markerless motion analysis: A pilot study," in Annual meeting of the European Society for Movement analysis in Adults and Children (ESMAC), 2021.

[21] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Development of a deep-learning based markerless motion capture approach: A pilot study," in 9th World Congress of Biomechanics (WCB), 2021.

[22] A. Leardini, Z. Sawacha, G. Paolini, S. Ingrosso, R. Nativo, and M. G. Benedetti, "A new anatomically based protocol for gait analysis in children," Gait & posture, vol. 26, no. 4, pp. 560–571, 2007.

[23] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 11, pp. 1330–1334, 2000.

[24] I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions," in IJ Schoenberg Selected Papers, Springer, 1988, pp. 3–57.

[25] M. Harrington, A. Zavatsky, S. Lawson, Z. Yuan, and T. Theologis, "Prediction of the hip joint centre in adults, children, and patients with cerebral palsy based on magnetic resonance imaging," Journal of Biomechanics, vol. 40, no. 3, pp. 595–602, 2007.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

[27] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693–5703.

[28] P. Sturm, "Pinhole camera model," in Computer Vision: A Reference Guide, K. Ikeuchi, Ed. Boston, MA: Springer US, 2014, pp. 610–613.

[29] E. S. Grood and W. J. Suntay, "A Joint Coordinate System for the Clinical Description of Three-Dimensional Motions: Application to the Knee," Journal of Biomechanical Engineering, vol. 105, no. 2, pp. 136–144, May 1983. eprint: https://asmedigitalcollection.asme.org/biomechanical/article-pdf/105/2/136/5599866/136_1.pdf.

[30] A. Ferrari, A. G. Cutti, and A. Cappello, "A new formulation of the coefficient of multiple correlation to assess the similarity of waveforms measured synchronously by different motion analysis protocols," Gait & posture, vol. 31, no. 4, pp. 540–542, 2010.

[31] L. Michelon, "Human pose and skeleton reconstruction with deep neural networks from mmwave radar point clouds,"

[32] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," IEEE Signal Processing Magazine, vol. 34, no. 2, pp. 22–35, 2017.

[33] M. A. Richards, J. Scheer, W. A. Holm, and W. L. Melvin, "Principles of modern radar," 2010.

[34] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," IEEE Transactions on Aerospace and electronic systems, vol. 42, no. 1, pp. 2–21, 2006.

[35] T. Xu, D. An, Y. Jia, and Y. Yue, "A review: Point cloud-based 3d human joints estimation," Sensors, vol. 21, no. 5, p. 1684, 2021.

[36] E. Camuffo, D. Mari, and S. Milani, "Recent advancements in learning algorithms for point clouds: An updated overview," Sensors, vol. 22, no. 4, p. 1357, 2022.

[37] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "Mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," IEEE Sensors Journal, vol. 20, no. 17, pp. 10 032–10 044, 2020.

[38] M. Zhao, Y. Tian, H. Zhao, et al., "Rf-based 3d skeletons," in Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, 2018, pp. 267–281.

[39] A. Sengupta and S. Cao, "Mmpose-nlp: A natural language processing approach to precise skeletal pose estimation using mmwave radars," IEEE Transactions on Neural Networks and Learning Systems, 2022.

[40] S.-P. Lee, N. P. Kini, W.-H. Peng, C.-W. Ma, and J.-N. Hwang, "Hupr: A benchmark for human pose estimation using millimeter wave radar," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 5715–5724.

[41] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," Advances in neural information processing systems, vol. 30, 2017.

[42] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.

[43] H. Fan and Y. Yang, "Pointrnn: Point recurrent neural network for moving point cloud processing," arXiv preprint arXiv:1910.08287, 2019.

[44] A. Leardini, Z. Sawacha, G. Paolini, S. Ingrosso, R. Nativo, and M. G. Benedetti, "A new anatomically based protocol for gait analysis in children," Gait & posture, vol. 26, no. 4, pp. 560–571, 2007.

[45] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 4, pp. 2994–3009, 2020.

[46] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in kdd, vol. 96, 1996, pp. 226–231.

[47] P. Zhao, C. X. Lu, J. Wang, et al., "Mid: Tracking and identifying people with millimeter wave radar," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), IEEE, 2019, pp. 33–40.

[48] Z. Meng, S. Fu, J. Yan, et al., "Gait recognition for co-existing multiple people using millimeter wave sensing," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 849–856.

[49] T. Wagner, R. Feger, and A. Stelzer, "Radar signal processing for jointly estimating tracks and micro-doppler signatures," IEEE Access, vol. 5, pp. 1220–1238, 2017.

[50] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, Mar. 1960. eprint: https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf.

[51] D. Lerro and Y. Bar-Shalom, "Tracking with debiased consistent converted measurements versus ekf," IEEE Transactions on Aerospace and Electronic Systems, vol. 29, no. 3, pp. 1015–1022, 1993.

[52] H. W. Kuhn, "The hungarian method for the assignment problem," Naval Research Logistics (NRL), vol. 52, no. 1, pp. 7–21, 2005.

[53] R. J. Fitzgerald, "Development of practical pda logic for multitarget tracking by microprocessor," in 1986 American Control Conference, 1986, pp. 889–898.

[54] C. M. Bishop, Neural networks for pattern recognition. Oxford university press, 1995.

[55] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[56] Z. Wu, S. Song, A. Khosla, et al., "3d shapenets: A deep representation for volumetric shapes," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1912–1920.

[57] N. K. Gupta, Inside Bluetooth low energy. Artech House, 2016.

[58] D. O. Camacho, "Redfang: A high-level bluetooth library for building distributed android applications," in 2015 Asia-Pacific Conference on Computer Aided System Engineering, IEEE, 2015, pp. 337–341.

[59] L. Guo-Cheng and Y. Hong-Yang, "Design and implementation of a bluetooth 4.0-based heart rate monitor system on ios platform," in 2013 International Conference on Communications, Circuits and Systems (ICCCAS), IEEE, vol. 2, 2013, pp. 112–115.

[60] P. Pierleoni, L. Pernini, A. Belli, and L. Palma, "An android-based heart monitoring system for the elderly and for patients with heart disease," International journal of telemedicine and applications, 2014.

[61] Y. A. Djawad, S. Suhaeb, H. Jaya, et al., "Development of an intelligent mobile health monitoring system for the health surveillance system in indonesia," IRBM, vol. 42, no. 1, pp. 28–34, 2021.

[62] K. Malhi, S. C. Mukhopadhyay, J. Schnepper, M. Haefke, and H. Ewald, "A zigbee-based wearable physiological parameters monitoring system," IEEE sensors journal, vol. 12, no. 3, pp. 423–430, 2010.

[63] P. Valsalan, T. A. B. Baomar, and A. H. O. Baabood, "Iot based health monitoring system," Journal of critical reviews, vol. 7, no. 4, pp. 739–743, 2020.

[64] Y.-j. Park and H.-s. Cho, "Transmission of ecg data with the patch-type ecg sensor system using bluetooth low energy," in 2013 International Conference on ICT Convergence (ICTC), IEEE, 2013, pp. 289–294.

[65] K. Shahzad and B. Oelmann, "A comparative study of in-sensor processing vs. raw data transmission using zigbee, ble and wi-fi for data intensive monitoring applications," in 2014 11th International Symposium on Wireless Communications Systems (ISWCS), IEEE, 2014, pp. 519–524.

[66] A. M. Chan, N. Selvaraj, N. Ferdosi, and R. Narasimhan, "Wireless patch sensor for remote monitoring of heart rate, respiration, activity, and falls," in 2013 35th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2013, pp. 6115–6118.

[67] S. Zampato, C. A. Bernardini, Z. Sawacha, and M. Rossi, "Open-mbic: An open-source android library for multiple simultaneous bluetooth low energy connections," in 2022 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT), 2022, pp. 333–337.

[68] S. Zampato, C. A. Bernardini, M. Rossi, and Z. Sawacha, "Remocop - a home telemedicine system for rehabilitation monitoring of covid-19 survivors and chronic obstructive pulmonary disease patients," in 9th World Congress of Biomechanics (WCB), 2022.

[69] S. Zampato, C. A. Bernardini, M. Rossi, and Z. Sawacha, "Open-mbic: An open-source android library for multiple simultaneous bluetooth low energy connections," in Annual meeting of Gruppo Telecomunicazioni e Tecnologie dell'Informazione (GTTI), 2022.

[70] E. Pegolo, M. Romanato, S. Zampato, et al., "A novel experimental paradigm to investigate neural and muscular activities during normal gait: Proof of concept coupling eeg, semg and kinematics measures," Gait Posture, vol. 105, S35–S36, 2023.

[71] M. G. Benedetti, E. Beghi, A. De Tanti, et al., "Siamoc position paper on gait analysis in clinical practice: General requirements, methods and appropriateness. results of an italian consensus conference," Gait Posture, vol. 58, pp. 252–260, 2017.

[72] R. Baker, "Gait analysis methods in rehabilitation," Journal of NeuroEngineering and Rehabilitation, vol. 3, 2006, Cited by: 347; All Open Access, Gold Open Access, Green Open Access.

[73] H. D. Eberhart, "Fundamental studies of human locomotion and other information relating to the design of artificial limbs," A Report to the National Research Council, 1947.

[74] H. Adeli and S. Ghosh-Dastidar, Automated EEG-based diagnosis of neurological disorders: Inventing the future of neurology. CRC press, 2010.

[75] T. D'Alessio and S. Conforto, "Extraction of the envelope from surface emg signals," IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 6, pp. 55–61, 2001.

[76] M. Meng, Q. She, Y. Gao, and Z. Luo, "Emg signals based gait phases recognition using hidden markov models," in The 2010 IEEE International Conference on Information and Automation, 2010, pp. 852–856.

[77] M. Romanato, W. Piatkowska, F. Spolaor, D.-K. To, D. Volpe, and Z. Sawacha, "Different perspectives in understanding muscle functions in parkinson's disease through surface electromyography: Exploring multiple activation patterns," Journal of Electromyography and Kinesiology, vol. 64, p. 102 658, 2022.

[78] P. Wei, J. Zhang, F. Tian, and J. Hong, "A comparison of neural networks algorithms for eeg and semg features based gait phases recognition," Biomedical Signal Processing and Control, vol. 68, p. 102 587, 2021.

[79] C. D. Joshi, U. Lahiri, and N. V. Thakor, "Classification of gait phases from lower limb emg: Application to exoskeleton orthosis," in 2013 IEEE Point-of-Care Healthcare Technologies (PHT), 2013, pp. 228–231.

[80] R. Luo, S. Sun, X. Zhang, Z. Tang, and W. Wang, "A low-cost end-to-end semg-based gait sub-phase recognition system," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 1, pp. 267–276, 2020.

[81] W. J. Piatkowska, F. Spolaor, M. Romanato, et al., "A supervised classification of children with fragile x syndrome and controls based on kinematic and semg parameters," Applied Sciences, vol. 12, no. 3, p. 1612, 2022.

[82] S. Tortora, L. Tonin, C. Chisari, S. Micera, E. Menegatti, and F. Artoni, "Hybrid human-machine interface for gait decoding through bayesian fusion of eeg and emg classifiers," Frontiers in Neurorobotics, vol. 14, p. 582 728, 2020.

[83] S. Nakagome, T. P. Luu, Y. He, A. S. Ravindran, and J. L. Contreras-Vidal, "An empirical comparison of neural networks and machine learning algorithms for eeg gait decoding," Scientific reports, vol. 10, no. 1, p. 4372, 2020.

[84] D. De Venuto, V. Annese, G. Defazio, V. Gallo, and G. Mezzina, "Gait analysis and quantitative drug effect evaluation in parkinson disease by jointly eeg-emg monitoring," in 2017 12th International Conference on Design Technology of Integrated Systems In Nanoscale Era (DTIS), 2017, pp. 1–6.

[85] G. Cisotto, M. Capuzzo, A. V. Guglielmi, and A. Zanella, "Feature selection for gesture recognition in internet-of-things for healthcare," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–6, 2020.

[86] J. Tryon, E. Friedman, and A. L. Trejos, "Performance evaluation of eeg/emg fusion methods for motion classification," in 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada: IEEE Press, 2019, pp. 971–976.

[87] J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, no. 6, pp. 679–698, 1986.

[88] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," IEEE Comm. Magazine, 2020.

[89] B. Dong, V. Prakash, F. Feng, and Z. O'Neill, "A review of smart building sensing system for better indoor environment control," Energy and Buildings, 2019.

[90] J. Al Dakheel, C. Del Pero, N. Aste, and F. Leonforte, "Smart buildings features and key performance indicators: A review," Sustainable Cities and Society, 2020.

[91] M. Benzaghta, B. Y. Gokdogan, R. B. Coruk, and A. Kara, "Modeling and measurement of human body blockage loss at 28GHz," Jour. of Electromag. Waves and Applications, 2023.

[92] B. Ji, Y. Han, S. Liu, et al., "Several key technologies for 6G: challenges and opportunities," IEEE Comm. Standards Magazine, 2021.

[93] L.-H. Shen, K.-T. Feng, and L. Hanzo, "Five Facets of 6G: Research Challenges and Opportunities," Assoc. for Comp. Machinery, November 2023.

[94] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless RSSI fingerprinting localization," Signal Processing, 2017.

[95] E. Elnahrawy, X. Li, and R. Martin, "The limits of localization using signal strength: a comparative study," in 2004 First Annual IEEE Commun. Society Conf. on Sensor and Ad Hoc Commun. and Net., 2004. IEEE SECON 2004., 2004.

[96] P. Susarla, M. Jokinen, N. Tervo, et al., "Learning human-blockage direction prediction from indoor mmwave radio measurements," in 2023 IEEE Int. Conf. on Commun. Workshops (ICC Workshops), 2023.

[97] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. R. Liu, "The Promise of Radio Analytics: A Future Paradigm of Wireless Positioning, Tracking, and Sensing," IEEE Sig. Proc. Magazine, 2018.

[98] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," IEEE Commun. Magazine, vol. 55, no. 10, pp. 98–104, 2017.

[99] Y. Zheng, Y. Zhang, K. Qian, et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in Proceedings of the 17th annual int. conf. on mobile systems, app., and services, 2019.

[100] W. Qi, J. Huang, J. Sun, Y. Tan, C.-X. Wang, and X. Ge, "Measurements and modeling of human blockage effects for multiple millimeter wave bands," in 2017 13th Int. Wireless Comm. and Mobile Comput. Conf. (IWCMC), IEEE, 2017.

[101] U. T. Virk and K. Haneda, "Modeling human blockage at 5g millimeter-wave frequencies," IEEE Trans. on Antennas and Propagation, 2020.

[102] D. Marasinghe, N. Rajatheva, and M. Latva-aho, "Lidar aided human blockage prediction for 6g," in 2021 IEEE Globecom Workshops (GC Wkshps), IEEE, 2021.

[103] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in 2020 IEEE 91st veh. technol. conf. (VTC2020-Spring), IEEE, 2020.

[104] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," IEEE Signal Processing Magazine, 2019.

[105] P. Susarla, B. Gouda, Y. Deng, M. Juntti, O. Silven, and A. Tölli, "Learning-based Beam Alignment for Uplink mmWave UAVs," IEEE Trans. on Wireless Comm., 2022.

[106] E. Dahlman, S. Parkvall, and J. Skold, 5G NR: The next generation wireless access technology. Academic Press, 2020.

[107] M. E. Leinonen, M. Jokinen, N. Tervo, O. Kursu, and A. Pärssinen, "System evm characterization and coverage area estimation of 5g directive mmw links," IEEE Trans. on Microw. Theory and Techniques, 2019.

[108] O. Rioul and M. Vetterli, "Wavelets and signal processing," IEEE Sig. Proc. Magazine, 1991.

[109] J. Opitz and S. Burst, Macro f1 and macro f1, 2021.

List of publications

Conferences - orals and posters

[17] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Proof-of-concept exploratory study markerless motion capture methodology based on deep learning," in Congress of European Society of Biomechanics (ESB), 2021.

[18] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Cnn-based markerless motion capture approach: A pilot study," in Congress of International Society of Biomechanics (ISB), 2021.

[19] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Smart motion capture based on deep learning: A proof-of-concept," in Congress of Società Italiana di Analisi del MOvimento in Clinica (SIAMOC), 2021.

[20] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Deep learning-based markerless motion analysis: A pilot study," in Annual meeting of the European Society for Movement analysis in Adults and Children (ESMAC), 2021.

[21] S. Zampato, A. Bouleimen, F. Piemontese, et al., "Development of a deep-learning based markerless motion capture approach: A pilot study," in 9th World Congress of Biomechanics (WCB), 2021.

[67] S. Zampato, C. A. Bernardini, Z. Sawacha, and M. Rossi, "Open-mbic: An open-source android library for multiple simultaneous bluetooth low energy connections," in 2022 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT), 2022, pp. 333–337.

[68] S. Zampato, C. A. Bernardini, M. Rossi, and Z. Sawacha, "Remocop - a home telemedicine system for rehabilitation monitoring of covid-19 survivors and chronic obstructive pulmonary disease patients," in 9th World Congress of Biomechanics (WCB), 2022.

[69] S. Zampato, C. A. Bernardini, M. Rossi, and Z. Sawacha, "Open-mbic: An open-source android library for multiple simultaneous bluetooth low energy connections," in Annual meeting of Gruppo Telecomunicazioni e Tecnologie dell'Informazione (GTTI), 2022.

[70] E. Pegolo, M. Romanato, S. Zampato, et al., "A novel experimental paradigm to investigate neural and muscular activities during normal gait: Proof of concept coupling eeg, semg and kinematics measures," Gait Posture, vol. 105, S35–S36, 2023.

[110] M. Lecci, T. Zugno, S. Zampato, and M. Zorzi, "A full-stack open-source framework for antenna and beamforming evaluation in mmwave 5g nr," in ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1–6.

[111] S. S. Krishna Chaitanya Bulusu, S. Zampato, P. Susarla, N. Tervo, A. Pärssinen, and O. Sílven, "Hardware-friendly power amplifier linearization in next-generation broadcasting systems," in 2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 2023, pp. 299–304.

Acknowledgments

I thank Michele for leading me with kindness and respect in this path of professional and personal growth.
Zimi for challenging me with research problems that made me grow.
Giulia for being an older sister with advices and sharing experiences.
Marco for sharing all the steps, especially the deadlines.
Francesca, Jacopo e Giovanni, always a step ahead and sources of insipiration, but even more good mates.
I thank the whole SIGNET group that is so lively and all the woking neighbours from Mian, 219, ”sgabuzzino”, LTTM, and different rooms. A lot of different faces and personal stories that became part of my everyday life in these years. Moreover, I think it is not only about the individuals, but more about the community vibes shared in front of a life-saving coffee.
The whole Biomov lab that welcomed me even in the Christmas traditions. A special thank to Elena for the constant attempt to face together our common destiny.
And all the students I could guide during their thesis period because they always taught me something.
I thank all people I met during my research experience in Oulu. They had a strong impact on my path as a researcher and as an Italian in the cold Finland.
I thank the whole DEI, a place where the humanity and great research passion are strongly linked, especially when a bottle of wine is open in PalaDEI. Because you can always find a reason to celebrate.