POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Real-time evaluation of the human body's inertia tensor by using a stereo-depth camera and a deep-learning-based algorithm

Tesi di Laurea Magistrale in
Mechanical Engineering - Ingegneria Meccanica

Author: **Laura Viprati**

Student ID: 994060
Advisor: Prof. Hermes Giberti
Co-advisors: Marco Carnevale, Nicola Giulietti
Academic Year: 2023-24

Abstract

Researches on the dynamic contribution of the human body during motion have garnered significant interest over the years due to their applicability in various fields. Having a real-time method for estimating the inertia tensor of the human body enables the calculation of mutually exchanged forces during human-robot interactions, it promotes the development of dynamic simulators that consider the inertial contribution from body movement and allows for the analysis of body dynamics to create personalized rehabilitation exercises. The aim of this thesis is to develop a discretized model of the human body using a depth camera and a deep-learning-based algorithm to real-time estimate the individual's pose and, consequently, his inertial properties. The goal is to create a robust and versatile model capable of adapting to individuals with different physical characteristics and providing freedom of movement. The questions addressed in this research are: *Is it possible to obtain, through a discretization of the human body, close-to-reality inertial values? To what extent is it acceptable to simplify the model construction to achieve optimal real-time conditions?* To answer these questions, after defining the mathematical steps necessary for estimating the position of the center of mass and the inertia tensor during movement, various artificial intelligence algorithms are employed to more or less completely identify body poses. By comparing numerical results obtained through different algorithms and considering literature reference values, it is concluded that the developed method effectively estimates the inertial contribution associated with each human body pose. Furthermore, a more approximate discretization has limited influence on the results, allowing for the adoption of algorithms that reduce computational times and, consequently, increase the update frequency of the estimation.

Keywords: inertia tensor, human body, real-time estimate, deep-learning-based algorithm

Abstract in lingua italiana

Ricerche in merito al contributo dinamico dato dal corpo umano durante il movimento hanno destato particolare interesse nel corso degli anni in quanto usufruibili in vari campi applicativi. Avere a disposizione un metodo per la stima del tensore di inerzia del corpo umano in real-time consente di calcolare le forze mutualmente scambiate durante un interazione tra persona e robot, di sviluppare simulatori dinamici che tengano in considerazione il contributo inerziale dato dal movimento del corpo e di effettuare un'analisi della dinamica del corpo con il fine di creare esercizi personalizzati a scopo riabilitativo. L'obbiettivo di questa tesi è quello di sviluppare un modello discretizzato del corpo umano mediante l'utilizzo di una camera di profondità e di un algoritmo di intelligenza artificiale che consenta di stimare in real-time la posa dell'individuo e, conseguentemente, le sue proprietà inerziali. Si mira a sviluppare un modello robusto e versatile, in grado di adattarsi a persone con differenti caratteristiche fisiche e consentire libertà di movimento. Le domande alle quali si cerca risposta sono: *è possibile, mediante una discretizzazione del corpo, ottenere valori inerziali prossimi alla realtà? Fino a che punto è accettabile semplificare la costruzione del modello con lo scopo di approssimare al meglio la condizione di real-time?* Per rispondervi, dopo aver definito i passaggi matematici necessari alla stima della posizione del centro di massa e del tensore di inerzia durante il movimento, sono stati adottati vari algoritmi di intelligenza artificiale che consentono di individuare in maniera più o meno completa la posa del corpo. Mettendo a confronto i risultati numerici ottenuti applicando i vari algoritmi e considerando anche valori di riferimento in letteratura, si è potuti arrivare alla conclusione che il metodo sviluppato permette effettivamente di stimare il contributo inerziale associato a ciascuna posa del corpo umano. Inoltre, una discretizzazione più approssimativa ha una limitata influenza sui risultati, consentendo l'adozione di algoritmi che permettono di ridurre i tempi di calcolo e, conseguentemente, aumentare la frequenza di aggiornamento della stima.

Parole chiave: tensore di inerzia, corpo umano, stima in tempo reale, algoritmo di intelligenza artificiale

Contents

Abstract **i**

Abstract in lingua italiana **iii**

Contents **v**

1 Introduction **1**
1.1 Rationale . . . 1
1.2 Fields of application . . . 2
1.3 Human body inertia tensor estimation . . . 3
1.4 Human body pose estimation . . . 9
1.4.1 Sensors technology . . . 9
1.4.2 Representation of human body poses . . . 11

2 Materials and methods **15**
2.1 Image acquisition system . . . 16
2.2 3D human pose estimation . . . 17
2.3 Post-processing operations . . . 19
2.4 Definition of the geometric model . . . 20
2.4.1 Discretization of the human body using a 33-point model . . . 21
2.4.2 Discretization of the human body using a 17-point model . . . 24
2.5 Evaluation of the position of the center of mass . . . 25
2.6 Evaluation of the inertia tensor . . . 27
2.6.1 Body segment inertia tensor . . . 28
2.6.2 Rotation and translation of the inertia tensor . . . 29
2.6.3 Inertial ellipsoid . . . 31
2.7 Model validation . . . 32
2.8 Sensitivity analysis . . . 33

3 Results **35**
3.1 33 point model 35
3.1.1 'Open arms' and 'closed arms' configurations 35
3.1.2 Real-time updating of inertial properties 50
3.2 17-point models 66
3.2.1 Reduction of MediaPipe's model from 33 to 17 points 66
3.2.2 Comparison with TensorFlow and YOLOv8 78

4 Conclusions and future developments **85**

Bibliography **89**

A Appendix A **95**

List of Figures **103**

List of Tables **107**

List of Symbols **109**

1 | Introduction

1.1. Rationale

The estimation of the inertia tensor of the human body has been a central topic within the scientific community for numerous years. From the 1950s to the present days, it has been extensively explored and developed, leading to increasingly close-to-reality evaluations. The aim of the present thesis is to provide a reliable tool for the estimation of the most relevant inertial quantities, which include the position of the center of mass and the inertia tensor of the human body, through the usage of computer vision and artificial intelligence-based techniques. Previous studies are characterized by limitations associated with their not possible applicability in real-time or with their not so large versatility, caused by an instrumentation that can be utilized only in restricted and controlled environments or only with specific individuals performing determined movements. Consequently, even if large datasets providing values associated with different subjects under test performing a relatively wide range of motions are available, with the present research it is being addressed the need for a method capable of supplying updates in real-time, placing particular emphasis on its robustness and versatility in terms of adaptability to various human sizes and proportions, positioning and physical activity that is being performed.

The structure will be composed as follows: the present introductory Chapter provides an overview of the current state of the art and the most significant application fields for the topics covered in this thesis. It is in turn divided into three key subsections concerning 1) the main fields of application, 2) the models adopted for the evaluation of body dimensions and inertial quantities and 3) the research status on the estimation of the human body pose in real-time. The second Chapter is dedicated to the description of the adopted discretized model of the human body and of the mathematical passages followed for the calculus of its inertial quantities. The numerical results are shown in Chapter three and, finally, conclusions and descriptions of possible future developments are presented in Chapter four.

1.2. Fields of application

The estimation of the human body inertia tensor in real-time has significant applications in various fields, since it is fundamental for a deeper understanding of human movement dynamics and for the calculation of forces and moments exchanged during interactions with various objects. Some of the most important employments include the field of biomechanics, robotics and dynamic simulators. For what concerns biomechanics, inertial quantities are relevant for rehabilitation purposes since, through their estimation and the consequent dynamic analysis of the human body, it is possible to design adequate exercises aimed at a specific recovery. In this sector, as well as in orthopedic or sports sciences, a relevant often adopted study is the so called 'gait analysis', which involves the observation and interpretation of various parameters associated with body motion [1, 2].
Regarding robotics, the most important area of applicability is human-robot interaction occurring in collaborative robotics, which is receiving a growing attention from the scientific community due to its high flexibility and large applicability. In this context, the knowledge of human body motion, and consequently of the forces mutually exchanged with the robot, plays an important role for both control and safety reasons [3, 4]. By integrating the real-time estimation of the inertia matrix of the human body into the control algorithm, the robotic system can adapt more promptly to changes into its working region, ensuring a higher efficiency.

The main application field of our interest, anyway, is the world of dynamic simulators, which includes driving, aviation, fighting, virtual reality (VR), biomedical and many other sectors. The basic idea is to create an illusion of motion by recreating movements and interactions of various objects or systems in a simulated environment. More in details, "human-in-the-loop" identifies a specific kind of simulators, characterized by the not negligible influence of human motions and interactions [5, 6]. Human posture and inertial properties associated with it acquire a critical role in the dynamics of the action that is being simulated. These kind of simulators allow the user to interact with the virtual environment, causing a response of the system associated with his motion and making it possible to practice various tasks in a safe and controlled setting.
The main application examples we want to refer to are light-weight crafts simulators [7], for which the skipper's posture has a large influence on the inertial properties of the entire system, skiing simulators [8] and, primarily, driving simulators [9, 10]. It is important to notice that, in all cases, the relationship between the vehicle simulation and the output given by the controller of the platform on which the user is placed does not follow a direct 1:1 correspondence. This occurs because of structural and dynamic performance

limitations associated with such platforms, which does not allow the exact reproduction of the simulated vehicle movements. The output of the numerical model of the system, in fact, is manipulated through the usage of motion cueing techniques in order to generate a more adequate correspondence between the movement we are trying to simulate and the actual body motion [11]. This relationship depends on the perception of motion in terms of linear acceleration and angular speed of the head of the user, since it is driven by the vestibular system, which is a sensory system situated in the ears responsible for creating sense of balance. The closer these values of acceleration and speed are to the ones we are trying to simulate, the best the motion perception will be.
If the entire loop composed by the simulated environment, the human user, the numerical model of the system dynamics and the actuation of the platform is iterated in real-time, the operator will be able to actually feel the platform below his feet move accordingly to the motion that is trying to be simulated.
Dealing more in details with driving simulators, the driver modelling has been considered by various communities with different purposes. Some of them mainly focus on human factors, dealing with problems of driver's distraction and security for different tasks during driving [12, 13], others instead make use of sensors, which could be entire body sensors [14] or only facial trackers [15], for creating a model to predict the driver's intent [10]. In the majority of the cases, acceleration, deceleration, skid and steering are simulated taking into account also the variation of inertial properties associated with the change of posture of the driver. In a similar manner, also for craft or skiing simulators, each body motion is associated with a variation of the inertia tensor of the entire system, resulting in a different inclination of the platform on which the user is placed.
The objective of the present thesis is to provide a method for estimating human body inertial quantities in real-time regardless of the specific application. The strength of the approach that is going to be defined in the following Chapter relies on its great robustness and versatility with respect to the physical activity that is being performed and to the subject under test. Dealing with dynamic simulators as well as human-robot interaction mechanisms or studies for rehabilitation purposes, the method for the identification of the human body inertia tensor is the same, leading to results whose precision and accuracy depends mainly on the way the human body is discretized and on the instrumentation adopted, but not on the utilization context.

1.3. Human body inertia tensor estimation

Over the years, advancements in technology have led to the development of increasingly powerful methods for the identification of the inertia matrix of the human body. Starting

from older studies and moving toward more recent times, it is possible to recognize two main approaches. The first one is based on the definition of scaling factors and regression equations, which allow to relate proportions and body parts' dimensions of the individual under analysis with some generalized inertia values. The second method, instead, consists in the discretization of the human body through the definition of a geometrical model and the consequent evaluation of the inertial properties associated to it.

Let's start analysing more in details the first mentioned approach, which, for simplicity, will be referred to as the 'regression-based method'. Equations and scaling factors necessary for its application are derived from a large set of data, which include easy-to-collect anthropometric measurements and the corresponding inertial quantities that are being investigated. For the creation of such datasets, various approaches have been adopted throughout the years. In the 1950th and 1970th anatomical studies mainly relied on direct measurements of male and female sectioned cadavers [16–18]. Starting from these accurate studies, it was possible to derive the dimension of each body part and consequently, integrating mass and density contributions within each part, to derive all the other inertial quantities. This approach is highly time-consuming, making it challenging to acquire an amount of data representative of the entire human kind. Anthropometric quantities, in fact, are accurate and specific only for the actual population that is being investigated, but they are not extendable to individuals characterized by different body proportions. More recent approaches for the identification of human body dimensions rely instead on medical diagnostic technologies such as magnetic resonance imaging (MRI) [19], computer tomography (CT) [20, 21], gamma radiation scanning [22, 23] and X-ray absorptiometry [24, 25]. Also these techniques require a significant amount of time and, besides, some of them can be risky due to the involved exposure to radiations, which can potentially cause damaging health effects.
Among all the various regression-based models that can be found in literature, the intend is to spotlight the main ones. In 1955, starting from a dataset built considering eight sectioned cadavers, Dempster developed a model for estimating segmental masses and positions of the center of mass starting from specific anthropometric measurements [16]. His research was funded by the Wright Air Development Center with the purpose of studying the range of motion for the seated individuals of the Air Force flying personnel. In the following years, also Zatsiorsky proposed a similar model: starting from the numerical results obtained applying gamma radiation scanning technologies, he developed a regression-based approach aimed at the identification of the inertia matrices associated with various body parts [26]. His objective was that of quantifying the most important

biomechanical parameters involved during human movement. Later on, his research was carried forward by another scientist, De Leva [22], which modified the parameters identified by Zatsiorsky redefining them with respect to slightly different key body points.
Each model has its own set of assumptions and predictions, thus the selection of one of them is generally based on the specific requirements of the study and on the set of available anthropometric data. The main drawback associated with the regression-based method lies in the fact that inertial quantities are obtained generalizing data associated with reference body types. Thus, in case in which the individual proportions highly differ from the initial dataset, the resulting values may be not so accurate. Despite that, predictive equations are still used for some applications thanks to their easy implementation and low computational effort. Besides, considering the correct proportions on which the model is based, the corresponding results are highly precise and accurate and can thus be utilized to validate other models.
All the approaches that are being investigated focus on estimating properties related to each body part or to the entire human body, but only considering very basic poses. For a more detailed study, it would be necessary to take into consideration also a real-time update of the position of each body part, thus requiring the implementation of cameras, sensors, IMUs or force plates. The majority of the methods discussed in this Section does not involve the usage of such devices, but they concentrate on fixed poses defined a priori. For now we are solely focusing on modeling the human body with the purpose of calculating inertial quantities associated with each body part, while in the next Section it will be done a more detailed analysis of the most commonly employed techniques for the update of the body parts' positions in real-time.

Nowadays the most commonly adopted method is no longer the regression-based one, but the estimate of the inertia tensor is principally obtained using geometric models, which rely on the discretization of the human body into three-dimensional geometric shapes. The general idea is that of dividing the body into a certain number of rigid parts, which may be the head, the torso, the hands, etc., and of associating to each of them a certain shape, whose dimension is evaluated starting from anthropometric measurements or through computer vision techniques. The difference with respect to the previously introduced approach is that, while before inertial properties were derived using statistical approaches based on regression equations, in this case they are integrated directly considering the human body as a tree structure composed of a certain number of rigid parts. The first geometric model was developed by Hanavan in 1964 [27] with the purpose of quantifying the differences among the inertial quantities evaluated for eight different body

poses. The human body is discretized into fifteen segments which are head, upper torso, lower torso, hand, upper and lower arms, upper and lower legs and feet. Their dimensions are specified basing on twenty-five anthropometric measurements and their masses are set using regression equations, which are functions of the total body mass and segmental length. For this study, as well as for the majority of the geometric models, the choice of the form to be associated to each body part falls on simple geometric shapes, which may be cylinders, spheres, ellipsoids or frustums of cone. Making a comparison with the previously introduced regression-based models, it is immediately possible to notice the advantage in terms of individuality and specificity of the results. Hanavan is no longer generalizing values obtained for a specific body, but he is adapting the dimension of each discretized body part to the actual person that is being tested. The results are compared with those obtained experimentally by Dempster [16], showing a good correspondence both in terms of position of the center of mass and of the inertia tensor associated with each body part. The segments which show the greatest deviation from the experimental data are hands and the feet, but also in this case they do not exceed a 10% variation from the original values.

A slightly different approach has been studied by Yeadon in 1990 [28]. The human body is modelled using forty geometric solids, which are defined by 95 dimensional measurements. Initially it is done a rougher distinction of only eleven segments, which are later further sectioned into forty parts by planes perpendicular to the longitudinal axes of the solids. Each part, with the only exception of the head, is shaped as a stadium solid, which is a solid whose section is a rectangle with an adjoining semi-circle at each end of its width. This kind of discretization allows to reproduce more faithfully the shape and the contour of the human body, but at the same time it requires many additional anthropometric measurements. For this reason it is mainly applicable in those fields where it is possible to consume time for the collection of such a large amount of data or where only a few individuals are considered.

The most detailed and accurate model was developed by Hatze [29] and subsequently adapted and simplified by Robertson [30] to enhance its practical applicability in biomechanical modelling and simulation fields. Hatze's approach consists in the breaking of every body segment into a higher number of slices, each with a specific geometry, size and density. Every shape requires multiple measurements for the definition of its length, diameter or perimeter, resulting in a very complicated and time-consuming procedure. Starting from that, Robertson proposed a simplified model reducing dramatically the number of measurements required and introducing the possibility of interpolating the segments' dimensions in order to get smoother surfaces and joints. The validation is carried out performing a comparative analysis with other methods, in particular with the

regression-based one introduced by Zatsiorsky. Due to the closer-to-reality shape of the body segments, the results show high coherence. The approaches introduced by Yeadon and Robertson, despite their recognizable accuracy, are not suitable for applications where a quick identification of the discretized model is necessary. The evaluation of the dimension of each body part, in fact, would require the identification of the position of many body points, which is difficult to achieve in real-time.

Across the years, several approaches similar to that adopted by Havanan in 1964 have been developed, differentiating the number of body parts in which the body is discretized and the required anthropometric measurements, leading to a large variability of results. In the last decade, for instance, Nikolova introduced a geometric model with the aim of providing a dataset of inertia values associated with different poses, which was used for the construction and control of walking humanoid robots [31, 32]. In this case the human body is schematized through sixteen segments, which differ from Hanavan's ones for the division of the torso into three parts rather than two. This choice allows to better distinguish male and female figures, leading to more specific and accurate results, but at the same time it requires a higher number of anthropometric measurements which may not always be available. For what concerns the validation of the results, it was performed considering the values obtained from the regression-based models developed by Zatsiorsky [26], Clauser [17], Chandler [18] and Dempster [16]. The agreement between the simplified-model values and the experimental results is quite good, even if it can be noticed, as for the Hanavan's model, a certain deviation. This may be caused by the fact that the individuals analyzed by Nikolova do not have the exact same physical characteristics as those considered by the experimental studies and, primarily, they are associated with the simplifying assumption on which the model is based. The majority of the most commonly adopted geometric models share the following simplifying hypothesis: each body segment is assumed to have isotropic density, segmental volumes are constant despite body motion and the entire body is symmetrical with respect to the sagittal plane, which means that no distinction is done between right and left limbs' dimensions. These hypotheses differ from reality. The human body, in fact, is not consistent in density due to the variable water content, tissues and bones composition, it is never perfectly symmetric because of genetic and functional factors such as the predominance of one hand over the other and, finally, segmental volumes are variable mainly as a consequence of muscles contraction and lengthening. The simplifications adopted are expected to cause some systematic errors, which may be very difficult to identify. Nevertheless, they are accepted since they offer relevant advantages: this kind of model, in fact, leads to close-to-reality results and is also suitable for real-time applications, since the determination of each body part's dimension requires low computational effort. Adopting more complex

approaches, characterized by a higher number of discretized body parts or by variable segmental densities, the estimate of the inertial properties would require much more time and thus it would not be updated in real-time.
The most recent geometric model based on a similar approach was introduced in 2021 by Jagadale and Agrawal [33] for the development of protective clothing and working tools for female agricultural workers of Central India. Segments' dimensions are defined basing on 30 averaged anthropometric measurements of a set of individuals, which are representative of the Indian population. The human body is discretized into 14 segments, which are given the shape of ellipsoid (head+neck), elliptical cylinder (torso), sphere (hands) or frustum of cone (limbs and feet). The validity is tested making as usual a comparison with specific regression-based models, more in details an analysis is carried out in SAS software using one-way analysis of variance (ANOVA). Considerable differences were observed in the values of volume, density and position of the center of mass associated with the lower part of the arms and the feet, while the remaining results show a better correspondence. This model shares many similarities with the one developed by Nikolova, making it suitable for real-time applications too.

In conclusion, both regression-based and geometric models, if adequately applied, can lead to satisfying results. The choice between them primarily depends on their intended application and consequently on the required level of accuracy and computational effort. Regression-based methods rely on predefined sets of data associated with specific populations and for this reason they may be more suitable for "person-kind" or population analyses, like the one conducted by Jagadale about the Indian population [33]. They may also be implemented to find statistical trends within the considered data. Their main disadvantages are associated with the sensitivity to outliers and with the fact that the resulting values are highly influenced by the selected regression equations. Different models and thus different equations, in fact, may lead to variable results. For what concerns geometric models, the human body is discretized as a kinematic chain of rigid bodies, whose volumes, masses and densities are defined specifically for the individual that is being tested. Accordingly, they provide individual and specific results. On the other side, their accuracy depends on the complexity of the three-dimensional shape associated with each body part and on the simplifying assumptions adopted. The higher is the complexity of the model, the more accurate the resulting inertial quantities will be, but at the same time more detailed anatomical measurements will be required. For the purpose of the present thesis, it was decided to adopt a geometrical model instead of a regression-based one in order to be as versatile as possible. With this method, in fact, reliable results as-

sociated with individuals characterized by different proportions can be obtained, leading to a more inclusive and precise study. Among the previously cited ones, the geometric model adopted is similar to those introduced by Nikolova and Jagadale. Due to their simplicity and low computational effort they are more suitable for real-time applications, in contrast with Yeadon and Hatze's models, which require a number of anthropometric measurements which is too large to be quickly identified.

1.4. Human body pose estimation

1.4.1. Sensors technology

One of the main aspects to be discussed for a correct estimation of the inertial properties consists in the evaluation of the human pose. Once the discretized model of the human body has been identified, it is necessary to locate all body parts in space in order to combine correctly their inertial contributions and, in this way, calculate the inertia tensor associated with the entire body. The more accurately this evaluation is done in real-time, the more specific and reliable the results will be. In recent years researches in this direction have been widely carried out due to its large range of applicability: some interesting real-world applications could be human-computer interaction to allow robotic arms interacting with the human user, autonomous driving to avoid collision with pedestrians or biomechanics for health status monitoring [34].
Among the several solutions available on the market for tracking whole-body motion, a primary distinction can be done between marker-based and markerless approaches. Marker-based motion-capture systems are characterized by the usage of markers attached to the subject's body, whose position in space is recognized by a plurality of cameras. By continuously tracking these positions, it is possible to reconstruct the individual's motion and thus his pose. The most commonly used marker-based technologies, like those produced by Vicon or MotionAnalisys [35], are mainly adopted for indoor applications [36]. For outdoor motion capturing the usage of Inertial Measurement Units (IMUs) is quite common [37]. They are wearable devices equipped with accelerometers and gyroscopes which, whenever are placed on different body parts, are able to provide real-time motion tracking. These approaches suffer from two main drawbacks: measuring acceleration values, in order to obtain the position of each body part it is necessary to integrate data in time, besides the system is susceptible to disturbances associated with the usage of IMUs and to electromagnetic interference, which make it unusable in proximity to electric motors. Another possible solution consists in the usage of force-places [38], even if in reality it is not so commonly adopted since it requires a deep knowledge on how center of

pressure and vertical reaction forces are influenced by the body posture.
In recent years, markerless solutions have become the predominant approaches for real-time human pose estimation. They are based on the usage of deep-learning algorithms, which are trained on extensive datasets to accurately estimate the position of key elements, such as body points or facial features, through the processing of 2D or 3D images. They are applied in multiple fields allowing the study of body motion for rehabilitation or sport motion analysis purposes [2, 39], besides they are commonly implemented to study human-robot interaction mechanisms [36, 40]. Due to the absence of physical markers, they have the advantage of permitting more unrestricted and natural movements, allowing also for the saving of time that would otherwise be spent for the correct positioning of the markers. On the other side, they may have lower precision in capturing small body motions and they generally deal with more complex algorithms, which increase computational time and effort. For both approaches, in-the-wild applications may lead to some difficulties concerning lighting, shadows or view obstruction. The choice among systems with or without markers depends on the specific application and has to take into account various aspects such as required precision, computation time limits or setup complexity.
Due to the multiple advantages associated with the usage of markerless approaches, which do not involve errors associated with integration of data or electromagnetic interference, for the purpose of the present work the choice has fallen on a deep-learning-based approach. Our aim is that of estimating how the inertia matrix varies during a fluid motion of the body, thus it is fundamental to guarantee a large freedom of movement for the individual under analysis. Besides the model adopted for representing the human subject is a discretization of the original body and, consequently, the target degree of accuracy is not so high to require necessarily the implementation of a marker-based approach.

Datasets

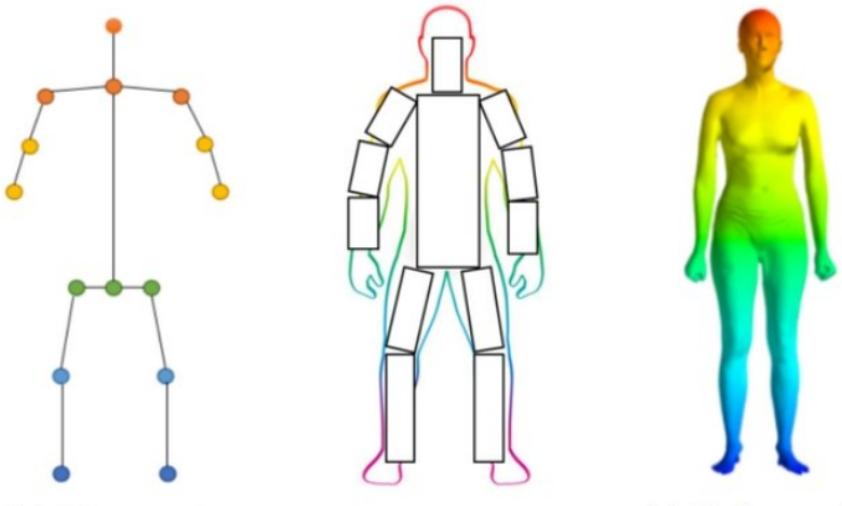
Datasets play a crucial role in the development, training and validation of deep-learning-based algorithms. Their input consists of videos, images or sequences usually gathered by a motion capture system, from which it is given the correct output in terms of body joints' positions, object labels or length of specific body parts. Starting from the association of every image to the correct values specified in the dataset, deep learning models are trained and their parameters are adjusted to meet the already identified outputs. In this way they will be able, given an input image, to automatically identify the body parts' positions and consequently the pose of the individual with a certain degree of accuracy.
The most commonly employed datasets include Human3.6M, HumanEva, MPII Human Pose and MPI-INF-3DHD. Human3.6M [41] is one of the largest scale datasets for 3D

human sensing in natural environments and, as the name suggests, it is based on 3.6 million images and corresponding poses acquired by recording the performance of 11 actors. This dataset, as well as the others, is based on images acquired during various human activities, such as talking, walking, smoking or taking photos, in order to consider a range of motions as large as possible. HumanEva dataset [42] is based on the tracking of motion of 4 individuals performing 6 predefined repeated actions. It employs a multi-camera setup and reflective markers, allowing for the capturing of synchronized videos and consequently for the reconstruction of human motion. Finally, MPII Human Pose [43] and MPI-INF-3DHD [44] provide informations on joint locations and movements considering a larger variability of motion with respect to the previously cited algorithms. The main difference with respect to HumanEva lies in the fact that data are collected using markerless motion capture systems, allowing the actors to wear everyday clothes instead of special suits required by the presence of markers. For MPI-INF-3DHD dataset it has also been conducted an outdoor scenario analysis, with the purpose of evaluating lighting and background influences on the human pose estimate.

1.4.2. Representation of human body poses

Human body poses can be represented in various ways, basing on the adopted technology, the purpose of the individual's analysis and the required accuracy. It is possible to distinguish three main kinds of models, which are called 'skeleton-based', 'contour-based' and 'volumetric' models. According to the first one, also called 'kinematic' model, the human body is represented as a tree structure composed of a set of key joints like shoulders, hips, wrists or ankles and by the connections among them (Figure 1.1a). It is frequently applied to capture the relations between different body parts and it can be used for both 2D and 3D pose estimations. Its capability to express informations concerning the texture or the external shape of the subject, anyway, is limited. Some of the most commonly used skeleton models include OpenPose [45], which estimates both 2D and 3D poses of single or multiple individuals, Microsoft Kinect Skeletal Tracking [46, 47], MediaPipe [48] and PoseNet [49], which is a real-time pose estimator developed by Google. These technologies are not able to recognize who is in an image, but they more simply estimate where key body points are placed in space. The main difference among these models lies is the number and exact positioning of the body points that are tracked. They are all based on a bottom-up approach, which means that the algorithm detects at first all body joints independently across the entire image and then assigns them to the corresponding person. Top-down approaches, instead, detect each person first and then locate their joints and

estimate their pose individually.
The second kind of model introduced is the contour-based one, also called 'planar' model (Figure 1.1b), which can be adopted only for 2D evaluations. It is generally utilized to visualize the shape of the human body, since it represents body parts as multiple rectangles for approximating the contour. It is the simplest to implement among the cited ones, but it is not equally precise and it is not applicable for 3D analyses. Among the multiple planar models existing, some of the most widely adopted are the Active Shape Model (ASM) [50], which captures the entire human body silhouette and its variability, the Deformable Part Model (DPM) [51], which is based on the decomposition of the body into deformable parts and the Constrained Local Model (CLM) [52], which is mainly used for facial landmark detection. Their main fields of applicability include anatomical structure analysis, human-computer interaction and, in particular for CLM, facial recognition and emotion analysis.
Finally, volumetric models operate by representing the human body as a three-dimensional volume (Figure 1.1c). This depiction allows to combine informations regarding the position of various body parts, like torso or limbs, taking into account also their 3D shapes. Volumetric models are simpler to implement and less sensitive to noises with respect to skeleton-based ones, but they suffer a larger inaccuracy in estimating joints' 3D positions, potentially leading to less realistic pose estimations. In recent years various volumetric models have been developed. Among them, some of the most widely spread are the Volumetric Occupancy Networks (VON) and the Volumetric Heatmap Networks (VHN).



(a) Kinematic model (b) Planar model (c) Volumetric model

Figure 1.1: Models for human body pose estimation [34]

For the present study it has been chosen to apply various kinematic models, thanks to their ability to provide accurate estimates of 3D body poses in real-time. More in details, different skeleton models will be identified using the body points made available by MediaPipe's, TensorFlow's and YOLOv8's algorithms. The identification of these points is fundamental for the definition of the geometric model: once their positions has been identified, considering appropriate proportions it is possible to derive the dimension of each body part and, consequently, the entire geometric model of the discretized human body.

2 | Materials and methods

The present Chapter provides the detailed description of the passages followed for the evaluation of the human body inertial properties in real-time, focusing at first on the instrumentation employed for human pose estimation and then on the mathematical procedure used to evaluate and update the position of the center of mass and the inertia tensor. To ensure a quicker understanding of the procedural flow, in Figure 2.2 is shown the flow chart of the sequence of passages followed at each time instant for the estimate of the human body inertial properties.

Initially the image is acquired through a depth camera, which provides the three-dimensional position associated with each pixel, then it is processed by a deep-learning-based algorithm for the evaluation of the skeleton model. The algorithm is able to recognize well-defined key body points and, thus, knowing their position in space, it is possible to reconstruct the entire three-dimensional kinematic model. It is fundamental to perform post-processing operations aimed at checking the correctness of the temporal sequence of each body point's position identified by the algorithm. There may be, for instance, out-of-trend data associated with a wrong positioning of the camera that must necessarily be readjusted. At this point each body part is associated with a specific geometrical shape for the identification of the geometric model. It is defined specifying the dimension of each body segment, which is set directly starting from the points identified by the algorithm, its mass and its density. Finally the inertial quantities associated with each segment are evaluated and, through the correct mathematical procedure, their contributions are combined in order to obtain the position of the center of mass and the inertia tensor of the entire body.

The procedure presented so far can be iterated every time a new image is made available by the camera, allowing for the real-time update of the inertial properties associated with the human subject under analysis.

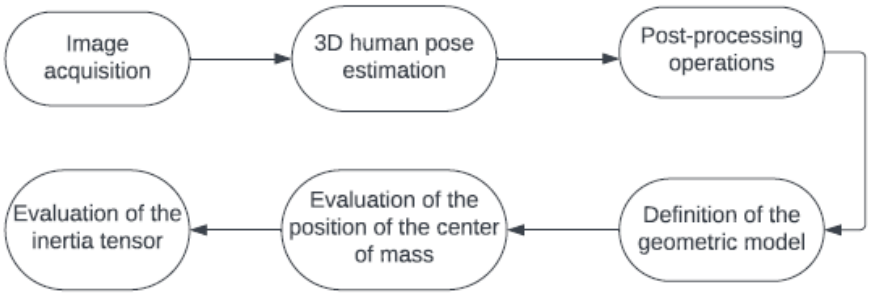


Figure 2.1: Flow chart showing the sequence of passages followed for the identification of human body's inertial properties instant by instant.

Now that the general structure of the Chapter has been depicted, let's delve into each element of the flow chart in more detail.

2.1. Image acquisition system

The image acquisition system utilized consists of the stereo-depth camera Intel RealSense D415. It is equipped with an infrared projector and two sensors to create a depth map allowing for the evaluation of the depth position associated with each pixel. Besides it utilizes a high-resolution RGB camera for the capturing of colourful images, even if for the purpose of this work it will be considered only a grey scale.

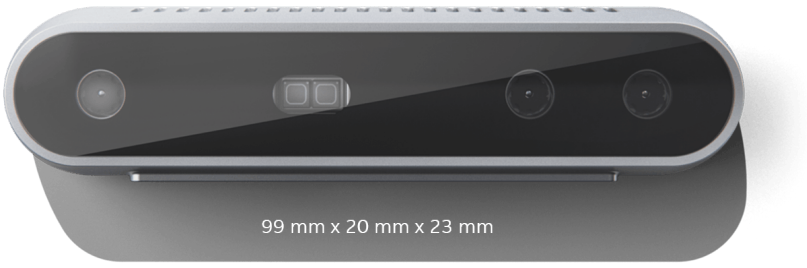


Figure 2.2: Front picture of the stereo depth camera Intel RealSense D415.

The maximum resolution supported is width x height = 1280 x 720 pixels at 30 frames per second (fps). Higher frequencies, such as 60 or 90 fps, can be supported too. In this

way the real-time condition is better approached, but at the same time higher frequencies are associated with a reduction of the maximum acceptable resolution, as shown in Table 2.1.

Resolution [pixel]	fps [-]
1280x720	30
848x480	60
640x480	90

Table 2.1: Frames per second recorded by the monochrome sensor of the RealSense D415 stereo-depth system at varying acquisition resolution.

The camera is placed in front of the participant at about one meter above the ground. It is fundamental to ensure that in the captured image there are no other individuals aside from the one undergoing the experiment, otherwise the deep-learning-based algorithm would have problems identifying the skeleton model. Besides the camera must be positioned sufficiently distant from the participant in order to let him move freely without his limbs going beyond the camera's field of view.

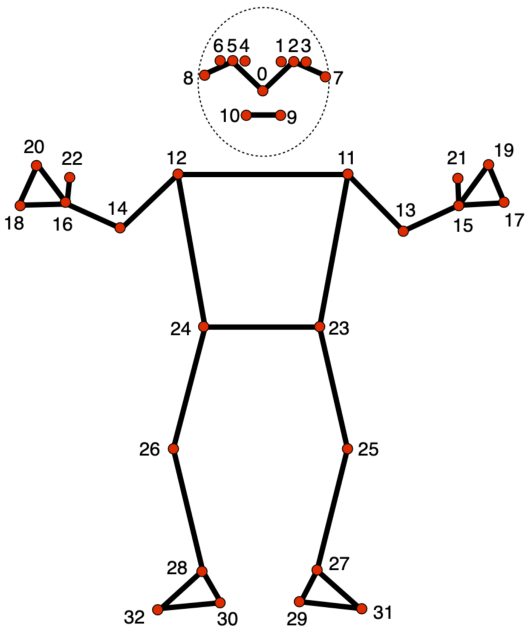
2.2. 3D human pose estimation

The skeleton three-dimensional model is generated through the adoption of deep-learning-based algorithms. At every frame a new image is made available by the camera and it is processed by the algorithm for the identification of key body points. Several alternatives are available, differing in the number of detectable points and their precise location on the body. The most commonly used in literature are:

- MediaPipe Pose, which identifies the position of 33 body landmarks,
- YOLOv8, which estimates the 3D human pose through the identification of 17 body points,
- TensorFlow, which detects the position of the same body landmarks as those estimated by YOLOv8.

MediaPipe Pose is a specific component within the MediaPipe framework focused on human pose estimate. It provides a pre-trained model for accurately estimating the position of 33 body landmarks. These points correspond to various body joints or facial features, which are listed in Figure 2.3.

YOLOv8 represents a set of pose models pre-trained on the COCO dataset, each of them

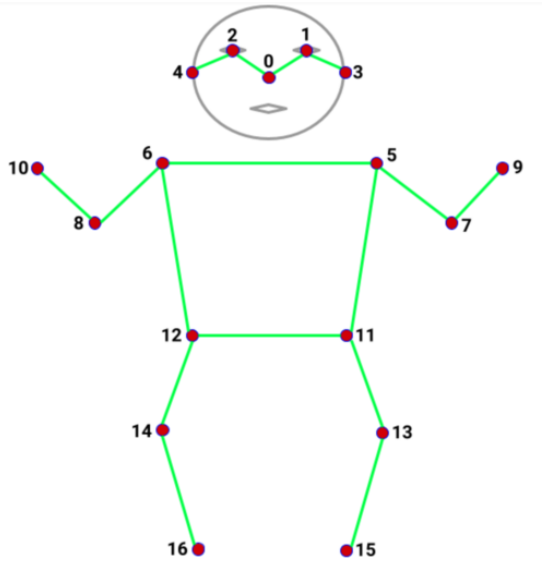


0 nose
1 left eye inner
2 left eye
3 left eye outer
4 right eye inner
5 right eye
6 right eye outer
7 left ear
8 right ear
9 mouth left
10 mouth right
11 left shoulder
12 right shoulder
13 left elbow
14 right elbow
15 left wrist
16 right wrist
17 left pinky
18 right pinky
19 left index
20 right index
21 left thumb
22 right thumb
23 left hip
24 right hip
25 left knee
26 right knee
27 left ankle
28 right ankle
29 left heel
30 right heel
31 left foot index
32 right foot index

Figure 2.3: 33 points identified by MediaPipe Pose [48].

characterized by a certain size in terms of resolution of the processed input image and by a CPU running speed. In ascending order of speed and decreasing size, calculations will be performed considering YOLOv8l-pose and YOLOv8m-pose [53]. For what concerns TensorFlow, it provides in turn two models called MoveNet and PoseNet, characterized by different sizes and running times too [54]. The 17 body landmarks identified by all these algorithms are shown in Figure 2.4. Even if the image processed and the detected body landmarks are the same, the resulting skeleton model obtainable with each algorithm is different. It depends on the resolution of the input image and it is also associated with the fact that different algorithms are not trained on the same dataset.

The main difference between the 17-point models and the one developed by MediaPipe lies in the identification of the position of head, hands and feet. While previously the head pose was defined through the identification of 10 body landmarks, using TensorFlow or YOLOv8 models this is achieved considering only 5 points. Hands and feet, whose fingers' positions were previously evaluated, are now no longer detectable. From one perspective, this entails a loss of informations making it challenging to precisely estimate the pose of the entire body. However, on the other side, it is also true that estimating the position of points such as fingers and toes is generally associated with larger errors: there is a higher probability for these points to go outside the field of view of the camera, resulting in approximate or even completely wrong models. Besides, some of the points identified by MediaPipe are disregarded in the creation of the geometric model. For instance, it



0 nose
1 left eye
2 right eye
3 left ear
4 right ear
5 left shoulder
6 right shoulder
7 left elbow
8 right elbow
9 left wrist
10 right wrist
11 left hip
12 right hip
13 left knee
14 right knee
15 left ankle
16 right ankle

Figure 2.4: 17 body landmarks identified by TensorFlow's and YOLOv8's algorithms [55].

is evident that assigning 6 points for the detection of eye positions may be excessive. Consequently, despite the loss of accuracy and the obtaining of more approximate results, opting for the 17-point models could be advantageous in mitigating errors associated with extremities' positioning.

2.3. Post-processing operations

Considering the temporal sequence of each body point's position as assessed by the algorithm, it was observed that certain out-of-trend data may occur, particularly in relation to body extremities such as hands and feet. Such deviation could be attributed to subjects making excessively broad movements, consequently going beyond the camera's visibility range. The best way to counteract this undesired effect is to accurately place the camera with respect to the individual ensuring the best trade-off between width of the field of vision and resolution and, since this may not be enough, implement a filtering mechanism. The starting point for the application of the desired filter is the correct identification of out-of-trend data. In order to define a reasonable threshold for their evaluation, a preliminary test is conducted taking into account the points' positions identified by the algorithm in a time interval of about 5 seconds. Given this large set of data, considering individually the temporal sequence of the position of each point, it is evaluated the average variation of position among consecutive time instants (μ) and its standard deviation (σ). After conducting several tests, is was decided to adopt a maximum acceptable variation from

the average value equal to 2.5 times the standard deviation:

$$\Delta max = \mu \pm (2.5\sigma). \tag{2.1}$$

Applying this reasoning to all body points, it is possible to identify the presence of out-of-trend data at each time instant. Finally, the approach chosen for their correction consists in the identification of the polynomial function that best fits the trend of correct points and the consecutive substitution of the scatter data with the corresponding position in time of this polynomial fitting. In this way it is possible to obtain a set of points that best represents the motion of the human body in time keeping the number of data associated with a certain time interval unchanged. If we had simply deleted the scatter values, in fact, we would have obtained a different number of points for each time interval depending on the correctness of the data identified by the algorithm.

2.4. Definition of the geometric model

Body parts are discretized into three-dimensional shapes for the identification of the human body geometric model. The number and size of these parts depends on the points that constitute the kinematic model, thus they are defined differently according to the deep-learning-based algorithm adopted. Before differentiating the various models applicable, let's list the simplifying assumptions on which they are based:

- each body segment has isotropic density;
- mass division among body segments and segmental volumes are constant, despite the variation of body position;
- body is assumed to be symmetric with respect to the sagittal plane, which means that no distinction is made between the dimension of right and left limbs.

Even if these hypothesis do not precisely correspond to reality, they are necessary to perform calculations quickly enough to ensure a real-time update of the resulting inertial quantities. Taking into account the variable density along each segment, or differentiating left and right limbs, the quantity of data to be processed would increase significantly. Indeed, for the same reason, these assumptions are applied in many other reliable studies, such as those conducted by Zatsiorsky [26], Chandler [18] and Dempster [16].

2.4.1. Discretization of the human body using a 33-point model

Taking into account the 33-point kinematic model, it was decided to divide the human body into 14 three-dimensional geometrical shapes for the definition of the geometric model: head and neck are discretized as an ellipsoid, the torso as a cylinder with elliptical cross section, upper arms, lower arms, upper legs and lower legs as frustums of cone and, finally, hands as sphere. The choice of the number of segments, as well as the assignment of the shape to each of them, depends on the number of points made available by the algorithm and it is also made with reference to some important models, in particular the one developed by Jagadale and Agrawal [33]. In the present study, in fact, the shape associated to each body part is the same as the one adopted by Jagadale and this decision is mainly driven by the simplicity in the extrapolation of each segment's dimensions. The main difference with respect to other studies ([32, 56]) lies in the choice of not sectioning the torso and it is driven by the fact that the 33-point algorithm provides only the positioning of shoulder and hips, making it difficult to estimate the correct proportions for the division of this part into two or three segments.
To have a clearer idea of how this geometrical model is structured, it has been reproduced in the CAD software "Inventor" (see Figure 2.5).

Figure 2.5: Reproduction of the human body geometrical model in CAD.

The dimension of each discretized body part is defined starting from the 33 points identified by the algorithm and using appropriate scaling rules, which are selected in accordance with anthropometric proportions defined in relevant studies [57, 58]. Body segments dimensions are also adjusted in order to meet the requirements in terms of segmental volumes, masses and densities.

Starting from the upper part of the body and moving downward, the first part encountered is the head, whose shape is discretized as an ellipsoid. Its center is placed in correspondence of the midpoint between the ears. The semi-minor axis is set as the ratio between the distance among the ears and 1.5, while the semi-major one is defined as the product between the distance among the eyes and the mouth and 2.2. The coefficients used to scale distances (1.2 and 2.2) are selected considering the detailed anthropometric analysis carried out in [59]. For what concerns the torso, which is represented as a cylinder with elliptical cross-section, its height is set directly as the distance between the axes identified by the shoulders and the hips. The minor axis is equal to 60% of the major one, whose length corresponds to half the distance between the shoulders. The center of the elliptical cylinder is placed in correspondence of the midpoint between shoulders and hips. All body arts, together with the feet, are given the shape of frustums of cone. The height of each frustum is directly derived from the points' positions identified by the algorithm: the length of the upper arm is equal to the distance between the shoulder and the elbow, the length of the lower arm to the distance between the elbow and the wrist, the length of the upper part of the leg to the distance between the hip and the knee, the length of the lower leg to the distance between the knee and the ankle and, finally, the length of the foot to the distance between the ankle and the foot index. As already mentioned, one of the hypothesis on which the geometrical model is based concerns the symmetry of the body with respect to the sagittal plane, which means that left and right limbs have coincident dimensions. To define the length of each of these body parts is has been decided to take into account the points' positions associated with the left part of the body and to extend these values also to the right limbs. The dimension of the bases of the frustums of cone is defined starting from the length of each segment using appropriate scaling factors. The ratio between the height of the segment and the major and minor base circumferences is set correspondingly equal to 1.1 and 1.2 for the upper arms, 1.2 and 1.7 for the lower arms, 0.85 and 1.1 for the upper legs, 1,1 and 1.55 for the lower legs, 1.55 and 0.8 for the feet. At this point, the last part of the body that has yet to be defined is the hand, which is discretized as a sphere. Its radius is set as the ratio between the distance from wrist to thumb and 1.5, while its center is placed in the midpoint between the wrist and the index.

When evaluating the dimension of the body segments, it is necessary to take into consideration some critical aspects associated with the temporal sequence of the body points' positions identified by the algorithm. Every time a new image is made available by the camera, the algorithm processes it and gives as output the position of 33 body points. Since it has not been designed with the purpose of maintaining fixed body segments' lengths or proportions, at every new acquisition the resulting distance among consecu-

tive points, and thus the corresponding segment dimension, varies. To counteract this problem, it has been decided to consider multiple acquisitions and set the length of each segment making the average. The higher is the number of frames considered, the better will be the estimate, provided that the individual does not perform broad movements that may damage the trend.

Once the dimension of each body segment has been properly evaluated, it is necessary to assign volume, mass and density values. The segmental volumes can be calculated considering well-known formulas associated with each geometrical shape. Starting from the ellipsoid, its volume is calculated considering the following expression:

$$V_{ellipsoid} = \frac{4}{3}\pi r^2 R, \tag{2.2}$$

in which R and r are correspondingly the lengths of the major and minor semi-axes. For a frustum of cone with height h, major semi-axis R and minor semi-axis r, the formula used is:

$$V_{frustum} = \frac{1}{3}\pi h(R^2 + Rr + r^2). \tag{2.3}$$

Finally, the volumes of a sphere with radius r and of an elliptical cylinder with height h and rays r and R are defined as:

$$V_{sphere} = \frac{4}{3}\pi R^3, \tag{2.4}$$

$$V_{cylinder} = \pi h R r. \tag{2.5}$$

The mass of each body segment is set as a percentage of the total body mass, using values defined by Dumas in [60]. This article, where previous studies concerning anthropometric and mass distribution characteristics of adult individuals are analyzed, provides reliable distinct mass percentages for male and female subjects. These values are listed in Table 2.2. Segmental densities, finally, are determined by making the ratio between corresponding mass and volume values.

	Female	Male
Head and neck	6.7 %	6.7 %
Torso	45 %	47.5 %
Upper arm	2.2 %	2.4 %
Lower arm	1.3 %	1.7 %
Hand	0.5 %	0.6 %
Upper leg	14.6 %	12.3 %
Lower leg	4.5 %	4.8 %
Foot	1 %	1.2 %

Table 2.2: Body segments' mass percentages adopted for the 33-point model.

2.4.2. Discretization of the human body using a 17-point model

Considering the 17-point models associated with the usage of the deep-learning-based algorithms TensorFlow and YOLOv8, the discretization of the human body for the definition of the geometrical model is slightly different. In this case, in fact, the position of the mouth as well as the one of fingers and toes is not identified by the algorithm, leading to the not possibility of evaluating the dimension of hands and feet body parts. Consequently the human body is divided into 10 rather than 14 segments, which are the same as the ones characterizing the 33-point model excluding hands and feet. For what concerns the size of each part, calculations are the same as those defined in the previous section with the only exception of the head: it is given the shape of an ellipsoid, whose minor radius is defined as the ratio among the distance between the ears and 1.5 and it is equal to 58% of the major radius. Defining the 33-point model, the major radius was defined considering the position of the mouth, which is now not available anymore and for this reason it is simply defined as a certain percentage of the minor one. The scaling factors are defined making reference to the anthropometric studies conducted in [59]. Loosing informations about the positioning and size of hands and feet may have repercussions on the precision of the estimate of the human pose, leading to approximate results. On the other side, anyway, the evaluation of the position of the indexes is generally associated with larger errors, due to the fact that these parts are more likely to fall outside the field of view of the camera. If this happened, the algorithm would not be able to estimate their location in space and it would be obtained a completely wrong estimate of hands or feet proportions. Besides, another advantage is associated with the lower computational effort related to the fact that, at every time instant, 17 rather than 33 points must be

processed by the algorithm leading to quicker evaluations of body dimensions and poses. The shape and the volume associated with each body part are the same as the ones adopted for the 33-point configuration. For what concerns segmental masses, instead, due to the discretization of the body into a lower number of segments, they must be re-scaled. The choice made consists in the association of hands' and feet's percentages correspondingly to the lower part of the arms and the lower part of the legs. It means that the mass of the lower arm will be equal to (1.3+0.5)% for female individuals and (1.7+0.6)% for male ones, while the mass of the lower leg will correspond to (4.5+1)% for females and (4.8+1.2)% for males. These values are extrapolated from Table 2.2. The remaining percentages remain invariant.

2.5. Evaluation of the position of the center of mass

The center of mass (COM) associated with each body segment and with the entire human body is the first fundamental inertial quantity to be evaluated. The positions of the points provided by the deep-learning-based algorithm are initially expressed with respect to a reference frame centered on the plane identified by the camera; thus, in order to evaluate the position of the COM with respect to a more easily analysable frame, they must be re-scaled. More in detail, the choice was that of adopting the right-handed frame shown in Figure 2.6 for the 33-point model and in Figure 2.7 for the 17-point ones. The y axis is vertical and perpendicular to the ground, the x axis is directed toward the left of the body and the z axis is directed outside the page. Since the output image obtained by the camera is mirrored, in both figures the x axis seems to be directed toward the right. The origin of the y axis is set in correspondence of the lower among the points identified by the algorithm, while the origins of x and z axes are identified coincidentally with the mid-values of the positions of left and right hips. Looking at the pictures below it is possible to notice that, since the 17-point model does not include the positioning of heels and foot indexes differently from the 33-point one, the origin of the y axis is set in correspondence of the lower among the two ankles.

To prevent the global reference frame's origin from varying with time, it is considered fixed and coincident with the one evaluated for the initial position of the human body. Every time a new image is acquired by the camera, in fact, the position of the body points estimated by the algorithm changes leading to variable positioning of the origin of the reference frame. Keeping it constant, it is possible to better analyze how the position of the COM varies in time in association with the motion of the body.

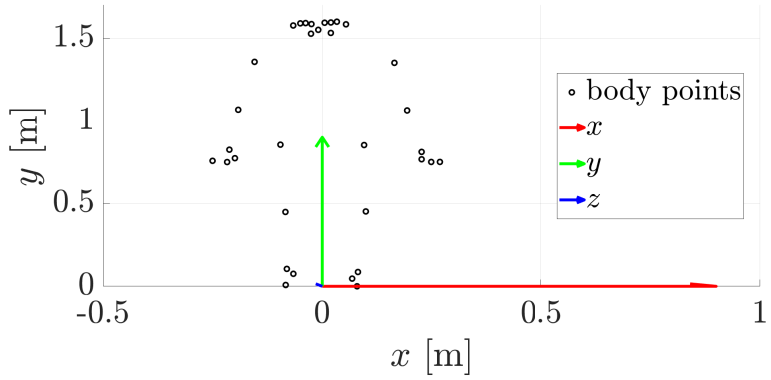


Figure 2.6: Human body global reference frame for a 33-point model.

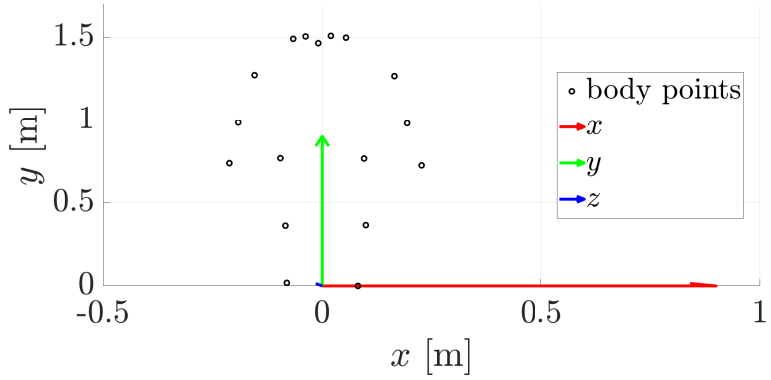


Figure 2.7: Human body global reference frame for a 17-point model.

To evaluate the position of the COM of the entire body it is necessary at first to consider every body part individually. Basing on the assumption that each segment has isotropic density, it is possible to use specific formulas associated with each geometrical shape. Regarding the sphere, the ellipsoid and the elliptical cylinder, their COMs are located exactly at half of their length along their symmetry axes. For the frustum of cone, instead, the following formula is adopted:

$$COM_{frustum} = \frac{h(1/4R^2 + 1/2Rr + 3/4r^2)}{R^2 + Rr + r^2}, \tag{2.6}$$

where h represents the height, R the radius of the major base and r the radius of the minor base. In this way is it obtained the distance of the COM from the bigger base of the frustum. These positions are re-expressed with respect to the global reference frames

defined above and the resulting values are collected in vectors:

$$\boldsymbol{x} = [x_1, x_2, ..., x_n], \tag{2.7a}$$

$$\boldsymbol{y} = [y_1, y_2, ..., y_n], \tag{2.7b}$$

$$\boldsymbol{z} = [z_1, z_2, ..., z_n]. \tag{2.7c}$$

n corresponds to the number of segments adopted to discretize the human body, thus it is equal to 14 for the 33-point models and to 10 for the 17-point ones.
Finally, the position of the COM of the entire body can be determined performing a mass weighted mean of the body segments' COM positions along x, y and z directions:

$$X_{COM} = \frac{\sum_{i=0}^{n} m_i x_i}{M}, \tag{2.8a}$$

$$Y_{COM} = \frac{\sum_{i=0}^{n} m_i y_i}{M}, \tag{2.8b}$$

$$Z_{COM} = \frac{\sum_{i=0}^{n} m_i z_i}{M}. \tag{2.8c}$$

In these final expressions M is the total body mass, m_i is the mass of the i^{th} segment, x_i, y_i and z_i are respectively the positions along x, y and z of the COM of the i^{th} segment expressed in the global reference frame.

2.6. Evaluation of the inertia tensor

The inertia tensor is a symmetric tensor generally written using the following formulation:

$$\boldsymbol{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix}. \tag{2.9}$$

The elements on the diagonal are the moments of inertia with respect to x, y and z axes and they signify the resistance to changes in rotational speed of the body about these axes. Off-diagonal terms, instead, are called 'products of inertia' and they are indices of an imbalance in the distribution of the mass.
The procedure followed for the identification of the inertia tensor of the entire body is similar to that adopted for the positioning of the COM: initially it is evaluated the matrix associated with each body segment individually and then adequate transformations, which include rotation and translation, are applied to find the overall inertia tensor.

2.6.1. Body segment inertia tensor

The segmental inertia tensors are initially evaluated with respect to local reference frames, defined individually for each body part. The y axis of each local frame is aligned with the axis of symmetry of the corresponding body segment. Considering axisymmetric parts, which are the ellipsoid, the frustums of cone and the sphere, the orientation of x and z axes is indifferent as long as it defines a right-handed 3D reference frame with the y axis already identified. A visual representation is provided in Figure 2.8 and 2.9.

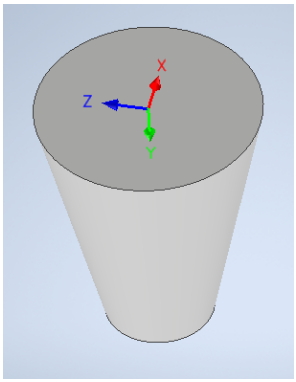


Figure 2.8: Local reference frame of the frustum of cone.

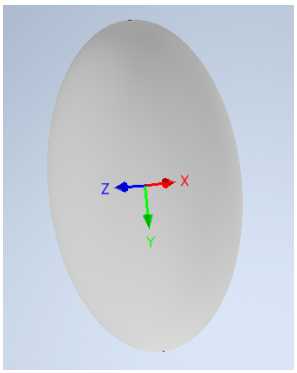
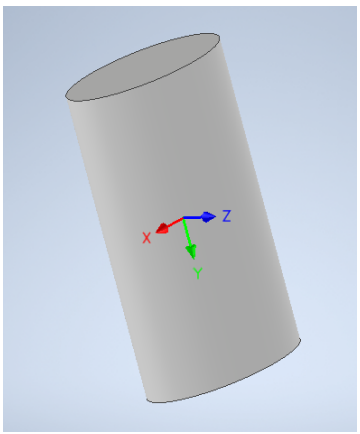
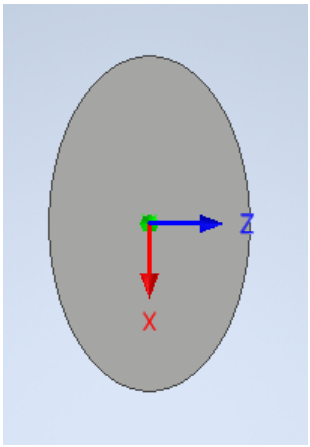


Figure 2.9: Local reference frame of the ellipsoid

Regarding the elliptical cylinder, since it is the only non-axisymmetric segment, x and z axes are required to be aligned with its semi-axes, as shown in Figure 2.10a and 2.10b.



(a) Lateral view



(b) Top view

Figure 2.10: Local reference frame of the elliptical cylinder.

In order to obtain diagonal inertia tensors, whose elements are called 'principal moments

of inertia', the origin of the local reference frames is placed in the COM of each body part. In this way calculations result quicker and easier.
Considering each geometrical shape adopted for the identification of the geometric model, it is possible to define segmental inertia matrices using the following formulas. Starting from the ellipsoid, which is used to model 'head and neck' body part, the expression used is:

$$\boldsymbol{I}_{ellipsoid} = \frac{m}{5}\begin{bmatrix} r^2+R^2 & 0 & 0 \\ 0 & 2r^2 & 0 \\ 0 & 0 & r^2+R^2 \end{bmatrix}. \tag{2.10}$$

For the elliptical cylinder, which represents the torso, and the sphere, which is used to model hands, the inertia matrices are:

$$\boldsymbol{I}_{cylinder} = \frac{m}{12}\begin{bmatrix} 3r^2+h^2 & 0 & 0 \\ 0 & 3(R^2+r^2) & 0 \\ 0 & 0 & 3R^2+h^2 \end{bmatrix} \text{ and} \tag{2.11}$$

$$\boldsymbol{I}_{sphere} = \frac{2}{5}m\begin{bmatrix} r^2 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \end{bmatrix} \tag{2.12}$$

Finally, the principal moments of inertia of the frustum of cone, due to their length, are defined separately along each local reference frame axis through the following expressions:

$$I_{xx,frustum} = \frac{1}{30}m\frac{3R^2+3r^2+h^2+(3R^4+3r^4+6r^2R^2+4h^2(R^2+r^2))}{3(R^2+r^2)}$$
$$I_{yy,frustum} = 3m\frac{R^4+R^3r+R^2r^2+Rr^3+r^4}{10(R^2+Rr+r^2)}$$
$$I_{zz,frustum} = I_{xx,frustum}$$

2.6.2. Rotation and translation of the inertia tensor

In order to evaluate the inertia tensor representative of the entire human body, it is necessary to combine all the previously assessed segmental contributions re-expressing them with respect to a Common Reference Frame (CRF), which is chosen as the one centered in the COM of the body with the axes having the same direction as those defined in Figure 2.6 and 2.7
The first passage to be executed is the rotation of the segmental inertia matrices so to align them with the orientation of the CRF. For this purpose, each body segment is associated with a direction cosine matrix (DCM), which indicates the relative orientation among its

local frame and the CRF. Considering a generic local reference frame L identified by the three unit vectors $\hat{l_x}$, $\hat{l_y}$, $\hat{l_z}$ and a global reference frame G identified by $\hat{g_x}$, $\hat{g_y}$, $\hat{g_z}$, the DCM is defined as follows:

$$\boldsymbol{DCM} = \begin{bmatrix} \hat{g_x} \cdot \hat{l_x} & \hat{g_x} \cdot \hat{l_y} & \hat{g_x} \cdot \hat{l_z} \\ \hat{g_y} \cdot \hat{l_x} & \hat{g_y} \cdot \hat{l_y} & \hat{g_y} \cdot \hat{l_z} \\ \hat{g_z} \cdot \hat{l_x} & \hat{g_z} \cdot \hat{l_y} & \hat{g_z} \cdot \hat{l_z} \end{bmatrix}. \tag{2.14}$$

To have a clearer idea about how the rotation is performed, in the image below (Figure 2.11) is shown a local reference frame before and after the rotation, which is performed using the DCM. The axes of the frame, which were oriented in accordance with the position of the corresponding body segment, after the rotation are perfectly aligned with the global reference frame.

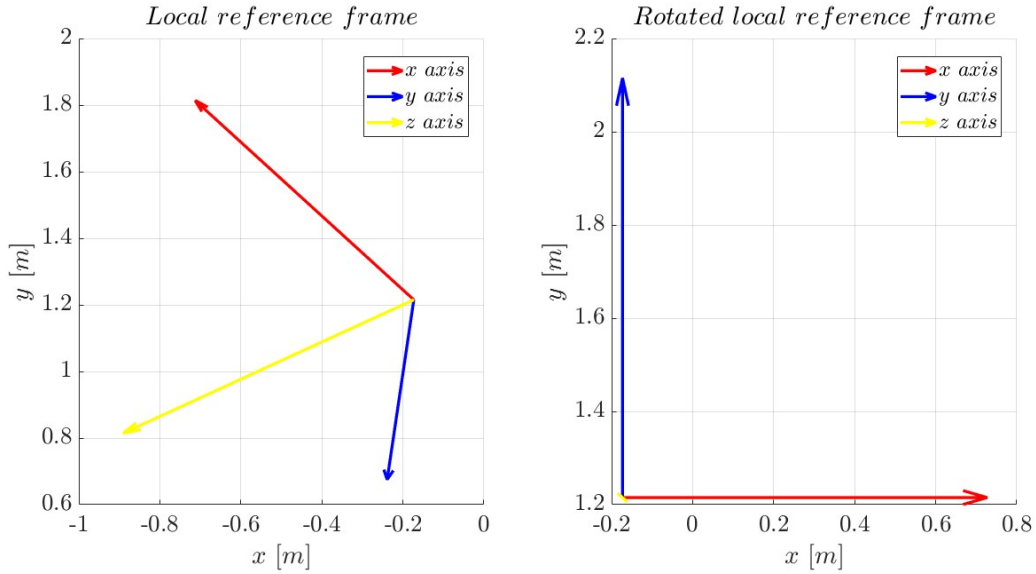


Figure 2.11: Example of rotation of a local reference frame using the direction cosine matrix (DCM).

The rotation of each segmental inertia tensor is consequently carried out performing the multiplication shown in Equation 2.15.

$$\boldsymbol{I}_{rotated} = \boldsymbol{DCM} \cdot \boldsymbol{I}_{local} \cdot \boldsymbol{DCM}' \tag{2.15}$$

The final transformation to be executed it the translation of the rotated inertia matrix into the origin of the Common Reference Frame, which coincides with the position of the COM of the entire body. If the distance among the origin of the local and the global

frames is identified by the vector $\mathbf{r} = [r_x, r_y, r_z]$, the translation is obtained applying the following summation:

$$\boldsymbol{I}_{translated} = \boldsymbol{I}_{rotated} + \boldsymbol{I}_{translation}, \tag{2.16}$$

where

$$\boldsymbol{I}_{translation} = m \begin{bmatrix} r_y^2 + r_z^2 & -r_x r_y & -r_x r_z \\ -r_y r_x & r_x^2 + r_z^2 & -r_y r_z \\ -r_z r_x & -r_z r_y & r_x^2 + r_y^2 \end{bmatrix}. \tag{2.17}$$

The overall human body inertia tensor is finally obtained summing up all the segmental contributions obtained after the rotation and translation of the axes. Besides, it is noteworthy that, since the order of the transformations does not affect the final results, rotation and translation operations can be interchanged.

2.6.3. Inertial ellipsoid

The inertial ellipsoid constitutes an alternative way of representing graphically the inertia tensor. The length of its principal semi-axes and their directions depends on the body's principal moments and directions of inertia. It could be helpful for a better visual understanding of the body's mass distribution and resistance to rotation around various axes. The general expression of the inertial ellipsoid is the following:

$$I_{xx}x^2 + I_{yy}y^2 + I_{zz}xz + 2I_{xy}xy + 2I_{xz}xz + 2I_{yz}yz = 1. \tag{2.18}$$

Taking into account that the equation representing a generic ellipsoid having semi-axis' lengths equal to a, b and c is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{2.19}$$

and considering the principal moments of inertia indicated as $I_{pr,xx}$, $I_{pr,yy}$ and $I_{pr,zz}$, it is possible to define the dimension of the specific inertial ellipsoid under analysis through the following expressions:

$$a = \sqrt{\frac{1}{I_{pr,xx}}}, \quad b = \sqrt{\frac{1}{I_{pr,yy}}}, \quad c = \sqrt{\frac{1}{I_{pr,zz}}}. \tag{2.20}$$

Looking at the ellipsoid geometry and shape, it is possible to make multiple considerations. Since the length of the semi-axes and consequently the volume of the ellipsoid are inversely proportional to the principal moments of inertia, the bigger their values are, the smaller the resistance to rotation about these axes will be. For what concerns the shape, an

elongated ellipsoid is indicative of an asymmetric mass distribution of the associated body, while a rounder shape suggests that the mass distribution is more symmetrical and thus the resistance to rotation is more balanced about different axes.
Another important aspect to highlight is the orientation, which should be representative of the direction of the principal axes of inertia. The ellipsoid, in fact, is aligned with these axes and usually its COM is placed in correspondence of the COM of the entire body. Since the aim of the present study is to show the differences mainly in terms of volumes and orientations of the inertial ellipsoids associated with different body poses, its COM will be simply placed in the origin.

2.7. Model validation

The results are validated through a dual approach, involving both literature reviews and the creation of a CAD model. The first mentioned method, namely the comparison with values found in literature, has been the most commonly utilized over the years to validate various geometric models. The first ever built geometrical model, which was introduced by Havanan in 1964 [27], in fact, was developed also taking into consideration the resulting inertial values obtained experimentally by Dempster [16]. Furthermore, in more recent years, Nikolova [56] and Robertson [30] validated their models comparing their results with those obtained using the regression equations developed by Zatsiorsky [26]. Considering the accuracy of its results and its well-established use as a validation method, Zatsiorsky's regression-based approach has been chosen to validate also the numerical outcomes obtained using the model developed in the present thesis. As Zatsiorsky conducted a comprehensive study on male individuals, but not on females, reference will be made to the analysis conducted by De Leva in [22]. In this work, in fact, De Leva extended all regression equations to include female figures, making it a more suitable reference.
The second method, namely the reproduction of the geometric model in a CAD software, is adopted to check the correctness of the calculations concerning the position of the COM and the inertia tensor for each body segment and for the entire human body. This choice has been employed in multiple studies, in particular in those conducted by Nikolova in [31] and [61]. Her purpose was that of providing reliable data of mass-inertial parameters associated with the human body in several basic positions such as standing erect, standing with arms over the head or spreading arms and legs. In the present thesis it has been decided to adopt the CAD software 'Inventor'. Considering a specific body pose, the discretized model of the human body is reproduced making reference to the positions of the points identified by the deep-learning-based algorithm and to the dimension of each body segment evaluated applying the procedure defined in Section 2.4. In Inventor the origin

of the reference frame is placed in correspondence of the mid-point between the hips, thus to locate all the segments in the assembly it is created an Excel file containing the values of distance of each segment's extremity from this origin. In this way it is possible to guarantee that the pose is faithfully reproduced. Once all body segments have been correctly positioned in space and they have been associated with the mass percentages adopted for the model, the CAD software automatically provides the exact positioning of the center of mass and the values of the inertia tensor for each body part and for the entire discretized body.

2.8. Sensitivity analysis

A sensitivity analysis is conducted for both the 17-point and the 33-point models with the purpose of assessing how changes in the input points' positions detected by the algorithm impact the results in terms of position of the COM and inertia tensor of the entire human body. This kind of analysis can be helpful in understanding how much sensible the inertial values are to a relatively small change of position of the various body parts, which may be associated with the resolution of the camera. The smaller is the resolution, in fact, the higher is the probability that the position of each one of the 17 or 33 body points deviate from the actual value, resulting in slightly different inertial quantities. Quantifying this variation allows to have a better understanding of the system behaviour, improving its reliability and it supports a more robust decision-making. Besides it provides insights into the areas where the model is particularly sensitive or insensitive to changes, estimating the uncertainty associated with the resulting values.
The sensitivity analysis has been conducted considering a well-defined fixed body pose and generating normally distributed random numbers with null mean and different standard deviations. For each standard deviation, which is made vary between 0 and 0.5 m for the estimation of the variability of the position of the COM and between 0 and 0.5 $kg\ m^2$ for the moments of inertia, the randomly generated values are summed to the positions of the points provided by the deep-learning-based algorithm, leading to a set of input values which deviate from the original one. This procedure is iterated 1000 times and the outcomes are presented considering the Gaussian distribution of the resulting values of position of the COM and moments of inertia about the three axes. Additionally, also the histograms associated with each distribution are plotted in order to provide a graphical depiction of the probability of the resulting values to fall within a certain range.

3 | Results

In this Chapter, the numerical outcomes obtained from the procedures described in detail in the previous section are presented and discussed. The primary objective of this research is to establish a method for the real-time assessment and update of inertial properties of the human body, distinguishing among different types of models. The results presented first pertain to the 33-point model. In a preliminary analysis, two well-defined body poses, that will be indicated as 'open arms' and 'closed arms', are examined with the aim of highlighting their differences in terms of position of the center of mass and of inertia tensor, considering additionally their representation through inertial ellipsoids. Besides a validation of the model and, more specifically, of the mathematical procedures adopted, is conducted through a bibliographic comparison and the development of a CAD model in Inventor. At this point, the results concerning the variation of inertial properties over time are presented. To perform the update correctly, it is fundamental to take into consideration both the filtering mechanism for the elimination of out-of-trend data, as defined in Section 2.3, and the averaging of segments' lengths necessary for the correct identification of the geometrical model, as better explained in Section 2.4.1. The same sequence is followed for the presentation of the resulting values associated with the reduced 17-point model identified using MediaPipe and, finally, a comparison with the models obtained through the adoption of TensorFlow's and YOLOv8's algorithms is performed.

3.1. 33 point model

3.1.1. 'Open arms' and 'closed arms' configurations

The test has been conducted on a female individual with height and mass respectively equal to 1.58 m and 55 kg. Starting from the 33 three-dimensional points identified by MediaPipe and following the procedure defined in Chapter 2.4.1, it has been estimated dimension, volume, mass and density associated with each body part. The resulting values are listed in Table 3.1. Regarding the nomenclature, L indicates the length of the

body part, R its major radius and r its minor radius.

Body segment	**Dimensions** (cm)	**Volume** $(m^3 \cdot 10^3)$	**Mass** (kg)	**Density** (kg/m^3)
Head and Neck	$R_{HE} = 13.30$ $r_{HE} = 8.08$	3.64	3.69	1013
Torso	$L_{TO} = 51.00$ $R_{TO} = 15.99$ $r_{TO} = 9.59$	24.58	24.81	1009
Upper arm	$L_{UA} = 29.13$ $R_{UA} = 4.22$ $r_{UA} = 3.86$	1.49	1.21	810
Lower arm	$L_{LA} = 25.59$ $R_{LA} = 3.86$ $r_{LA} = 2.40$	0.80	0.72	892
Hand	$R_{HA} = 3.55$	0.19	0.28	1468
Thigh	$L_{TH} = 40.71$ $R_{TH} = 7.62$ $r_{TH} = 9.59$	5.96	8.03	1347
Shank	$L_{SK} = 34.59$ $R_{SK} = 9.59$ $r_{SK} = 3.55$	2.53	2.48	978
Foot	$L_{FO} = 13.74$ $R_{FO} = 3.55$ $r_{FO} = 2.84$	0.44	0.55	1242

Table 3.1: Dimension, volume, mass, and density of each body segment for the 33-point model identified by MediaPipe.

All these quantities are not dependant on the pose of the human body, but only on the specific individual under test and for this reason they are considered fixed in time. Also the segmental inertial properties, which include the positions of segmental centers of mass and segmental inertia tensors, both defined in local reference frames, are not variable with time. In Table 3.2 are shown the positions of the segmental COMs expressed as a percentage of the total length of the associated body part. For instance, 50% indicates that the center of mass is placed along the axis of symmetry of the segment exactly in the

middle of its volume. Thanks to the hypothesis of segmental isotropic density, this is valid for 'head and neck' (ellipsoid), 'torso' (elliptical cylinder) and 'hand' (sphere) body parts. Regarding the other segments, whose shape is a frustum of cone, their COMs' positions are expressed as a percentage of the segments' length starting from their major bases, assuming as usual that this position is located along the axis of symmetry of the segment. It is possible to notice that, the bigger is the difference among major and minor rays, the smaller is this percentage, which means that the COM is closer to the major base of the frustum of cone. In the same Table are also shown the values of the principal moments of inertia of each body part, expressed with respect to the local reference frames defined in Section 2.6.1. The principal moments of inertia with respect to the y axis, which is the one corresponding with the axis of symmetry, are always less than or equal to the others. The inertial values along the other two axes, instead, are equivalent for all body segments with the only exception of the torso, which has a non-axisymmetric geometrical shape.

Body segment	**Center of mass**	**Ixx** ($kgm^2 \cdot 10^3$)	**Iyy** ($kgm^2 \cdot 10^3$)	**Izz** ($kgm^2 \cdot 10^3$)
Head and neck	50%	17.85	9.62	17.85
Torso	50%	594.81	215.63	696.29
Upper arm	48.55%	8.51	0.99	8.51
Lower arm	42.32%	3.84	0.38	3.84
Hand	50%	0.14	0.14	0.14
Thigh	46.05%	113.57	19.06	113.57
Shank	41.64%	24.63	3.12	24.63
Foot	46.31%	0.96	0.29	0.96

Table 3.2: Position of the COM and principal moments of inertia expressed in the local reference frame of each body segment for the 33-point model.

Until this point, since all the quantities discussed are pose invariant, it has been considered an individual in a generic position. If a deeper analysis concerning the inertial properties of the entire human body is required, more specific conditions must be taken into account. For this purpose, it is going to be conducted a comparative study of two different configurations: 'open arms' and 'closed arms' poses. The subject is standing in front of the camera with different arrangements of the limbs. In the first configuration arms are resting at the sides and legs are fully closed, while in the second one arms are raised and legs are slightly open. In Figures 3.1, together with the points identified by MediaPipe, are shown the positions of the segmental COMs and of the COM of the entire

human body for the 'closed arms' pose. The resulting positions associated with the 'open arms' pose, instead, are presented in Figure 3.2.

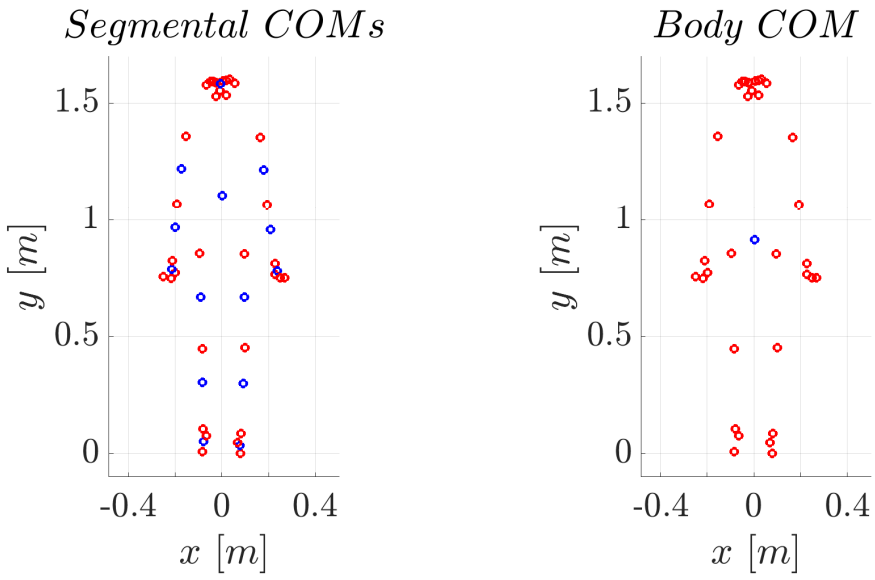


Figure 3.1: Position of the points identified by MediaPipe (red circles) and of the COMs of each body segment and of the entire human body (blue circles) for the 'closed arms' pose.

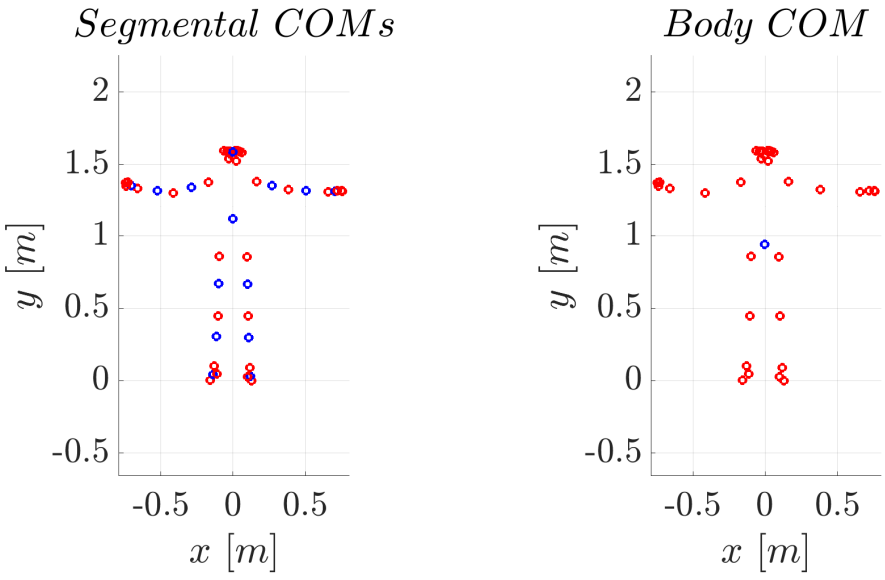


Figure 3.2: Position of the points identified by MediaPipe (red circles) and of the COMs of each body segment and of the entire human body (blue circles) for the 'open arms' pose.

The positions of the COM of the entire body for these two configurations are the following:

$$\boldsymbol{COM}_{closed} = [0.25 \quad 91.40 \quad 3.92]\,\text{cm},$$

$$\boldsymbol{COM}_{open} = [-0.21 \quad 94.16 \quad 4.00]\,\text{cm}.$$

The resulting inertia tensors, instead, are:

$$\boldsymbol{I}_{closed} = \begin{bmatrix} 7.38 & 0.03 & -0.04 \\ 0.03 & 0.66 & -0.17 \\ -0.04 & -0.17 & 7.79 \end{bmatrix} \text{kg m}^2,$$

$$\boldsymbol{I}_{open} = \begin{bmatrix} 8.14 & 0.03 & -0.10 \\ 0.03 & 1.45 & -0.36 \\ -0.10 & -0.36 & 9.28 \end{bmatrix} \text{kg m}^2.$$

The position of the COM is expressed with respect to the global reference frame defined in Chapter 2.5. The movement of the arms is almost symmetric with respect to the sagittal plane of the body and this is the reason for what the positions of the COM along x and z for the two configurations are very similar. For what concerns its values along y, instead, the arms' motion exerts a more substantial influence on the resulting quantities, leading to a more pronounced disparity between them.
The inertia tensor values are referred to a reference frame having the axis with the same orientation as those defined in Chapter 2.5, but with the origin centered in the COM of the entire human body. It is possible to notice that, in particular for the 'closed arm' configuration, the products of inertia are much smaller than the elements on the diagonal. This is another consequence of the high symmetry of the human pose under analysis. The moments of inertia along the axes, which are the diagonal values of the matrix, as well as the principal moments of inertia, are bigger for the first configuration than for the second. The outcome is perfectly in line with expectations: the segmental properties are invariant in the two configurations, thus, since the second pose is characterized by a farther distribution of the mass with respect to the COM of the body, it was expected that the resulting principal moments of inertia would have been higher.

In order to provide specific inertial values for a wider range of body poses, it is introduced a new configuration (Figure 3.3) characterized by a lower symmetry with respect to the previous ones.

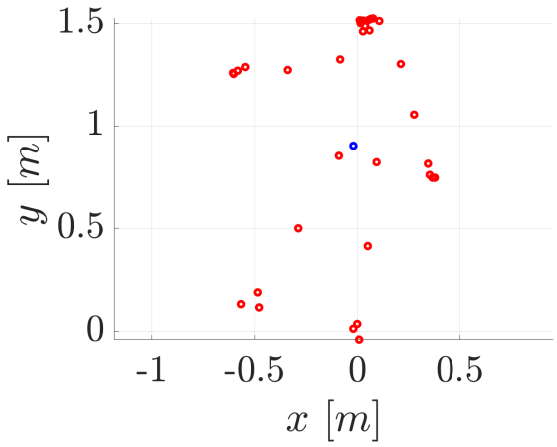


Figure 3.3: Position of the points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for an asymmetric body pose.

In this case the resulting quantities are:

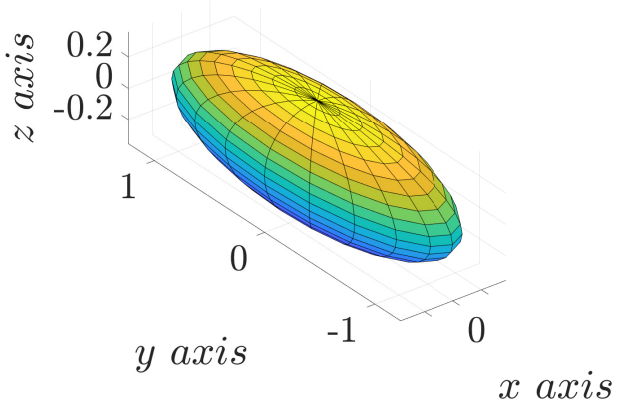
$$\boldsymbol{COM}_{asymmetric} = [-2.1 \quad 94.16 \quad 1.11]\,\text{cm},$$

$$\boldsymbol{I}_{asymmetric} = \begin{bmatrix} 7.20 & -1.03 & -0.18 \\ -1.03 & 1.69 & -0.27 \\ -0.18 & -0.27 & 8.61 \end{bmatrix} \text{kg m}^2.$$

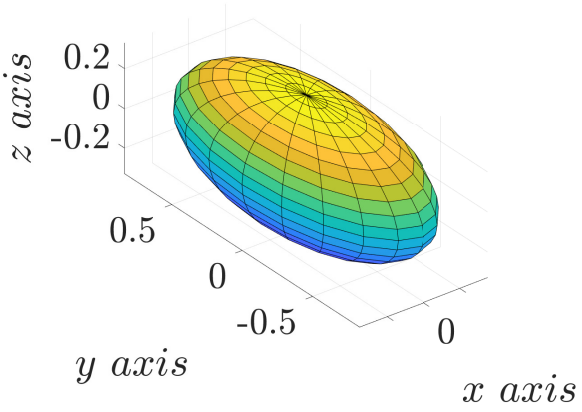
As expected, the position of the COM is farther from the origin of the global reference frame. Besides, the products of inertia are more significant in correlation with the moments of inertia about the axes, which entails their bigger influence on the resulting principal moments of inertia.
The same reasoning has been followed for various body poses, leading always to reasonable results. The analysis that has been conducted in this chapter is useful for verifying qualitatively the coherence of the inertial values obtained referenced to different body poses.

Inertial ellipsoid

The inertia tensors associated with the body poses analysed up to this point can be graphically visualized through the inertial ellipsoid representation. It is a powerful tool for the quick evaluation of various dynamic properties of the human body and it allows, through a dimensional comparison, to highlight the differences in terms of rotational resistance between different configurations. Taking into account the 'closed arms' and 'open arms' body poses, the resulting inertial ellipsoids are shown in Figure 3.4. This



(a) 'Closed arms' configuration



(b) 'Open arms' configuration

Figure 3.4: Inertial ellipsoid for two different body poses plotted without considering the rotation of the principal axes of inertia. The unit of measurement of x, y and z axes is $\sqrt{\text{kg} \cdot \text{m}^2}^{-1}$

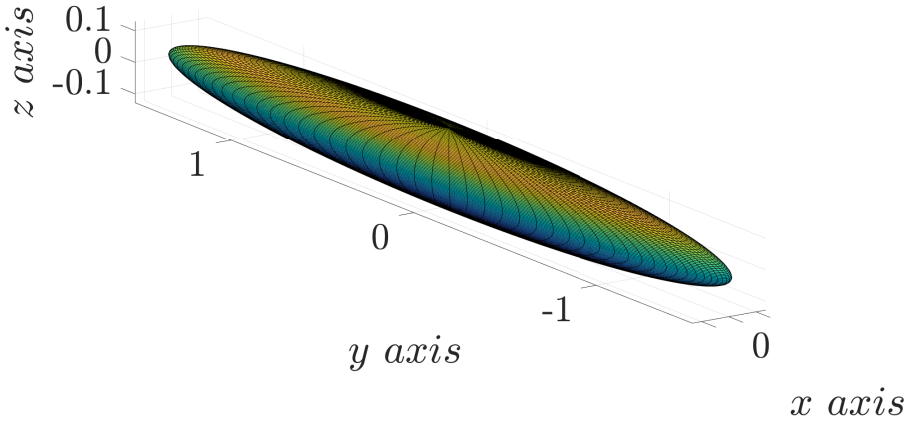
first representation does not take into consideration the rotation of the principal axes of inertia, but they are kept coincident with the axes of the global reference frame. Initially this simplification is adopted in order to better visualize the differences in terms of volume and semi-axis' lengths between the two configurations. For what concerns the dimensions of the ellipsoids, the 'closed arms' pose is characterized by the following values:

$$a = 0.36\sqrt{\text{kg m}^2}^{-1}, \quad b = 1.24\sqrt{\text{kg m}^2}^{-1}, \quad c = 0.36\sqrt{\text{kg m}^2}^{-1}, \quad V = 0.68\sqrt{\text{kg m}^2}^{-\frac{3}{2}},$$

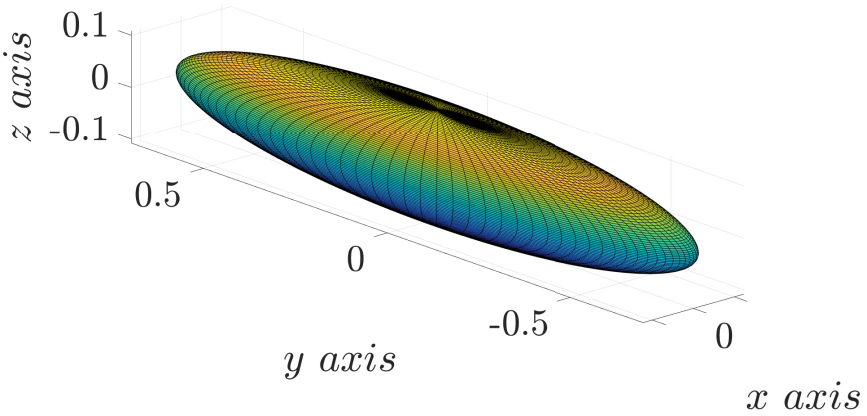
while the 'open arm' configuration by the following ones:

$$a = 0.35\sqrt{\text{kg m}^2}^{-1}, \quad b = 0.84\sqrt{\text{kg m}^2}^{-1}, \quad c = 0.33\sqrt{\text{kg m}^2}^{-1}, \quad V = 0.40\sqrt{\text{kg m}^2}^{-\frac{3}{2}}.$$

a, b and c are the lengths of the semi-axes respectively along x, y and z directions, while V is the volume of the ellipsoid. As expected in the second case, which is associated with bigger principal moments of inertia, the semi-axes lengths and consequently the volume have lower values. The main difference between the two cases is the farther distribution of mass with respect to the y axis in the 'open arms' configuration and, in fact, looking at the resulting values, the biggest variation is associated with the length of the semi-axis b.
For what concerns the orientation of the ellipsoid in space, it is necessary to take into account also the principal directions of inertia.



(a) 'Closed arms configuration'



(b) 'Open arms configuration'

Figure 3.5: Visual representation of the inertial ellipsoid considering the rotation of the principal axes of inertia. The unit of measurement of x, y and z axes is $\sqrt{\text{kg m}^2}^{-1}$

The poses considered until this point are characterized by a good symmetry, thus the actual orientation of the semi-axes will not be so much distant from the one shown in

Figure 3.4. For the 'closed arms' configuration, the rotation around x, y and z axes is respectively equal to 11.4°, -5.25° and -10.16°, while for the 'open arms' case it is equal to 12.62°, -4.89° and 10.03°. The inertial ellipsoids aligned with the principal directions of inertia are shown for both poses in Figure 3.5. Looking at these plots, even if the rotation is not so much evident, it can be observed that the axes of the figure are not anymore perfectly aligned with the orientation of the global reference frame.

Comparison with bibliographic values

A comparative analysis between the outcomes obtained in the present study and the documented findings in the existing literature is performed with the aim of discerning similarities and divergences and of confirming the robustness and generalizability of the applied methodology. To perform this analysis, the research conducted by De Leva in [22] has been taken as a reference. The first elements to be compared are the percentages of mass associated with each body segment, which are listed for the two methods in Table 3.3. These percentages are evaluated making reference to the overall mass of the human body and, as can be noticed looking at the resulting values, the correspondence is quite good. The biggest divergence corresponding to 2.43 kg is achieved for the torso, while the remaining body segments are characterized by smaller differences, in the order of tenths of hundredths of kg.

	Present study (kg)	**De Leva's study** (kg)
Entire body	55	55
Head and neck	3.69 (6.7%)	3.67 (6.68%)
Torso	24.81 (45%)	23.41 (42.57%)
Upper arm	1.21 (2.2%)	1.40 (2.55%)
Lower arm	0.72 (1.3%)	0.76 (1.38%)
Hand	0.28 (0.5%)	0.31 (0.56%)
Upper leg	8.03 (14.6%)	8.13 (14.78%)
Lower leg	2.48 (4.5%)	2.65 (4.81%)
Foot	0.55 (1%)	0.71 (1.29%)

Table 3.3: Comparison of segmental masses between the current study and De Leva's research.

The comparison in terms of positioning of the segmental COMs is presented in Table 3.4. The values defined by Zatsiorsky and then readjusted by De Leva have been evaluated starting from multiple anthropometric measurements of various human subjects and adopting gamma-ray scanning techniques. Instead of associating simplified shapes to each part of the body, they have considered their real dimensions and the position of the COM has been expressed with respect to well-defined proximal or cranial endpoints. It means that the percentages presented in the table are not simply referenced to each body part's length, but they depend on the location of the joint centers, which may not coincide with the extremity of the corresponding segment. With the method adopted in the present work, instead, the shape of each body part has been largely simplified and the location of the joint centers has been assumed coincident with the extremities of the corresponding segments, leading to a different relative positioning of the COM. The major difference is achieved for the COM of the hand, whereas the other segments exhibit variations that remain within a range of no more than 10%.

	Present study	**De Leva's study**
Head and neck	50%	58.94%
Torso	50%	41.51%
Upper arm	48.55%	57.54%
Lower arm	42.32%	45.59%
Hand	50%	74.74%
Upper leg	46.05%	36.12%
Lower leg	41.64%	44.16%
Foot	46.31%	40.14%

Table 3.4: Comparison of the positions of the segmental COMs between the current study and De Leva's research.

Finally, in Table 3.5 are listed the resulting segmental principal moments of inertia. De Leva in his work does not provide directly inertia values, but instead he defines the percent ratios between segments' radii of gyration and lengths (relative radii of gyration). In order to obtain values which are more suitable to perform a comparison, the following

calculations are performed:

$$I_{xx} = m(l \cdot k_{x,rel})^2, \tag{3.1a}$$

$$I_{yy} = m(l \cdot k_{y,rel})^2, \tag{3.1b}$$

$$I_{zz} = m(l \cdot k_{z,rel})^2. \tag{3.1c}$$

$$\tag{3.1d}$$

In this way, starting from the relative radii of gyration ($k_{x,rel}$, $k_{x,rel}$ and $k_{x,rel}$) and the segmental masses (m), it is possible to derive the segmental principal moments of inertia. Looking at the resulting quantities shown in the table, it is possible to notice that the correspondence is quite good. The highest difference is observed for the torso, where we are talking about a variation on the order of tenths of $kg\ m^2$. For what concerns the remaining body parts, instead, there is a closer agreement with the original values provided by De Leva.
For both models the smallest contribution is given by hands a feet, whose principal moments of inertia are at least one order of magnitude smaller than those associated with all other segments. This serves as confirmation that, in case in which a simplification of the geometrical model would be required to reduce computational times and thus to approach better the real-time condition, these components would be the ones that could be omitted. Considering their contribution to the overall inertia tensor, in fact, is does not exceed 10%.

	Present study			De Leva's study		
	Ixx	**Iyy**	**Izz**	**Ixx**	**Iyy**	**Izz**
Head and neck	17.85	9.62	17.85	18.98	14.89	16.04
Torso	594.81	215.63	696.29	753.82	191.81	836.00
Upper arm	8.51	0.99	8.51	7.18	2.32	8.20
Lower arm	3.84	0.38	3.84	3.50	0.47	3.61
Hand	0.14	0.14	0.14	0.39	0.21	0.53
Upper leg	113.57	19.06	113.57	146.26	28.97	150.30
Lower leg	24.63	3.12	24.63	35.25	4.28	36.31
Foot	0.96	0.29	0.96	2.80	0.71	3.31

Table 3.5: Comparison of segmental principal moments of inertia between the current study and De Leva's research. The unit of measurement for all the values is $kg\ m^2 \cdot 10^3$.

It has been chosen to compare quantities whose values are independent of the pose, but depend solely on the characteristic dimensions of the body under analysis. While some

differences exist, they are inevitable given the significant simplification adopted in constructing the geometrical model and the association of De Leva's regression equations with human bodies characterized by specific proportions. One of the considerable advantages of the present study, in fact, lies in its applicability to a broader and more diverse population. Despite these disparities, anyway, a good correspondence between the two methodologies has been identified and consequently it is possible to conclude that the geometrical model developed, despite representing a simplification of human body dimensions and proportions, provides valid values. Thus, it can effectively be utilized to estimate the inertial properties of the human body.

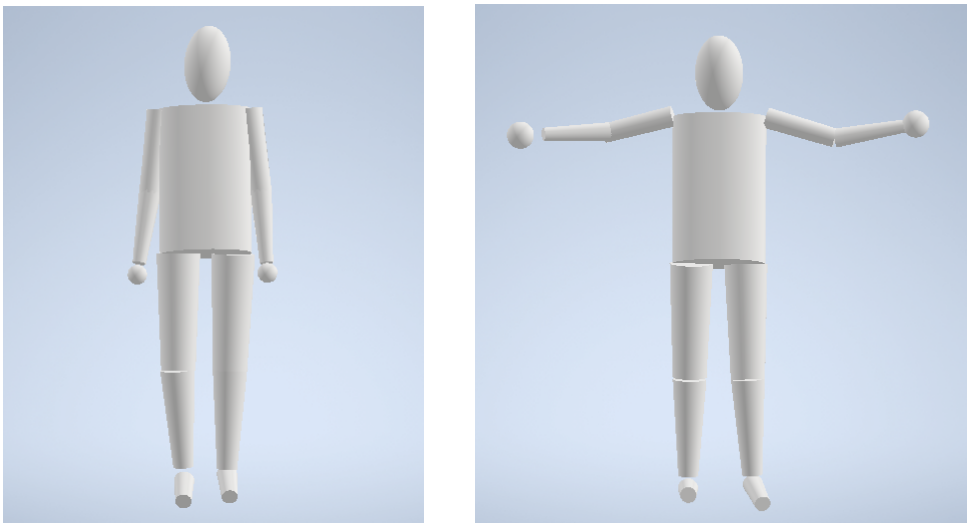
Comparison with the CAD model

The 'open arms' and 'closed arms' poses have been reproduced in the CAD software Inventor in order to check the correctness of the calculations concerning the position of the COM and the inertia tensor for each body segment and for the entire discretized body. The actual points' positions associated with the two configurations are reported in Table 3.6.

Body segment	**Xclosed** (cm)	**Yclosed** (cm)	**Zclosed** (cm)	**Xopen** (cm)	**Yopen** (cm)	**Zopen** (cm)
nose	-0.91	155.23	-0.40	-0.36	156.16	0.45
left eye	-3.80	159.30	3.80	-3.04	159.03	1.05
right eye	1.97	159.66	2.40	2.94	159.39	1.35
left ear	-6.62	157.85	6.10	-6.40	159.35	6.65
right ear	5.41	158.56	4.90	5.99	158.27	3.35
left mouth	-2.57	152.90	3.20	-2.82	154.00	4.85
right mouth	1.96	153.37	2.60	2.04	152.46	0.15
left shoulder	-15.44	135.96	7.50	-17.01	137.85	8.75
right shoulder	16.46	135.36	9.60	16.27	138.39	9.75
left elbow	-19.17	107.08	6.60	-41.55	130.58	6.65
right elbow	19.36	106.69	8.70	38.36	132.85	13.25
left wrist	-21.19	82.46	-0.10	-66.05	133.73	3.05
right wrist	22.67	81.13	1.50	65.87	131.23	12.85
left index	-21.70	75.00	-1.20	-74.63	137.44	5.75
right index	24.86	75.05	2.90	75.35	132.02	19.85
left hip	-9.55	85.49	-0.10	-9.68	85.94	0.45
right hip	9.55	85.27	0.10	9.68	85.64	-0.45
left knee	-8.40	44.88	2.40	-10.40	44.85	0.45
right knee	9.94	45.19	5.50	10.32	44.62	6.05
left ankle	-8.08	10.51	6.30	-13.06	10.27	2.75
right ankle	8.14	8.63	11.10	11.79	8.89	6.65
left heel	-6.60	7.54	5.80	-11.16	4.75	9.75
right heel	6.82	4.65	14.50	9.95	2.75	17.15
left foot index	-8.34	0.79	-4.90	-15.69	0.26	-7.85
right foot index	7.93	0	-0.70	13.00	0	-5.15

Table 3.6: Position of the body segments required to run the Matlab code and to reproduce the poses in Inventor.

Not all 33 points have been taken into account, but only those required to build the geometrical model and to reproduce exactly the human pose in Inventor; thus, for instance, the three-dimensional positions of pinkies or inner eyes are not reported in the Table. Each body segment is reproduced in the CAD software considering the segments' dimensions, masses and densities evaluated previously (Table 3.1) and following the procedure described in Chapter 2.7. Once all body parts have been recreated, the assembly is composed. One of the hypothesis on which the model relies is the symmetry of the body with respect to the sagittal plane, which means that left and right limbs have equal dimensions. This simplification does not correspond to reality and it could be noticed considering the different segments' length obtainable for right and left limbs starting from the positions listed in Table 3.6. This means that recreating body segments in Inventor, the extremities of consecutive body parts will not be perfectly coincident, but there will always be a gap or an overlap, as can be noticed in Figure 3.6.



(a) Closed arms (b) Open arms

Figure 3.6: Reproduction in Inventor of the human pose for 'closed arms' and 'open arms' configurations associated with the 33-point model.

At this point the exact inertial values obtained with the discretized model and with Inventor are considered in order to highlight possible differences or similarities. Starting from the 'closed arms' body pose, looking at Table 3.7 it is immediately possible to notice that the resulting values, both in terms of position of the COM and of principal moments of inertia, are very close. The maximum difference, which is obtained for the position of the COM along direction z and for the principal moment of inertia along x, in fact, is respectively smaller than 1 mm and 0.1 $kg\ m^2$. These quantities represent the 0.1%

and the 1.4% of the actual inertial values obtained via Inventor. They may be associated with numerical errors or with the not perfect positioning of the discretized human body as an assembly in CAD. Consequently, it is possible to assume that this discrepancy is not caused by mistakes in the mathematical passages defined in the code, but only by intrinsic errors associated with the reproduction of the model in the CAD software.

Closed arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** $(kg\ m^2)$	**Iyy** $(kg\ m^2)$	**Izz** $(kg\ m^2)$
Model	0.25	91.40	3.92	7.37	0.06	7.79
Inventor	0.24	91.37	3.85	7.39	0.06	7.80
Difference	0.01	0.03	0.07	0.02	0.00	0.01

Table 3.7: Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'closed arms' configuration.

For what concerns the inertial values obtained for the 'open arms' configuration, looking at Table 3.8 similar considerations could be done. Differently from the previous case, the maximum difference is obtained for the position of the COM along y and for the principal moment of inertia along z, but these value are still respectively smaller than 1 mm and 0.1 $kg\ m^2$.

Open arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** $(kg\ m^2)$	**Iyy** $(kg\ m^2)$	**Izz** $(kg\ m^2)$
Model	-0.21	94.16	4.00	8.12	1.43	9.31
Inventor	-0.21	94.24	4.06	8.10	1.43	9.26
Difference	0	0.08	0.06	0.02	0.00	0.05

Table 3.8: Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'open arms' configuration.

In conclusion, the comparison between the results obtained via Matlab and those derived recreating the body pose in Inventor shows a very good agreement, being characterized by an error which is too small to be influential. This alignment does not represent an actual validation of the numerical values obtained, but it confirms the reliability of the mathematical steps followed for the calculation of the position of the COM and of the inertia tensor. It means that, thanks to this comparison, it is possible to announce that, starting from the discretized model of the human body that has been previously described, the resulting inertial values are exactly those obtained with the Matlab code,

with a minimal margin of error (smaller than 1 mm for the position of the COM and 0.1 $kg\ m^2$ for the principal moments of inertia).

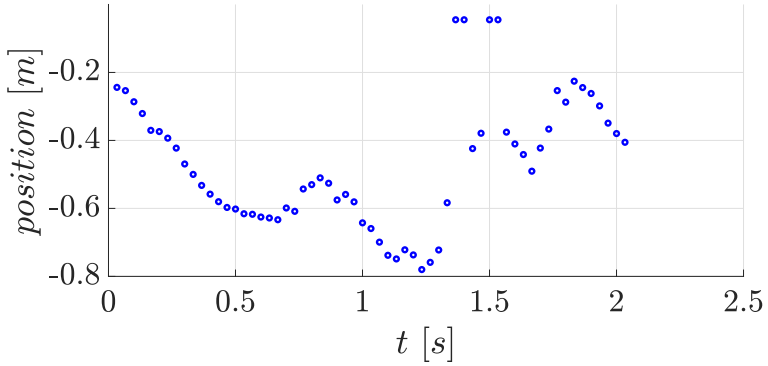
3.1.2. Real-time updating of inertial properties

Iterating the procedure for the evaluation of the inertial properties of the human body every time a new image is made available by the depth camera, it is possible to update the estimate in real-time. Before analysing in detail the numerical outcomes obtained for a moving body, let's discuss the most critical aspects that had to be faced. The first one concerns the possibility of having out-of-trend data in the temporal sequence of the body points' positions evaluated by the algorithm. It may be caused by the resolution of the camera or by its too close positioning with respect to the individual. If this occurs, in fact, performing broad movements the extremities of the body of the subject under test would go outside the visibility range of the camera, not allowing the algorithm to detect their correct positioning. The second relevant possible source of errors is the variability of the length of each segment over time. This is caused by the fact that MediaPipe, as well as all the other algorithms employed in this thesis, have not been designed with the purpose of maintaining constant distances among consecutive points and consequently, every time a new image is made available by the camera, different distances among consecutive points may be detected by the algorithm, leading to variable segments' dimensions.

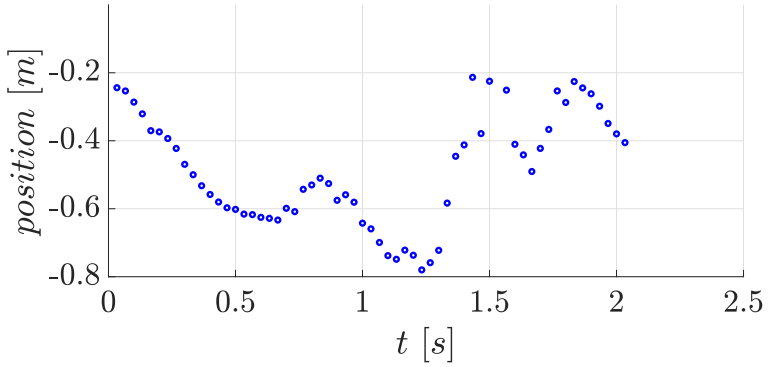
Filtering of out-of-trend data

Considering the body points' positions evaluated by MediaPipe over time, it is possible to detect the presence of out-of-trend data, associated in particular with body extremities such as hands and feet. Taking into account, for example, the variation of position of the left hand along x in time, which is shown in Figure 3.7a, it is clearly noticeable how at some time instants the position of the hand deviate too much from the moving mean, thus requiring a readjustment. Following the procedure described in detail in Section 2.3, it is possible to detect the presence of these scatter points and to substitute them with the corresponding value of the polynomial function that bets fits the correct temporal sequence, leading to the resulting trend shown in Figure 3.7b.

Another example is provided in Figure 3.8a, showing the variation of position along x of the left hand. In this case the deviation, which is visible at the time instant equal to 1.67 s, is much smaller with respect to the one detected for the right hand.
Another difference among the two situations is that, in the first case, the out-of-trend positions tend to approach the origin. This is caused by the fact that, at those critical time instants, the hand of the individual goes outside the range of visibility of the camera



(a) Before filtering



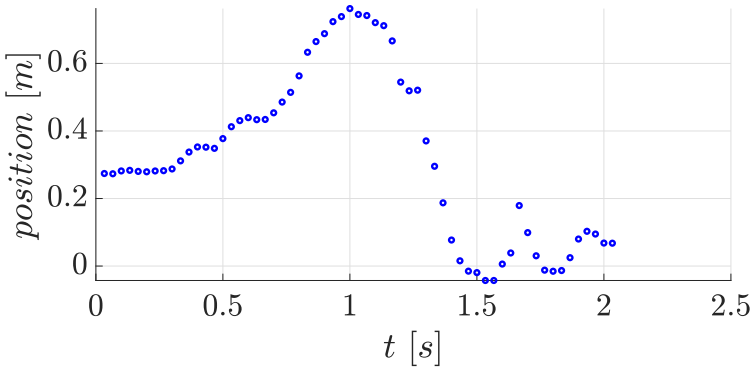
(b) After filtering

Figure 3.7: Position of the left hand along x in time before and after filtering.

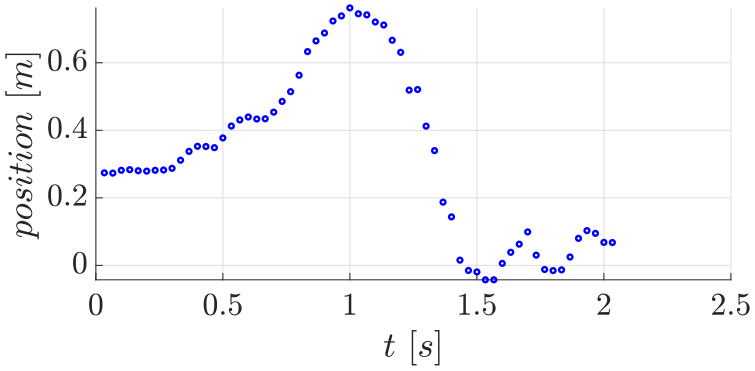
and thus, not being able to correctly estimate its position, MediaPipe places the point close to the origin. For what concerns the right hand, instead, the scatter point is not so critically distant from its position at the previous and following time instants, thus it may be associated with the resolution of the camera. The smaller the resolution is and the farther from the camera the subject is, the larger will be the area associated with a single pixel and thus the higher will be the probability of identifying out-of-trend data. This kind of errors are way less consistent with respect to those represented by the first case. Looking at the filtered temporal sequences, as could have been expected, the readjustment is more effective for the right hand than it is for the left one, since there are less scatter data and they are also closer to the correct trend.

In conclusion, it is possible to say that the application of the filtering mechanism is necessary to avoid taking into consideration wrong estimates of body points' positions, which may lead to the evaluation of erroneous inertial properties. In particular, it is fundamental in those cases where there is the risk for the individual to go outside the range of visibility of the camera, since the error associated with hands' and feet's positioning

would be much more consistent.



(a) Before filtering



(b) After filtering

Figure 3.8: Position of the right hand along x in time before and after filtering.

Body segments' dimensions variability

Evaluating the size of each body segment considering the point's positions made available by MediaPipe at each time instant, it would be obtained a discretized body whose dimensions and proportions vary over time. In Figure 3.9 is represented as an example the variability of the length of the torso. The graphs associated with other important elements of the discretized human body, such as major and minor rays of the ellipsoid representing the head, length and rays of the frustums of cone representing all body arts as well as rays of the cross-section of the elliptical cylinder representing the torso are reported in Appendix A. Placing particular attention on the variability of the length of the upper arm and of the radius of its bigger section, which are reported in Figure 3.10 and 3.11, it is possible to notice that their behaviour in time is the same. This occurs because the radius of the bigger cross section of the upper arm is not defined directly starting from the points identified by MediaPipe, but it is set as equal to a certain percentage of

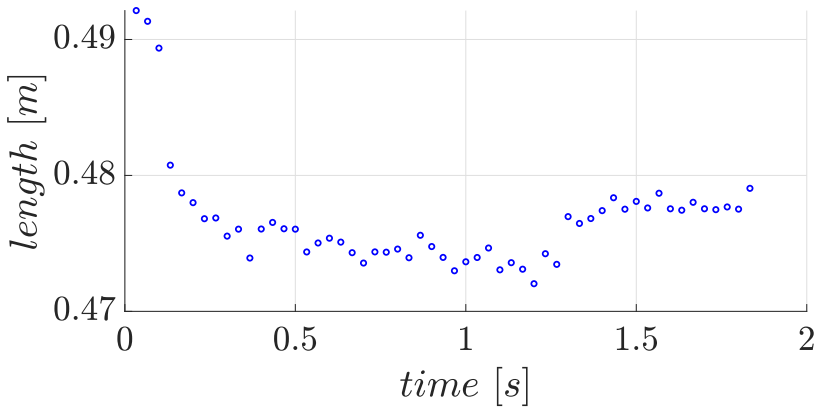


Figure 3.9: Variability of the length of the torso.

the length of the upper arm. The values among which the radius varies are different from those of the length of the arm, but this distinction is not particularly relevant and for this reason the data associated with the variability of the radius over time will not be taken into account for a deeper analysis. The same reasoning can be done for other body parts' characteristic dimensions: the radius of the smaller section of the lower arm has the same trend as its length, bigger and smaller rays of the thigh have the same trend as its length, the smaller radius of the shank has the same trend as its length and, finally, the smaller radius of the foot has the same trend as the bigger one.

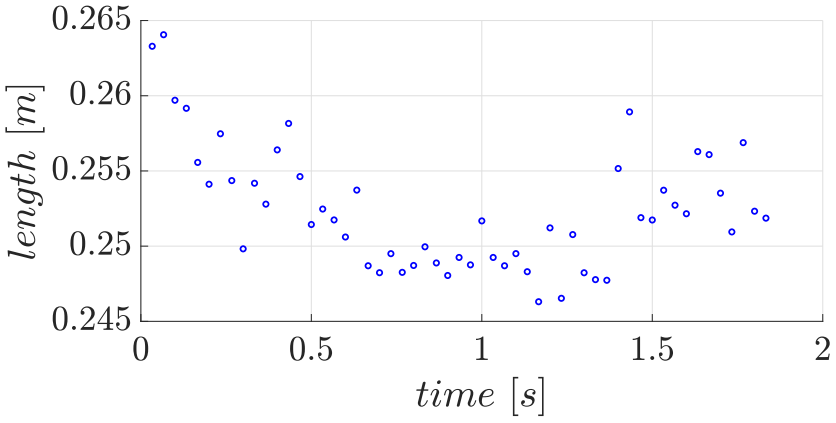


Figure 3.10: Variability of the length of the upper part of the arm.

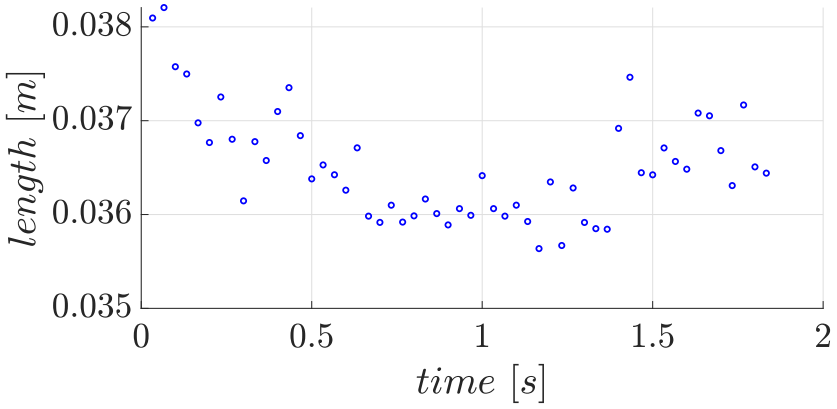


Figure 3.11: Variability of the radius of the bigger section of the upper part of the arm.

In Table 3.9 are reported the values of mean, maximum deviation from the mean and standard deviation of the most significant segments' dimensions.

Body part	**mean** (m)	**max dev** (m)	**max error** (%)	**std dev** (m)
r_{HE}	0.0684	0.0045	6.63	0.0016
R_{HE}	0.1326	0.0086	6.49	0.0030
L_{TR}	0.4767	0.0154	3.24	0.0039
R_{TR}	0.1536	0.0056	3.64	0.0016
L_{UA}	0.2524	0.0117	4.62	0.0041
L_{LA}	0.2225	0.0142	6.38	0.0043
R_{HA}	0.0425	0.0082	19.33	0.0037
L_{TH}	0.3940	0.0180	4.57	0.0076
L_{SK}	0.3588	0.0243	6.77	0.0085
L_{FO}	0.1482	0.0519	35.05	0.0199

Table 3.9: Mean value, max deviation from the mean and standard deviation of the variability of the length of the most significant segments' dimensions.

The max deviation for all the elements is in the order of centimeters or tenth of centimeters and, taking in account also the maximum error expressed as a percentage, it can be observed that the most critical results are obtained for the radius of the hand and the length of the foot. This is probably due to the fact that the extremities of the human body, such as the position of the fingers or of the tip of the toes, are the most difficult to be precisely estimated using the experimental devices at our disposal. Considering the standard deviation, instead, the higher values are associated with lower body parts,

in particular with the length of the thigh, of the stank and of the foot, while the one characterizing the radius of the hand is smaller. It means that, for what concerns the radius of the hand, even if there are some values that deviate significantly from the average, since the standard deviation is small the majority of the elements estimated in time are reliable, thus leading to a good estimate of this segment's length. On the other side, the estimate of the dimension of the lower body parts is more critical due to the higher dispersion of values around the mean, in particular for those associated with the feet.

The values analyzed up to this point are related to a quasi-static person, which means that the recorded individual does not perform broad movements, but he tries to stays as still as possible. This is an important aspect to be taken into account, because if positions associated with more significant movements were taken into account, the results would be much less reliable. Looking at Figure 3.12 and 3.13, in which the variability of the radius of the hand and of the length of the foot for a largely moving person are reported, it is immediately possible to notice that the corresponding dispersion of the results is much wider. In this case, in fact, the standard deviation with respect to the mean is respectively equal to 0.0638 m and 0.0480 m, more than double compared to the previous case.

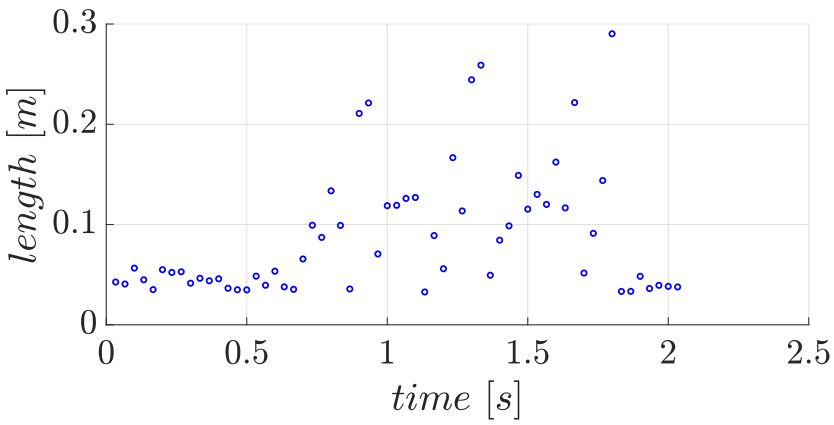


Figure 3.12: Variability of the radius of the hand considering an individual performing broad movements.

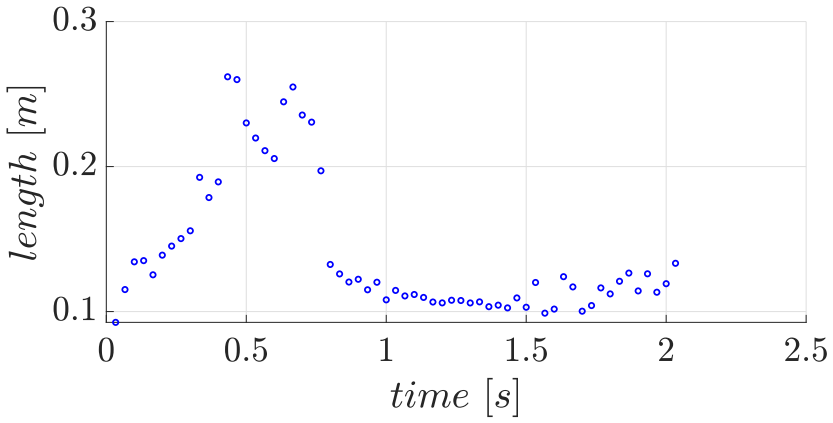


Figure 3.13: Variability of the length of the foot considering an individual performing broad movements.

In conclusion, in order to estimate the dimension of each body segment, it is necessary to perform a preliminary analysis considering the subject as still as possible. Due to the unavoidable variability of the results over time, instead of considering the values obtained instant by instant, the length of each segment is fixed considering the average. The results that will be shown consider a mean performed over the first 60 acquisitions. In general the higher is the number of frames considered, the better will be the estimate, provided that the individual does not perform broad movements that may damage the trend.

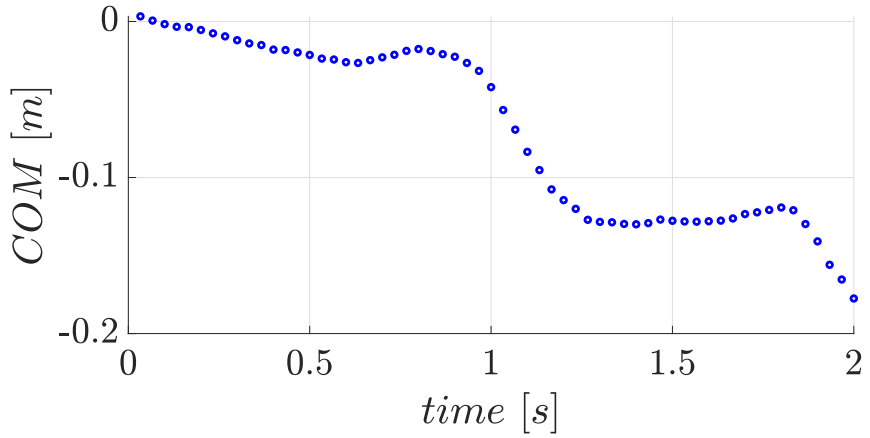
Resulting inertial properties for a moving body

The resulting real-time updated inertial values in terms of position of the COM and inertia tensor of the entire human body are plotted in the following figures. The individual under test is the same on which static tests, whose results have already been shown in Chapter 3.1.1, were based. Following the procedure depicted in detail in the previous subsection, at first it is evaluated the dimension of each body part and consequently the corresponding values of volume, mass and density, which are reported in Table 3.10. These values remain constant over time, while the position of the COM and the elements of the inertia tensor differ at each time instant.

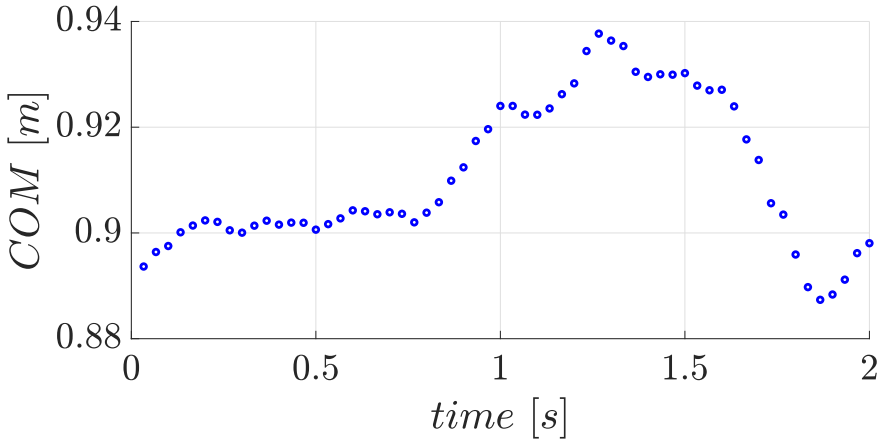
Body segment	**Dimensions** (cm)	**Volume** $(m^3 \cdot 10^3)$	**Mass** (kg)	**Density** (kg/m^3)
Head and Neck	$R_{HE} = 11.89$ $r_{HE} = 6.21$	3.64	1.92	1916
Torso	$L_{TO} = 46.84$ $R_{TO} = 14.94$ $r_{TO} = 8.97$	24.58	19.71	1258
Upper arm	$L_{UA} = 5.34$ $R_{UA} = 3.67$ $r_{UA} = 3.36$	1.49	0.98	1230
Lower arm	$L_{LA} = 24.45$ $R_{LA} = 3.36$ $r_{LA} = 2.29$	0.80	0.62	1153
Hand	$R_{HA} = 4.68$	0.19	0.43	641
Thigh	$L_{TH} = 40.83$ $R_{TH} = 7.64$ $r_{TH} = 6.02$	5.96	6.01	1336
Shank	$L_{SK} = 37.61$ $R_{SK} = 6.02$ $r_{SK} = 3.86$	2.53	2.93	845
Foot	$L_{FO} = 13.85$ $R_{FO} = 3.86$ $r_{FO} = 3.09$	0.44	0.53	1042

Table 3.10: Dimension, volume, mass, and density of each body segment of the 33-point model evaluated for a moving body.

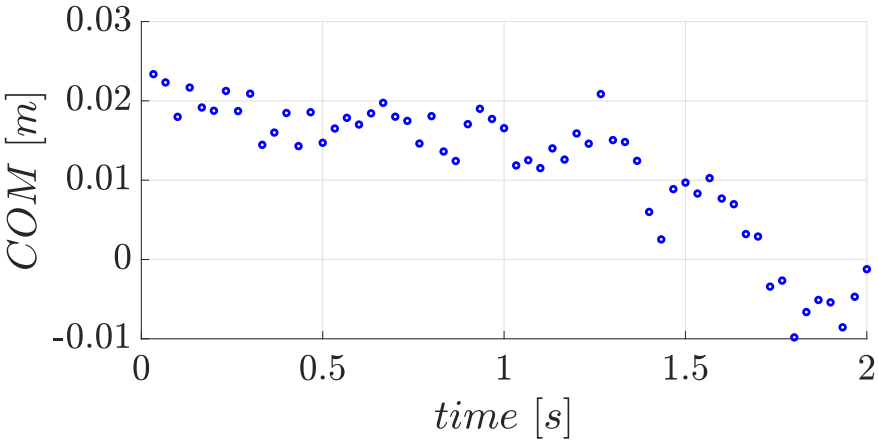
In Figure 3.14 is shown the variation of position of the COM along x, y and z directions during the motion of the human body. It is expressed with respect to the global reference frame defined in Chapter 2.5, having the origin of x and z axes placed in the mid-point between the hips and the one of the y axis in correspondence of the lower among the 33 points identified by MediaPipe. It is possible to notice that the trajectory over time along x and y is continuous, while looking at Figure 3.14c, which is representative of the z-axis, it exhibits a zigzag pattern. This may be associated with the fact that the accuracy of the estimation of the depth position of each pixel in the image captured by the camera is not so high. The minimum variation of position detectable by the camera along z, in fact, is equal to 1 mm, while along x and y it is able to take into account variations in the order of fractions of millimeters. In Figure 3.15 is represented the trend of the elements on the diagonal of the inertia tensor, which are called 'moments of inertia' about the three axes. The moments of inertia about x and z axes are characterized by similar trends: they vary among 7 kgm^2 and 11 kgm^2 and their bigger value is achieved close to 1.5 s. For what concerns the moment of inertia along y, instead, it is associated with a smaller variation (between 0.5 kgm^2 and 2 kgm^2) and the maximum value is obtained close to 0.5 s. Finally, in Figure 3.16 are shown the products of inertia, which are useful for evaluating imbalances in the distribution of the mass. Looking for example at the variation of I_{xy}, noticing that it increases quasi-monotonically over time, it may be concluded that the individual under test is tilting more and more with respect to the xy plane. I_{xz} and I_{yz} are interested by a smaller variation, meaning that they will have smaller influence on the value of the principal moments of inertia.



(a) Position of the COM along x

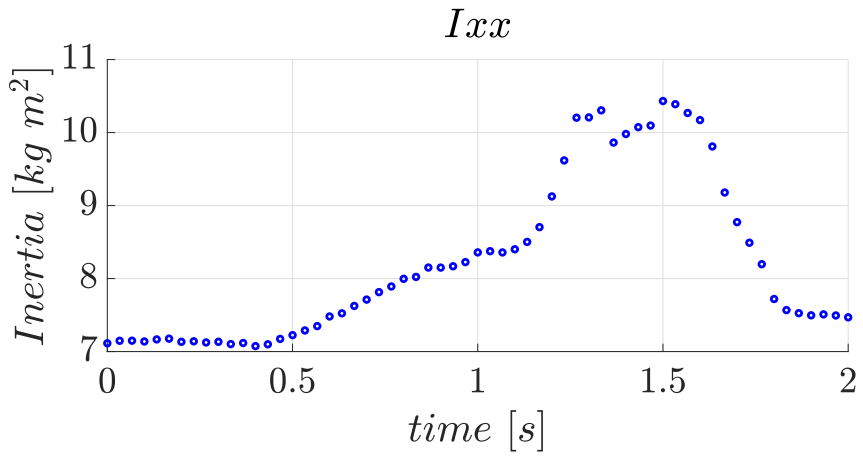


(b) Position of the COM along y

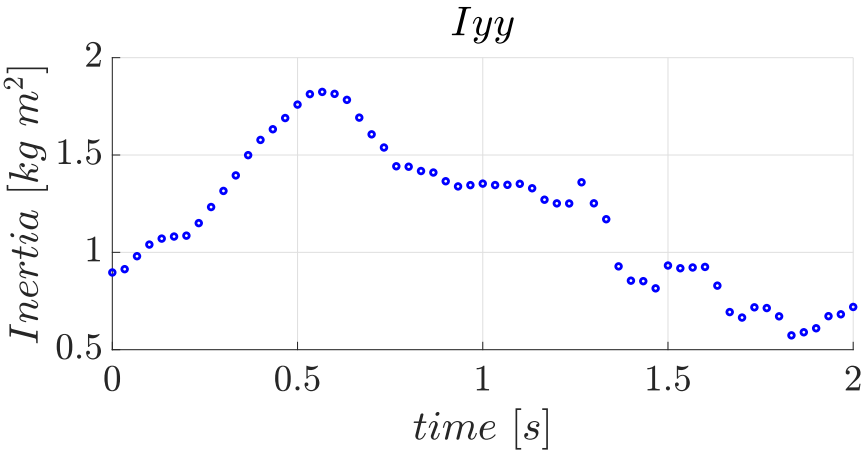


(c) Position of the COM along z

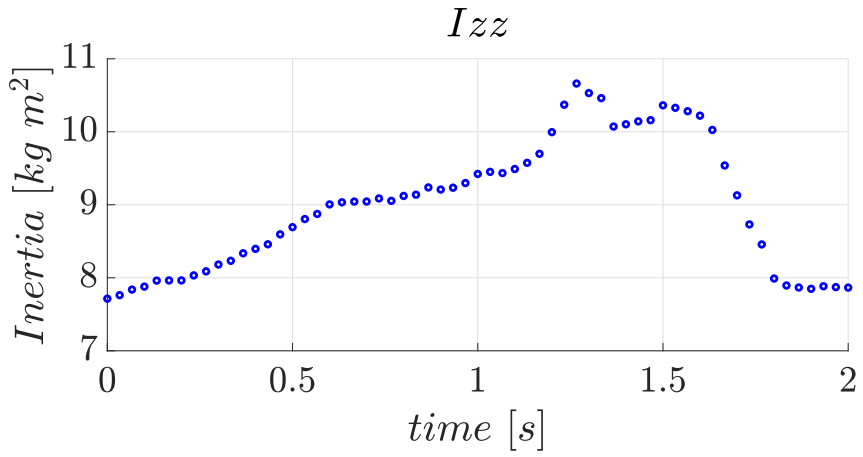
Figure 3.14: Variation of position of the COM during the motion of the body for the 33-point model.



(a) Moment of inertia along x

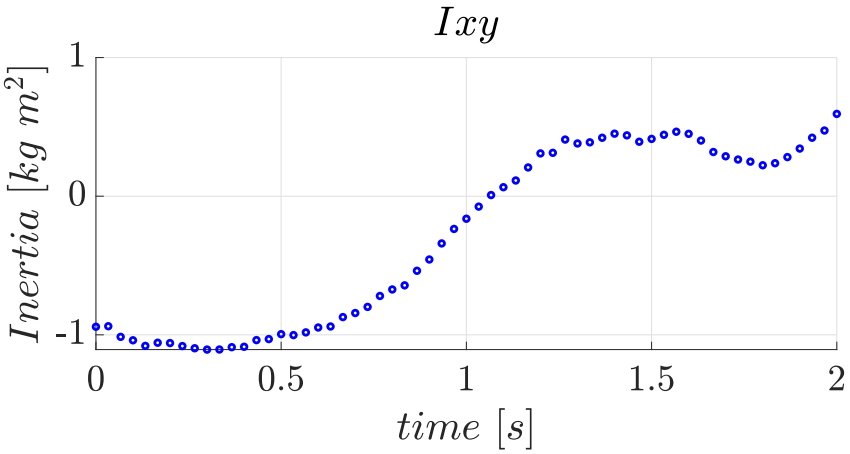


(b) Moment of inertia along y

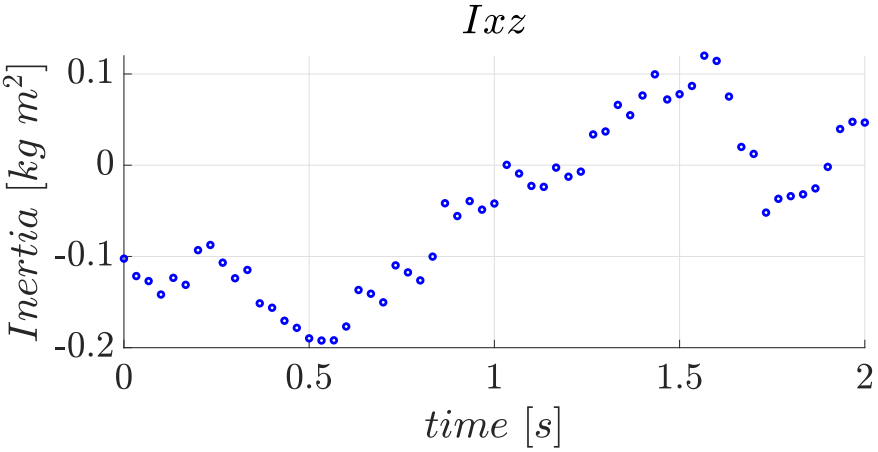


(c) Moment of inertia along z

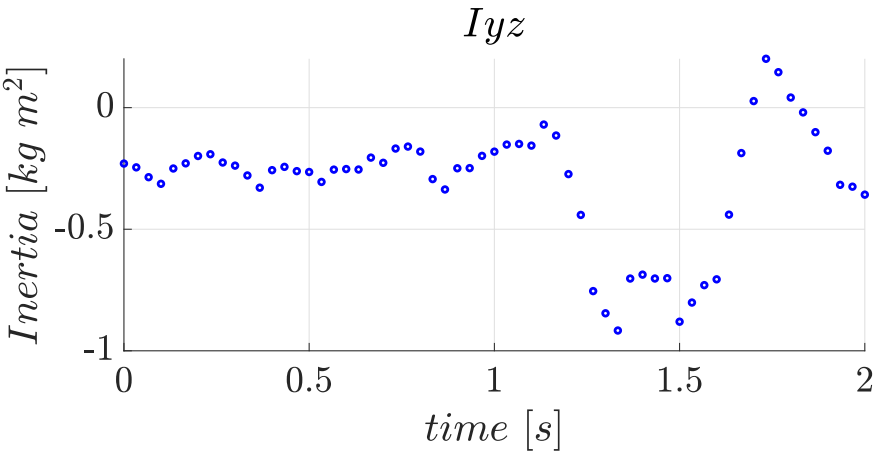
Figure 3.15: Variation of the diagonal terms of the inertia tensor during the motion of the body for the 33-point model.



(a) Product of inertia Ixy



(b) Product of inertia Ixz



(c) Product of inertia Iyz

Figure 3.16: Variation of the off-diagonal terms of the inertia tensor during the motion of the body for the 33-point model.

Sensitivity analysis

The sensitivity analysis for the 33-point model identified by MediaPipe has been conducted following the procedure depicted in detail in Section 2.8. The outcomes concerning the variability of position of the COM are shown in Figure 3.17, while those associated with the moments of inertia are presented in Figure 3.18. In each graph is represented the Gaussian distribution of the resulting data, which is fundamental for the identification of mean and standard deviation, and the histogram, which provides a graphical visualization of the probability for the resulting values to fall within a certain range. The higher is the histogram associated with a certain interval, the higher is the probability for the inertial property to assume one among that range of values. Comparing more in details the resulting graphs, it is immediately possible to notice that, differently from the histograms associated with the position of the COM which are almost symmetric with respect to the mean, those associated with the moments of inertia show a different behaviour. Increasing the standard deviation associated with the variation of the input data, in fact, the resulting moments of inertia do not vary symmetrically, but there is a tendency toward higher values. Imposing a maximum acceptable variation of the resulting quantities from the original values, it is possible to evaluate the corresponding maximum acceptable standard deviation. In Table 3.11 are reported the original values, which are associated with the 'open arms' configuration already adopted in Section 3.1.1, and the maximum deviations accepted. To perform a more comprehensive analysis, two different limits are considered: in the first case the maximum variation is set equal to 5 cm for the COM and 0.5 $kg\ m^2$ for the moments of inertia, while in the second case it is set respectively equal to 10 cm and 1 $kg\ m^2$.

	COMx (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** ($kg\ m^2$)	**Iyy** ($kg\ m^2$)	**Izz** ($kg\ m^2$)
original value	-0.21	94.16	4.00	8.14	1.44	9.28
max variation (case 1)	± 5	±5	±5	±0.5	±0.5	±0.5
max variation (case 2)	± 10	± 10	± 10	± 1	± 1	± 1

Table 3.11: Maximum acceptable variations of the resulting position of the COM and inertia tensor for the 33-point model applied on the 'open arms' body pose.

The resulting values of maximum standard deviation acceptable associated with the positions of the 33 body points identified by MediaPipe are shown in Table 3.12.

	COMx (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** (cm)	**Iyy** (cm)	**Izz** (cm)
max standard deviation (case 1)	8.5	5.5	4.5	2.0	4.0	2.0
max standard deviation (case 2)	10	8.5	10	4.0	6.5	3.5

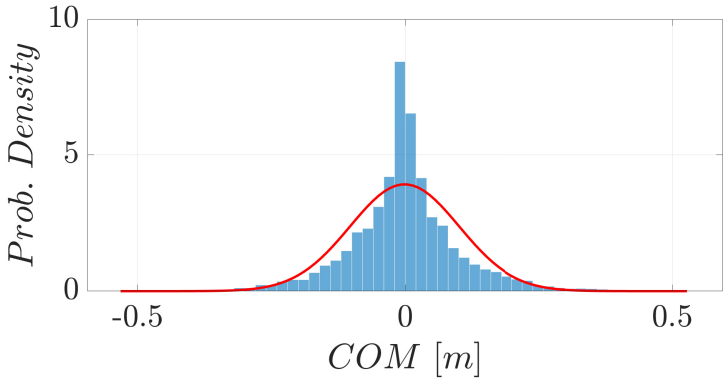
Table 3.12: Resulting maximum acceptable standard deviations associated with the input position data obtained performing the sensitivity analysis for the 33-point model on the 'open arms' body pose.

For what concerns the position of the COM, the most critical condition is associated with the z-axis in the first case and with the y-axis in the second one, since the corresponding standard deviations assume smaller values. It means that it is sufficient a smaller variation of position of the original 33 body points to make the inertia value fall outside the admissible range. Increasing the maximum acceptable variation from 5 cm to 10 cm, as could have been expected, the standard deviation increases along all three directions. Looking at the results associated with the moments of inertia, the outcomes show a different trend: in both cases Ixx and Izz are associated with smaller and close standard deviations, which represent the most critical condition, while those related to Iyy are almost double.
The standard deviation values indicate that there is a probability equal to 99% for the 33 points' positions identified by MediaPipe to fall within a range equal to:

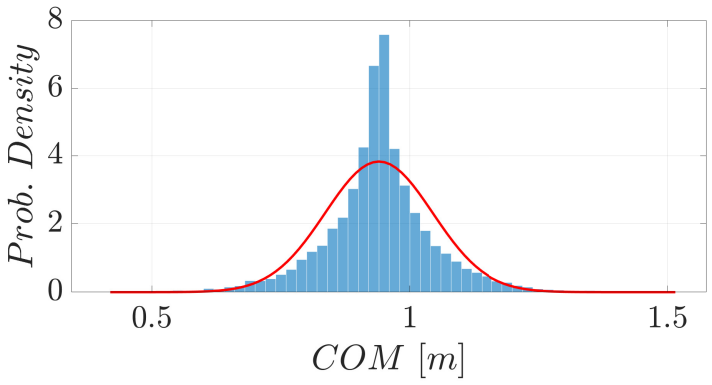
$$x_{varied} = x_{original} \pm (3\ stdev), \tag{3.2}$$

where $x_{original}$ represents the original position of the points identified by the algorithms, thus for a null standard deviation, $stdev$ indicates the actual standard deviation applied to the variability of the input data and x_{varied} is the resulting points' position. This consideration is valuable in understanding the robustness of the adopted model. Taking into account the first case, for example, by accepting a maximum variation of inertial properties equal to 5 cm and 0.5 $kg\ m^2$, it is possible to tolerate a maximum standard deviation of 5.5 cm for the COM and 2 cm for the inertia tensor. This implies that the input points given by MediaPipe can vary from their original positions up to three times these values. Considering the camera adopted to register human motion, despite its specific resolution it is able to detect the position of each pixel with a minimum accuracy in the order of millimeters, which is orders of magnitude smaller if compared to the maximum standard deviations acceptable. Therefore, it is possible to conclude that the geometrical model developed in the present thesis is capable of providing an accurate estimate of the inertial properties of the human body despite the uncertainty associated

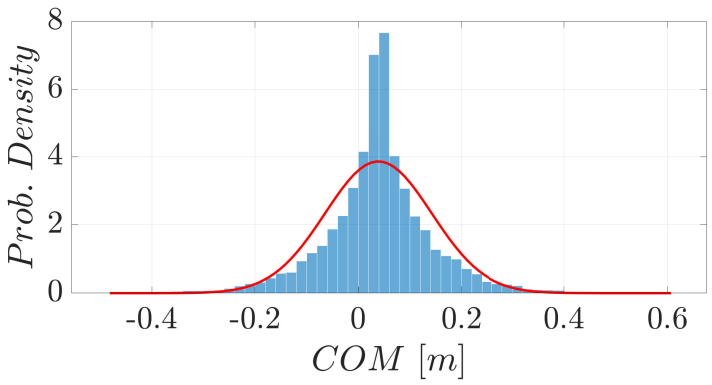
with the camera resolution, given its minimal impact on the achievable resulting values.



(a) COM along x

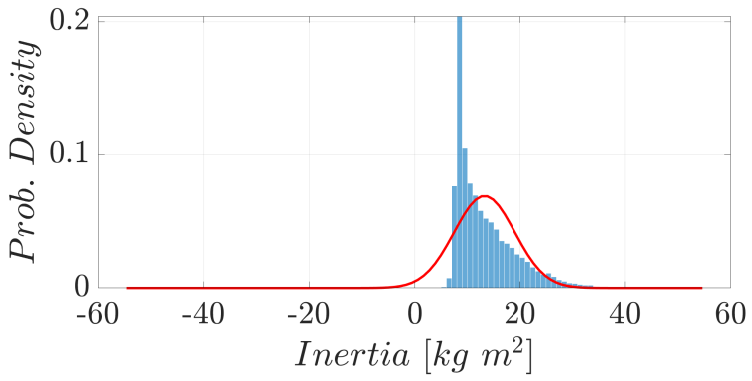


(b) COM along y

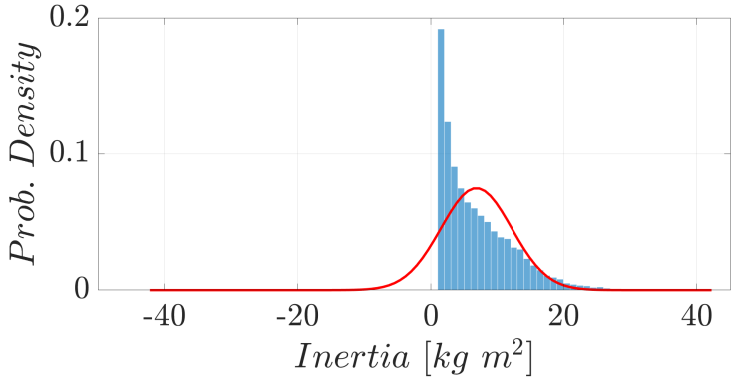


(c) COM along z

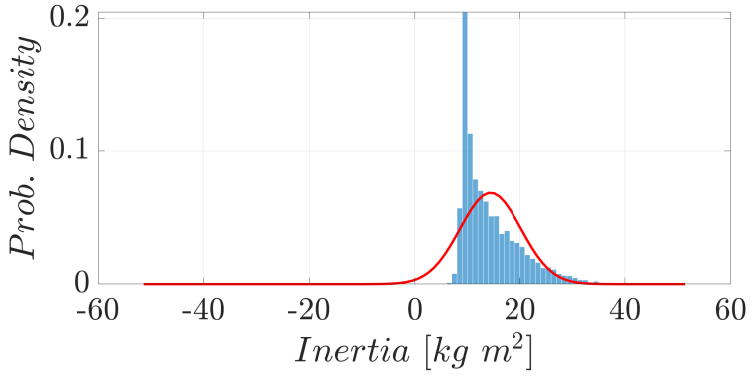
Figure 3.17: Sensitivity analysis conducted on the position of the COM of the entire body for the 33-point model. The outcomes are shown along each direction (x, y and z) in terms of Gaussian distributions and histograms.



(a) Ixx



(b) Iyy



(c) Izz

Figure 3.18: Sensitivity analysis conducted on the moments of inertia about the three axes of the entire body for the 33-point model. The outcomes are presented in terms of Gaussian distribution and histogram of the associated resulting inertia values.

3.2. 17-point models

The aim of this Section is to show the main differences and similarities among the various 17-point models considered, which do not include only those derived using TensorFlow's and YOLOv8's algorithms, but also the one obtainable reducing the number of points identified by MediaPipe from 33 to 17. At first it will be discussed how this reduced model can be obtained and it will be made a comparison with the original 33-point model, considering the same static poses analyzed at the beginning of the previous section ('open arms' and 'closed arms'). Then a comparative analysis between the 33-point model and the reduced one and between all the 17-point models derivable using MediaPipe, TensorFlow and YOLOv8 will be carried out, with the aim of underlying the main advantages and drawbacks associated with the simplification of the skeleton model adopted to estimate the human body inertial properties in real-time. Different frequencies, equal to 30 and 90 fps, will be considered.

3.2.1. Reduction of MediaPipe's model from 33 to 17 points

Starting from the 33 body points identified by MediaPipe, it is possible to reconstruct a 17-point model characterized by the same body landmarks as those identified by TensorFlow and YOLOv8. Once the unnecessary points (hands, feet, mouth and outer and inner corner of the eyes) have been deleted, the procedure described in Section 2.4.2 is applied for the estimate of the human body inertial properties. The main difference between the original 33-point model and the reduced one lies in the absence of 'hands' and 'feet' body parts and in the different procedure adopted for the identification of the dimension of the head. For the 17-point model, in fact, fingers, thumbs, heels and indexes are not identified by the algorithm, which means that the inertial contributions given by hands and feet are not included in the calculations. For what concerns the 'head and neck' body part, which is discretized as an ellipsoid, the different mathematical passages followed for the definition of its major semi-axis lead to dissimilar segmental inertia tensors.
To perform a comparative analysis between the two models identifiable by MediaPipe, the 'open arms' and 'closed arms' body poses represented in Figure 3.19 and 3.20 are considered.

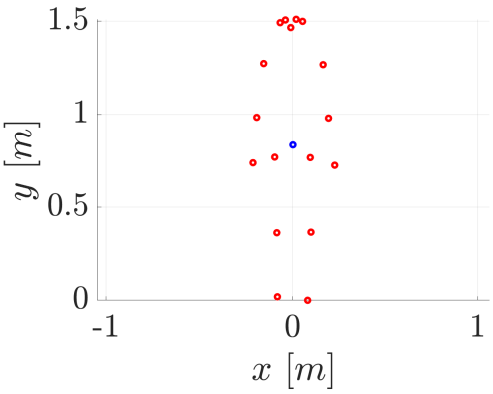


Figure 3.19: Position of the 17 points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for the 'closed arms' configuration.

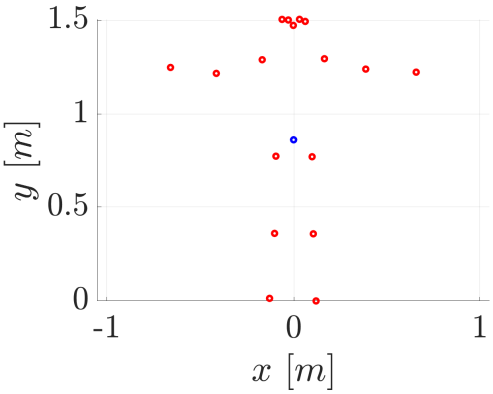


Figure 3.20: Position of the 17 points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for the 'open arms' configuration.

Taking into account the resulting segmental inertial quantities, which are presented in Table 3.13, it is possible to notice that the positions of the segmental COMs, which are expressed as percentages of each segment's length, are exactly the same as those obtained for the original 33-point model. The shape associated with each body part, in fact, has not been changed. For what concerns the inertia tensors, the contributions given by torso, upper arms and upper legs are invariant, while the matrices associated with lower arms and legs are slightly bigger. This is caused by the fact that eliminating hands and feet body parts, the mass percentages that were attributed to such extremities have been re-associated respectively to the lower part of arms and legs.
The inertia tensor characterizing the head is different from the one obtained for the original

Body segment	**Center of mass** (%)	**Ixx** ($kg\ m^2 \cdot 10^3$)	**Iyy** ($kg\ m^2 \cdot 10^3$)	**Izz** ($kg\ m^2 \cdot 10^3$)
Head and neck	50	19.10	9.62	19.10
Torso	50	594.81	215.63	696.29
Upper arm	48.55	8.51	0.99	8.51
Lower arm	42.32	5.32	0.53	5.32
Thigh	46.05	113.57	19.06	113.57
Shank	41.64	30.10	3.82	30.10

Table 3.13: Position of the COM and principal moments of inertia expressed in the local reference frame of each body segment for the 17-point model.

33-point model, which was:

$$\mathbf{I}_{head,33-point} = \begin{bmatrix} 17.9 & 0 & 0 \\ 0 & 9.62 & 0 \\ 0 & 0 & 17.9 \end{bmatrix} \text{ kg m}^2 \cdot 10^3.$$

The major difference is in the order of thousandths of $kg\ m^2$ and it is obtained for the moments of inertia about x and z axes, while those referenced to the y-axis are coincident. Ultimately, the contributions given by hands and feet are null. When examining the values derived from the original model, the segmental inertia tensors associated with these body extremities contributed less than 10% to the overall inertia matrix of the entire body. Hence, in order to obtain an even more simplified model it seems reasonable to exclude these segments, since they can be safely eliminated without excessively influencing the resulting inertia tensor.

In Table 3.14 is presented the comparison between the two models in terms of position of the COM for the 'open arms' and 'closed arms' configurations. For both poses the position of the COM along direction x is coincident for the two models, the position along z is characterized by a small difference (in the order of fractions of millimeter), while that along y is the most critical one. Along the y-axis, in fact, it is achieved a difference equal to 7.93 cm for the 'closed arms' configuration and 8.38 cm for the 'open arms' one.

Model	COMx (cm)	COMy (cm)	COMz (cm)
17-point model CLOSED ARMS	0.25	83.47	4.01
33-point model CLOSED ARMS	0.25	91.40	3.92
difference CLOSED ARMS	0.00	7.93	0.09
17-point model OPEN ARMS	-0.21	85.78	4.03
33-point model OPEN ARMS	-0.21	94.16	4.00
difference OPEN ARMS	0.00	8.38	0.03

Table 3.14: Comparison of the position of the COM between the 17-point and the 33-point models identified by MediaPipe for 'open arms' and 'closed arms' body poses.

A similar analysis can be carried out looking at the resulting inertia tensors shown in Table 3.15. The moments of inertia about the three axes are characterized by the biggest dissimilarities, while the products of inertia for the two models are very close. For both poses the less critical disparity is obtained for the moment of inertia about the y-axis, while the highest difference is achieved for the 'closed arms' pose along x and for the 'open arms' one along z, with value respectively equal to 0.42 $kg\ m^2$ and 0.57 $kg\ m^2$.

Model	Ixx	Iyy	Izz	Ixy	Ixz	Iyz
17-point model CLOSED ARMS	6.96	0.66	7.37	0.04	-0.04	-0.15
33-point model CLOSED ARMS	7.38	0.66	7.79	0.03	-0.04	-0.17
difference CLOSED ARMS	0.42	0.00	0.33	0.01	0.00	0.02
17-point model OPEN ARMS	7.68	1.31	8.71	0.03	-0.09	-0.33
33-point model OPEN ARMS	8.14	1.45	9.28	0.03	-0.10	-0.36
difference OPEN ARMS	0.46	0.14	0.57	000	0.01	0.03

Table 3.15: Comparison of the inertia tensor between the 17-point model and the 33-point one for 'open arms' and 'closed arms' body poses. All the elements of the inertia tensor are expressed in $kg\ m^2$.

To perform a more comprehensive comparison, the same video showing the motion of a single individual has been processed by both the original 33-point MediaPipe algorithm and by the reduced 17-point one, with the purpose of underlying the main differences between the resulting inertial quantities for a moving body. In Figure 3.21 and Figure 3.22 are shown respectively the position of the COM and the moments of inertia obtained

with both models. The COM is expressed with respect to the global reference frame associated with the original 33-point skeleton model defined in Section 2.5: the origins of x and z axes correspond to the mid-point between the hips, while the origin of the y-axis coincides with the lower among the 33 points identified by the algorithm. Even if the global reference frame usually identified for the 17-point model is different, with the purpose of having more easily comparable outcomes it is made coincident with the 33-point model's one. Analysing at first the results in terms of position of the COM, looking at Figure 3.21 it is possible to notice that there is a good agreement.

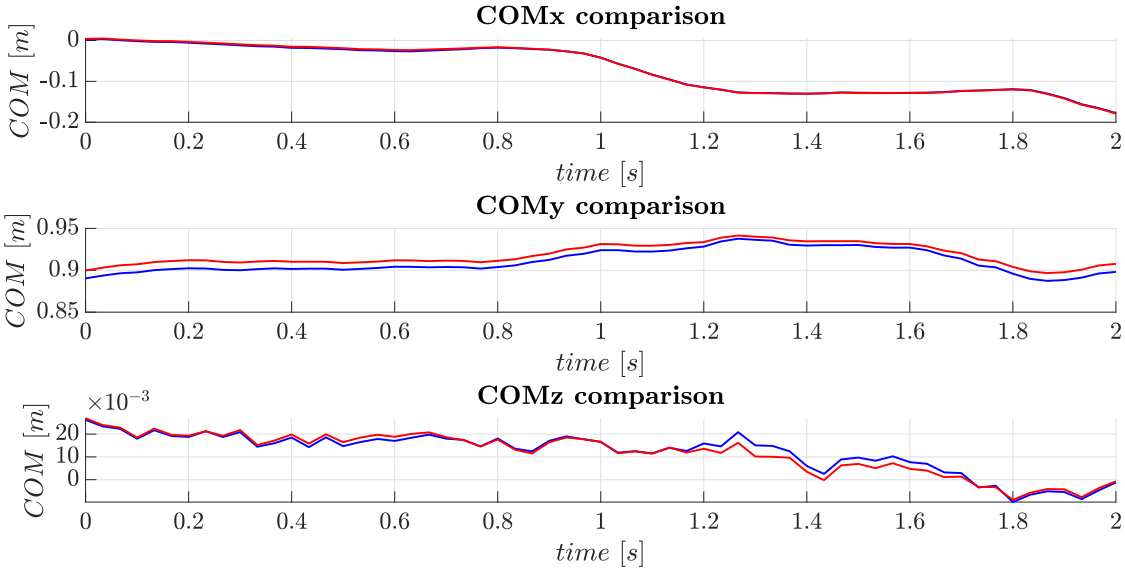


Figure 3.21: Comparison between the position of the COM evaluated using the original skeleton model identified by MediaPipe (blue line) and the reduced 17-point one (red line).

Considering the numerical difference of the positions obtained using the two models at each time instant, it can be calculated mean and standard deviation associated to it. The mean of the difference along x, y and z is respectively equal to 0.12 cm, 0.73 cm and 0.14 cm, while the standard deviation is respectively equal to 0.09 cm, 0.18 cm ad 0.12 cm. The mean along each direction is in the order of millimeters, as well as the standard deviation, meaning that using the 17-point model it is possible to obtain results comparable with those obtained with the original model. The highest difference is obtained along direction y, where the reduced model tends to slightly overestimate the position of the COM. Even in this case, anyway, we are dealing with a disagreement of the order of mil-

limeters, thus it can be concluded that despite it the 17-point model is a valid alternative for the real-time update of the estimate of this inertial quantity.
For what concerns the moments of inertia, which are shown in Figure 3.22, the graphs associated with the two models exhibit similar trends, even if they tend to assume slightly different values with a quasi-constant offset at each time step, in particular for the moments of inertia about x and z axes. The mean of the difference along x, y and z is

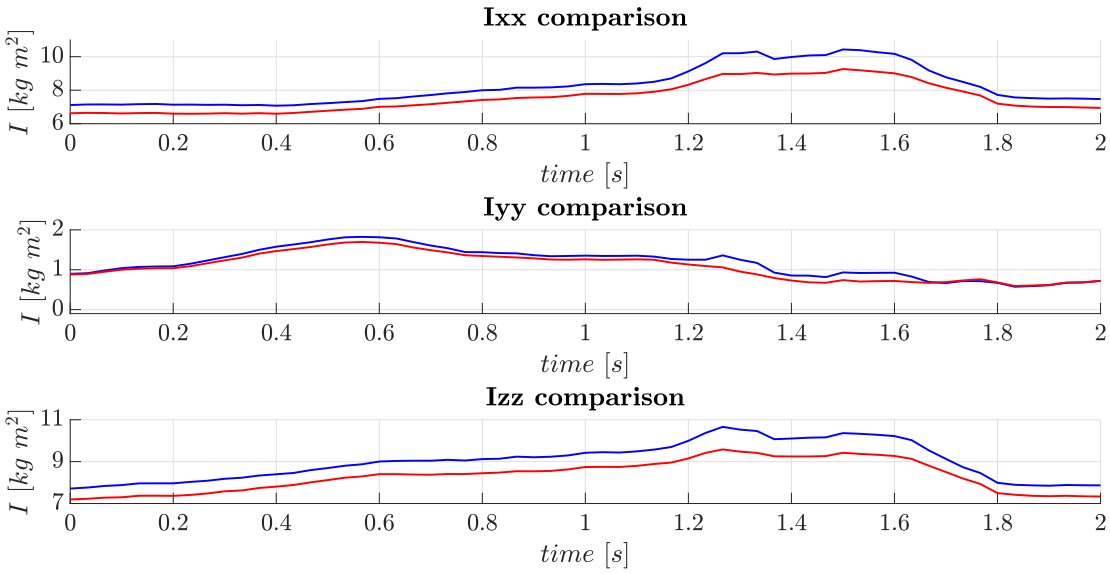


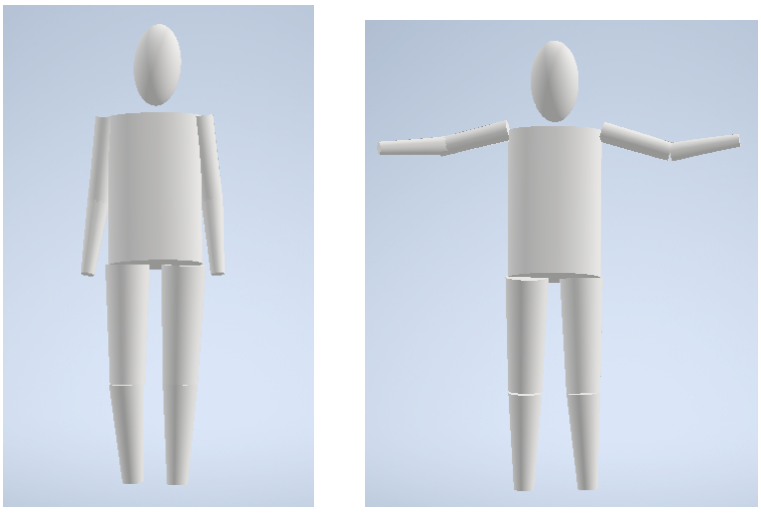
Figure 3.22: Comparison between the inertia moments evaluated using the original skeleton model identified by MediaPipe (blue line) and the reduced one (red line).

respectively equal to 0.67 kgm^2, 0.10 kgm^2 and 0.68 kgm^2, while the standard deviation is respectively equal to 0.25 kgm^2, 0.07 kgm^2 and 0.16 kgm^2. The highest difference, which is in the order of tenths of kgm^2, is associated with x and z axes. In these cases, in fact, looking also at the plots it is immediately possible to notice that the 17-point model tends to underestimate the moments of inertia. This is caused by the absence of hands' and feet's body parts and, consequently, by the fact that their associated inertia tensors are null, not being able to contribute positively to the overall human body inertia matrix. Besides, this difference could be partially related to the different positioning of the COM for the two models, with respect to which the inertia tensor is calculated. The values of the moment of inertia about the y-axis, instead, are closer; also the standard deviation of the difference is smaller, indicating a lower dispersion of the resulting values around the mean. The higher divergence is achieved between 1.2 s and 1.7 s, where peaks associated with the 33-point model are rounded by the 17-point one. This difference could be caused by the difficulty in the correct estimation of hands' and feet's positioning required by the

33-point model. Fingers, toes and heels, in fact, are points that are more likely to fall outside the field of view of the camera, causing an overestimation of the inertia tensor. From this point of view the adoption of the 17-point model may be advantageous for the mitigation of such undesirable effects.
In conclusion, it is possible to announce that the 17-point model represents a valid alternative for the update in real-time of the human body inertial properties. Even if it does not include hands and feet body parts, the resulting values are very close and are characterized by the same trend as those calculated using the 33-point model. Besides it provides additional advantages: not considering body extremities, the risk of erroneously estimating their position does not exist and, finally, identifying 17 rather than 33 points the computational cost associated with each update of the estimation is lower.

CAD comparison

Following the same procedure as the one adopted in Section 3.1.1, the 'closed arms' and 'open arms' body poses associated with the 17-point model have been reproduced in Inventor. Differently from the 33-point configuration, looking at Figure 3.23 it is immediately possible to notice the absence of hands and feet, which in fact do not contribute to the evaluation of the inertial properties of the human body.



(a) Closed arms (b) Open arms

Figure 3.23: Reproduction in Inventor of the human pose for 'closed arms' and 'open arms' configurations considering the 17-point model.

The inertial values obtained with Matlab and Inventor for these static poses are presented

in the following tables in order to highlight possible differences or similarities, also making reference to the original 33-point model obtained through MediaPipe. Considering at first the 'closed arms' body pose, whose associated values are presented in Table 3.16, looking at the last row it is possible to notice that for the COM the maximum difference is obtained for the position along z, while the values of the principal moments of inertia are almost coincident. These results are consistent, in terms of order of magnitude of the difference, with those obtained for the 33-point model.

Closed arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** ($kg\ m^2$)	**Iyy** ($kg\ m^2$)	**Izz** ($kg\ m^2$)
Matlab code	0.25	83.47	4.01	6.95	0.65	7.37
Inventor	0.24	83.45	3.90	6.96	0.65	7.37
Difference	0.01	0.02	0.11	0.01	0.00	0.00

Table 3.16: Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'closed arms' configuration of the 17-point model.

Analyzing, finally, Table 3.17 in which the values associated with the 'open arms' pose are reported, the maximum variation equal to 0.14 cm is achieved for the position of the COM along z. For what concerns the moments of inertia, the maximum difference is kept smaller than 0.1 $kg\ m^2$, as occurred for all the previously analyzed cases.

Open arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** ($kg\ m^2$)	**Iyy** ($kg\ m^2$)	**Izz** ($kg\ m^2$)
Matlab code	-0.21	85.78	4.03	7.68	1.30	8.73
Inventor	-0.21	85.85	4.17	7.66	1.29	8.69
Difference	0.00	0.08	-0.14	0.02	0.01	0.04

Table 3.17: Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'open arms' configuration of the 17-point model.

In conclusion, the comparison between the results obtained via Matlab and Inventor shows a good agreement: for both poses the maximum percentage deviation is equal to 3.3% for the position of the COM and 0.5% for the principal moments of inertia. These quantities may be associated with numerical errors caused by numerical approximations during the execution of the code or with the not exact positioning of the body parts in the CAD

software. The outcomes are also in agreement with those obtained for the 33-point model in Section 3.1.1, whose differences were characterized by the same order of magnitude. Consequently, since such a good correspondence among Matlab and Inventor is achieved for both models, it is possible to conclude that the mathematical passages followed for the estimation of the inertial properties are validated.

Sensitivity analysis

The sensitivity analysis for the 17-point reduced model has been conducted following the procedure described in Section 2.8. The outcomes associated with the variability of position of the COM are shown in Figure 3.24, while those concerning the moments of inertia about the three axes are presented in Figure 3.25. The trends of the Gaussian distributions as well as those of the histograms are aligned with the corresponding results obtained for the 33-point model: for the position of the COM it is obtained a distribution of data which is quasi-symmetric with respect to the mean, while for the moments of inertia the tendency is toward higher values, as can be noticed looking at the histograms in Figure 3.25. Imposing maximum acceptable deviations from the original inertial values equal to those utilized in Section 3.1.2, which are reported for the 'open arms' pose in Table 3.18, the maximum standard deviations acceptable shown in Table 3.19 are obtained.

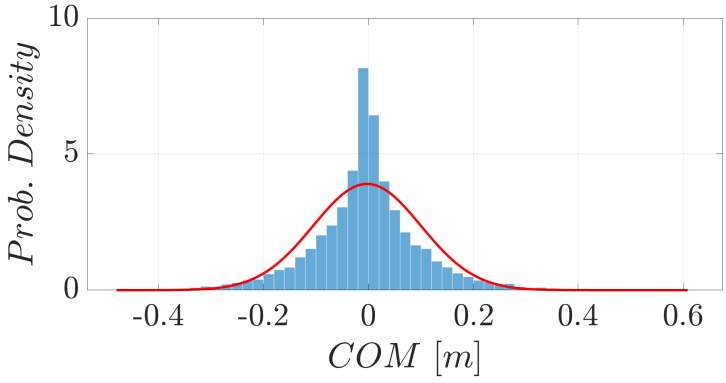
Open arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** $(kg\ m^2)$	**Iyy** $(kg\ m^2)$	**Izz** $(kg\ m^2)$
original value	-0.21	85.78	4.03	7.68	1.31	8.71
max variation (case 1)	± 5	±5	±5	±0.5	±0.5	±0.5
max variation (case 2)	± 10	± 10	± 10	± 1	± 1	± 1

Table 3.18: Maximum acceptable variations of the resulting position of the COM and inertia tensor for the 17-point model applied on the 'open arms' body pose.

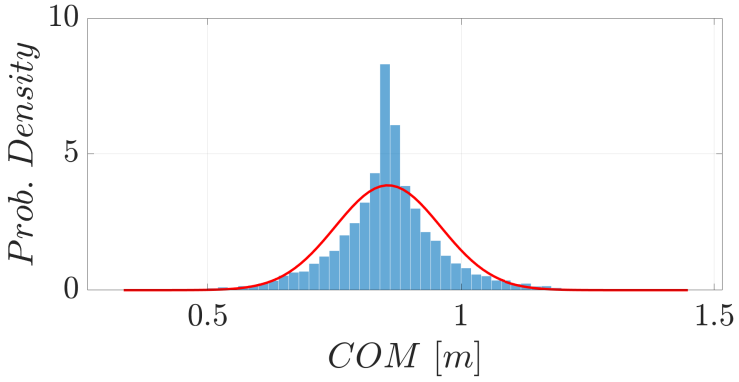
Open arms	**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** (cm)	**Iyy** (cm)	**Izz** (cm)
max standard deviation (case 1)	5.5	5.0	5.0	2.5	4.0	2.5
max standard deviation (case 2)	10.5	10.0	10.0	4.5	7.0	4.0

Table 3.19: Resulting maximum acceptable standard deviations associated with the input position data obtained performing the sensitivity analysis for the 17-point model on the 'open arms' body pose.

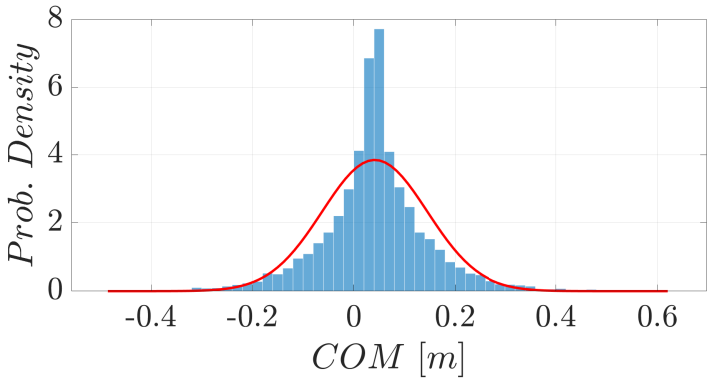
For what concerns the COM, the maximum standard deviation associated with its position along all three directions is similar, there is not such a big discrepancy as the one characterizing the 33-point model in Section 3.1.2. In the present case, in fact, there is not a direction along which the maximum standard deviation acceptable is significantly smaller. Looking at the results associated with the inertia tensor, instead, the behaviour is different: the most critical condition is attributed to Ixx and Izz for both cases, with values close to those resulting from the sensitivity analysis conducted on the 33-point model.
The worst condition differs for the two models: for the COM the smaller maximum standard deviation is equal to 5 *cm* for the first case and 10 *cm* for second case considering the 17-point model, while it is equal to 5.5 *cm* and 8.5 *cm* for the 33-point one. Thus, in both cases, increasing the maximum acceptable variation, the maximum acceptable standard deviation increases too, but the trend is not the same: in the first case the situation is worse for the 17-point model, while in the second one it is worse for the 33-point one. Considering the results associated with the inertia tensor, instead, the behaviour of the two models is more similar, since in both cases the worst condition is associated with the moments of inertia about x and z axes and the values resulting from the sensitivity analysis are closer.
Taking into account, as already done for the original model, the resolution of the camera adopted to register body motions, the uncertainty associated with it has a minimal impact on the resulting values (in the order of *mm* for the positioning of the COM and of hundredths of $kg\ m^2$ for the inertia tensor), given the capability of the camera to detect motions of millimeters (z axis) or fractions of millimeters (x and y axes).



(a) COM along x

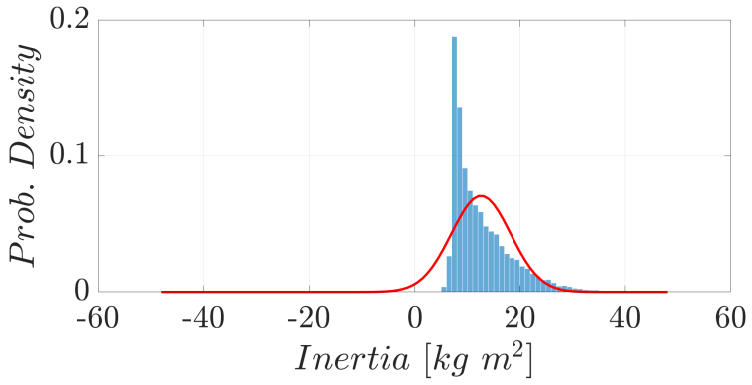


(b) COM along y

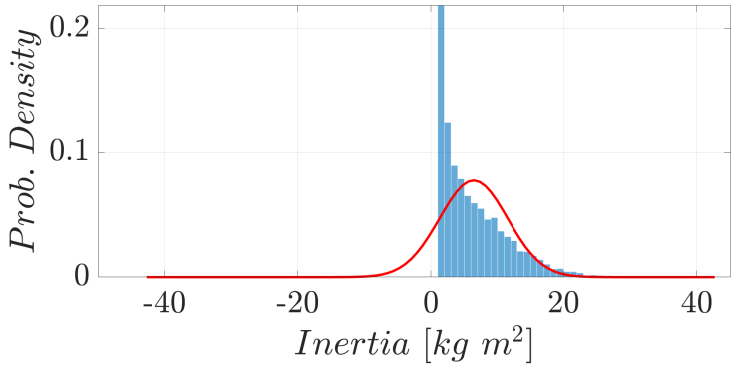


(c) COM along z

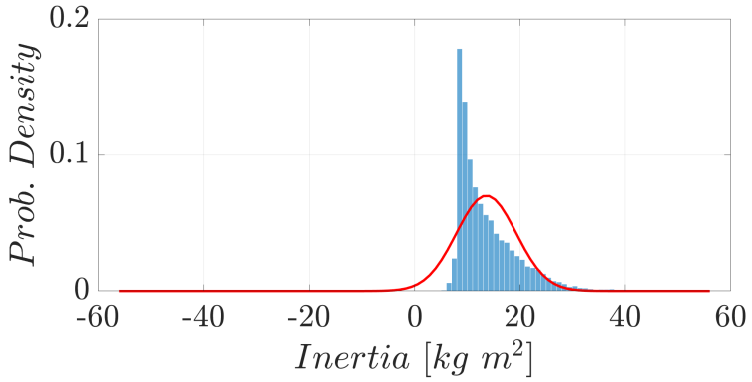
Figure 3.24: Sensitivity analysis conducted on the position of the COM of the entire body for the reduced 17-point model. The outcomes are shown along each direction (x, y and z) in terms of Gaussian distributions and histograms.



(a) Ixx



(b) Iyy



(c) Izz

Figure 3.25: Sensitivity analysis conducted on the moments of inertia about the three axes of the entire body for the reduced 17-point model. The outcomes are presented in terms of Gaussian distribution and histogram of the associated resulting inertia values.

3.2.2. Comparison with TensorFlow and YOLOv8

Adopting the procedure described in Section 2.4.2, it is possible to calculate human body's inertial properties starting from the body points' positions made available by TensorFlow's and YOLOv8's algorithms. In order to highlight the main differences among all the 17-point models considered, which include also the reduced MediaPipe's model analyzed in the previous Section, the same video showing the motion of a human subject is processed. Considering the same sequence of images, in fact, these 5 deep-learning-based algorithms allow for the estimate of 5 different skeleton models, leading finally to unalike inertial properties. In Figure 3.26 is shown the comparison among the results obtained with TensorFlow (MoveNet) and with the reduced MediaPipe's model in terms of position of the COM, considering a frequency equal to 90 fps. The numerical difference at each time instant is shown in Figure 3.27: along x and z directions the distribution of the difference is quasi-symmetric and the maximum deviation from the values obtained using MediaPipe's algorithm is lower than 2 cm. Along y, instead, the difference tends toward positive values, meaning that the position of the COM estimated using MoveNet is lower with respect to that estimated using MediaPipe's reduced algorithm.

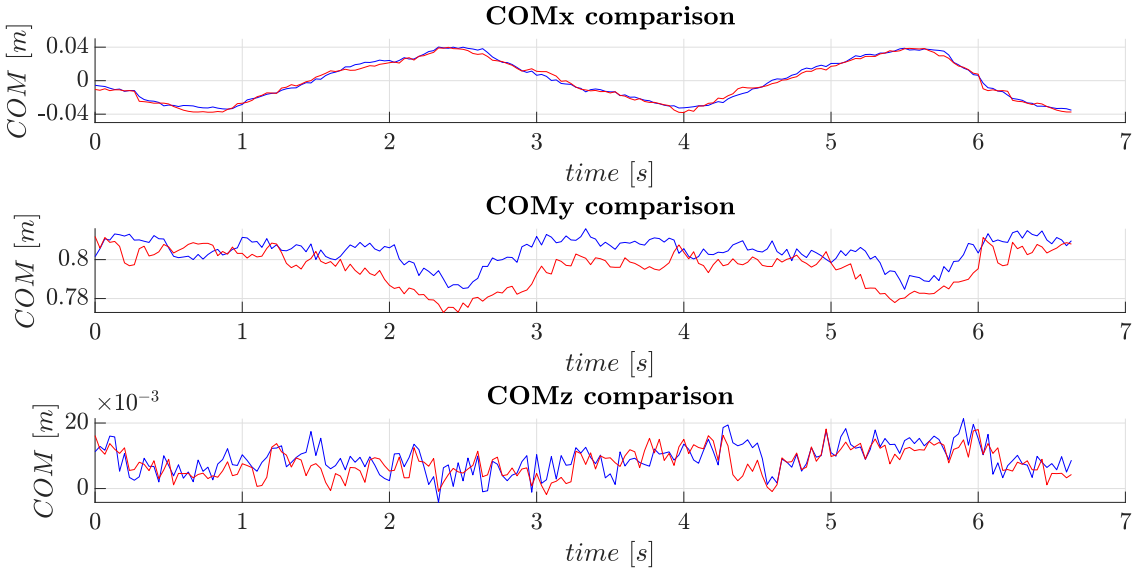


Figure 3.26: Comparison between the variation of position of the COM in time evaluated using the reduced skeleton model identified by MediaPipe (blue line) and the one identified by MoveNet (red line) for a frequency equal to 90 fps.

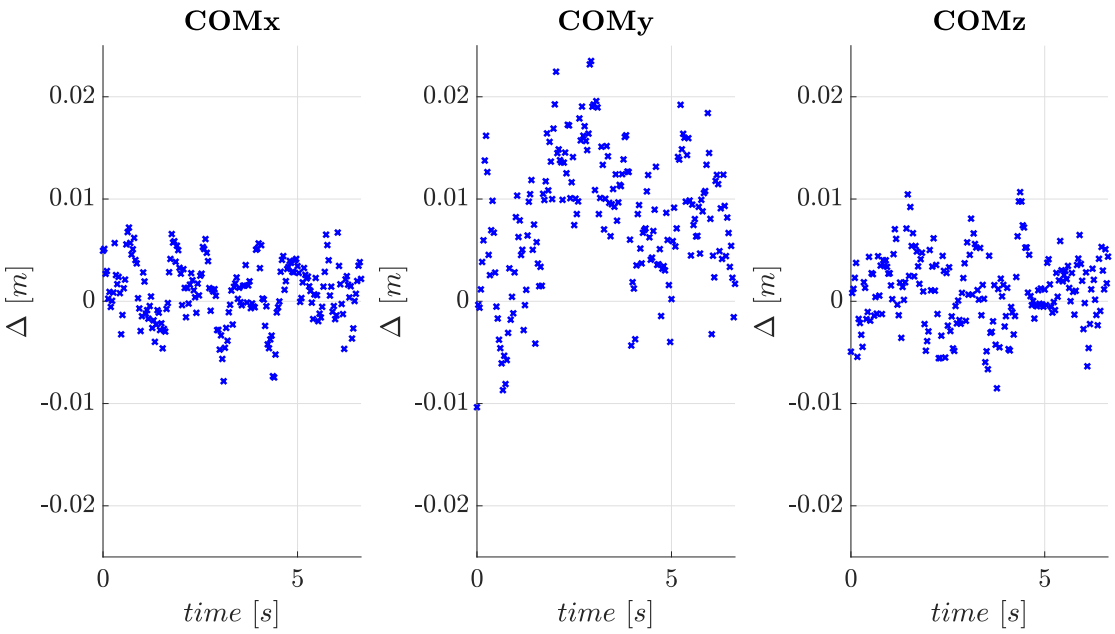


Figure 3.27: Difference between the position of the COM evaluated in time using the reduced skeleton model identified by MediaPipe and the one identified by MoveNet for a frequency equal to 90 fps.

Taking into account the values of difference associated with the position of the COM calculated using the 2 models at each time instant, it is possible to calculate mean and standard deviation: the mean of the difference along x, y and z is respectively equal to 1.06 cm, 1.81 cm and 0.46 cm, while the standard deviation is respectively equal to 0.38 cm, 1.17 cm and 0.37 cm. These values confirm that the difference along the y-axis is the most critical distinction between these two models.

Another comparison is done considering the moments of inertia, whose trends are shown in Figure 3.28. Looking at their numerical difference, which is presented in Figure 3.29, it is possible to notice that the behaviour is similar to that associated with the position of the COM: only for Iyy the difference is non-symmetric, but it tends toward positive values. This is reasonable, since the moments of inertia are related to the position of the COM of the entire human body and a deviation of the COM along y has a bigger influence on the moment of inertia about the y-axis than is has along x and z directions. The mean of the difference along x, y and z is respectively equal to 0.01 kgm^2, 0.09 kgm^2, and 0.01 kgm^2, while the standard deviation is respectively equal to 0.09 kgm^2, 0.05 kgm^2 and 0.09 kgm^2. Looking at these values it is possible to conclude that, since the mean associated with the y-axis is bigger while its standard deviation assumes the smaller value,

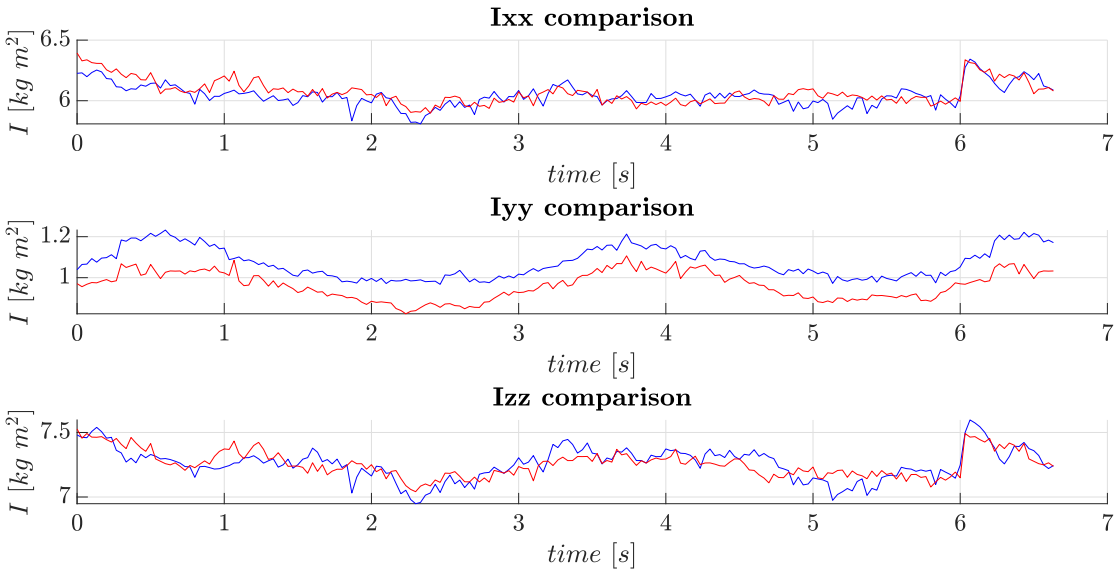


Figure 3.28: Comparison between the variation of the moments of inertia in time using the reduced skeleton model identified by MediaPipe (blue line) and the one identified by MoveNet (red line) for a frequency equal to 90 fps.

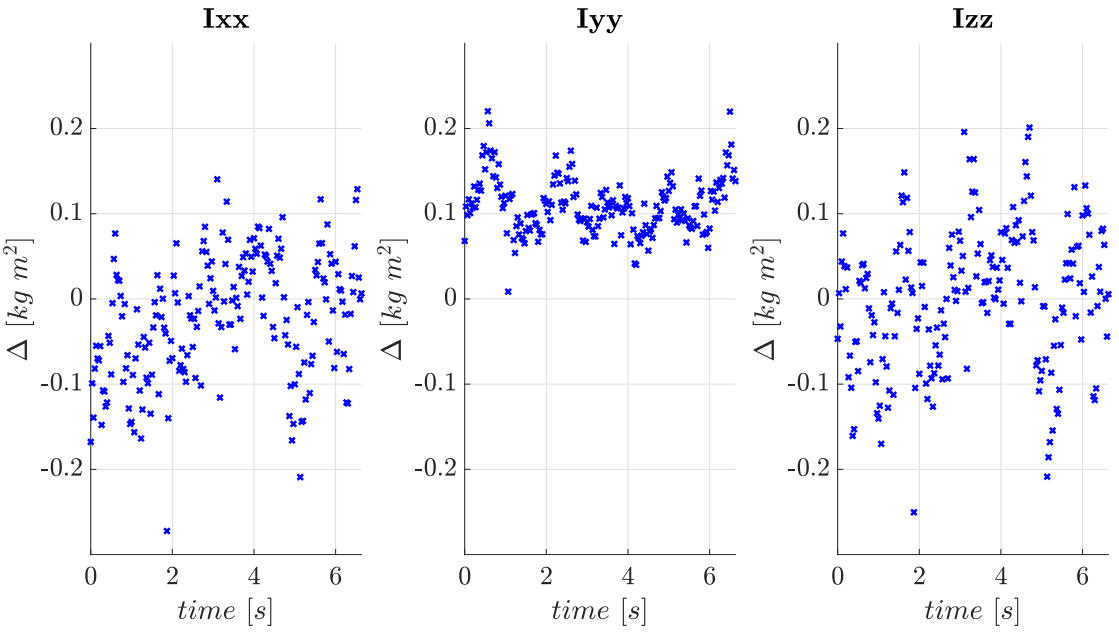


Figure 3.29: Difference between the moments of inertia evaluated in time using the reduced skeleton model identified by MediaPipe and the one identified by MoveNet for a frequency equal to 90 fps.

the moments of inertia about the y-axis obtained for the two models are characterized by

an offset, which is approximately equal to 0.1 $kg\ m^2$, but their trends in time are very similar. Given the small standard deviation, in fact, despite the visible offset, the data show a little dispersion around the mean. Along x and z directions, instead, the values of the mean are smaller while those of the standard deviation are bigger, meaning that the graphs associated with the two models are not characterized by an almost constant offset in time, but they cannot be overlapped given the larger dispersion of data around the mean.

The same analysis is carried out for all the remaining 17-point models, considering always as reference the reduced 17-point model obtained with MediaPipe. The results in terms of mean and standard deviation of the difference are presented in Table 3.20 for a frequency of 30 fps and in Table 3.21 for a frequency of 90 fps. Starting analysing the differences associated with the lower frequency, the position of the COM along y is characterized by the biggest mean and standard deviation for all the models. The bigger differences are obtained using PoseNet and YOLOv8l, while MoveNet's and YOLOv8m's resulting values are closer to MediaPipe's ones.

30 fps		**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** $(kg\ m^2)$	**Iyy** $(kg\ m^2)$	**Izz** $(kg\ m^2)$
MoveNet	mean	1.06	1.81	0.46	0.11	0.09	0.12
	std. dev.	0.38	1.17	0.37	0.09	0.05	0.09
Posenet	mean	0.58	3.12	0.55	0.16	0.09	0.23
	std. dev.	0.46	2.10	0.43	0.12	0.07	0.15
YOLOv8l	mean	0.67	4.76	0.53	0.55	0.10	0.58
	std. dev.	0.39	0.78	0.41	0.11	0.10	0.14
YOLOv8m	mean	0.34	3.36	0.39	0.53	0.09	0.56
	std. dev.	0.23	0.66	0.28	0.11	0.08	0.15

Table 3.20: Comparison between the results obtained with TensorFlow's and YOLOv8's models and the reduced MediaPipe's 17-point model for a frequency of 30 fps. In the table are listed means and standard deviations of the difference between the results obtained with these algorithms, using always as reference the MediaPipe's 17-point model.

For what concerns the difference associated with the moments of inertia, also in this case a similar behaviour is shared by all the models: along x and z both mean and standard deviation are higher, more in detail the bigger deviation from the reference is reached using YOLOv8l. We are dealing with differences in the order of centimeters or fraction of

centimeters for the positioning of the COM and of tenths or hundredths of kgm^2 for the moments of inertia. These values are associated with the fact that each model tends to identify the position of the 17 body points slightly differently, leading to the estimate of unalike inertial properties. Despite this numerical difference, the resulting values exhibit a similar trend: the increase or reduction of inertial properties in time is comparable across various models, as it is possible to notice looking at Figure 3.26 and 3.28.

Switching to a higher frequency equal to 90 fps, which is associated with a smaller resolution of the camera, it is possible to notice in general a better agreement between the 17-point models under analysis and the reference model obtained with MediaPipe. Considering Table 3.21, in fact, mean and standard deviation of the difference for the positioning of the COM are smaller, with the only exception of PoseNet's model, which deviates slightly more from the reference one. For what concerns the moments of inertia, instead, there is a decrease of discrepancy for all the models, included PoseNet's one. This improvement occurs despite the reduction of resolution associated with the depth camera and it may be due to the fact that updating the estimate more frequently allows the system to respond more promptly to changes in the input data. It is available, in fact, a larger amount of recent informations which permits to estimate closer-to-reality inertial properties over time.

90 fps		**COMx** (cm)	**COMy** (cm)	**COMz** (cm)	**Ixx** $(kg\ m^2)$	**Iyy** $(kg\ m^2)$	**Izz** $(kg\ m^2)$
MoveNet	mean	0.25	0.90	0.30	0.06	0.11	0.07
	std. dev.	0.19	0.53	0.22	0.05	0.03	0.05
Posenet	mean	0.67	3.37	0.80	0.18	0.06	0.18
	std. dev.	0.56	1.63	0.71	0.11	0.04	0.11
YOLOv8l	mean	0.35	1.16	0.55	0.25	0.10	0.29
	std. dev.	0.29	0.64	0.38	0.08	0.06	0.11
YOLOv8m	mean	0.42	0.56	0.37	0.16	0.07	0.18
	std. dev.	0.26	0.44	0.29	0.09	0.04	0.11

Table 3.21: Comparison between the results obtained with TensorFlow's and YOLOv8's models and the reduced MediaPipe's 17-point model for a frequency of 90 fps.

In conclusion, the differences among the various 17-point models for both frequencies remain in the order of centimeters or millimeters for the position of the COM and of tenths or hundredths of kgm^2 for the moments of inertia. They are associated with the

fact that each model, even starting from the same image made available by the camera, identify a slightly different skeleton model and, consequently, unalike inertial quantities. It is possible to mitigate such a variation considering higher acquisition frequencies (from 30 to 90 fps), which are associated with closer-to-reality estimations. Starting from the present study, given the small difference associated with each model with respect to the reference MediaPipe's one, it is not possible to identify the best absolute choice among the various alternatives. Anyway, keeping in mind that YOLOv8 models have been developed more recently, being associated with more advanced and modern technologies, they could represent the best choice for the update in real-time of the human body inertial properties.

4 | Conclusions and future developments

The primary objective of the present thesis was to establish a reliable method for the real-time estimation of human body's inertial properties, utilizing data made available by the stereo-depth camera Intel RealSense D415 and by the consequent processing of images performed by a deep-learning-based algorithm. Various algorithms were implemented to define the skeleton model associated with the individual under test. Starting from MediaPipe, which provides the three-dimensional coordinates of 33 key body landmarks, following the procedure described in detail in Chapter 2 it was possible to assess the dimension of each body part constituting the geometrical model and to calculate their inertial contributions. Firstly, the examination conducted on static poses revealed how specific inertial properties are indeed associated with each body pose. These properties enable a detailed analysis of human body motion and imbalance: interpreting the values of the inertia tensor and of the position of the COM, it becomes possible to scrutinize the body's spatial orientation and assess the stability of the assumed posture. The procedure conducted to validate the model and the mathematical passages employed to infer inertial values solely from the points' positions provided by the algorithm, consists of two consecutive steps. At first it was conducted a comparative analysis between the inertial properties' values associated with each body segment obtained using the model and those evaluated using the regression equations defined by Zatsiorsky [26] and De Leva [22]. As evident in the associated section (Section 3.1.1), this first validation yielded satisfactory responses. Despite the unavoidable differences associated with the fact that the geometric model presented in this thesis is a simplification of human body's real dimensions and shapes, while the equations defined in literature take into consideration more detailed anthropometric measurements and studies, the resulting values in terms of segmental mass percentages, positions of the COM and inertia tensors show a good agreement. The second part of the validation consists in the reproduction of the geometrical model in the CAD software Inventor. This was performed for two different poses and the results revealed a maximum error equal to 0.1% for the positioning of the COM and to 1.4%

for the inertia tensor. These percentages indicate a deviation from the original values derived in CAD of respectively less than 1 mm and 0.1 $kg\ m^2$, which may be associated with numerical errors or with the not perfect positioning of the discretized body as an assembly in Inventor. Thus, it was possible to conclude that this discrepancy is not caused by mistakes in the mathematical passages followed, but only by intrinsic errors associated with the reproduction of the model in the CAD software. In this way, not only the values of inertial properties associated with each body segment were confirmed, but it was possible to validate the mathematical calculations defined for the combination of such quantities aimed at defining the inertia tensor of the entire body.
To carry out the inertial properties' real-time update it was fundamental to address some critical aspects. First of all, considering the possibility that the individual might go out of the camera's visibility range during motion, leading to the erroneous detection of body landmarks' positions, it has been necessary to develop a filtering mechanism to make up for the presence of scatter data. This issue could also be attributed to the resolution of the camera. However, as concluded in Section 3.1.2, it was observed that in this second case the deviation of position from the trend is smaller and it does not significantly influence the resulting inertial values. The second important factor to be taken into account concerns the variability of each segment's length over time, arising from the fact that the algorithm was not designed with the purpose of maintaining consistent proportions of the human body, but every time a new image is processed these may vary. Throughout this analysis, it has been demonstrated that hands' and feet's body points are those associated with the largest positioning errors. This is caused by the fact that being at the extremities of the body, these points are more likely to fall outside the camera's visibility range, in particular if the individual under test is performing broad movements. In this case, in fact, as deduced in Chapter 3.1.2, variations in hands' and feet's lengths may reach magnitudes of decimeters, which are clearly not acceptable. Consequently, in order to estimate correctly the dimension of each body segment and its location in space, it is necessary to perform a preliminary analysis considering the subject as still as possible and setting the length of each segment performing an average of the values obtained over multiple acquisitions.

The inaccuracy related to the positioning of hands' and feet's body parts, whose contribution to the overall inertia tensor, furthermore, do not exceed 10%, is the main reason for what the usage of 17-point rather than 33-point models was taken into consideration. Although considering a geometrical model composed of 10 instead of 14 parts may entail a loss of informations, given the possible source of error associated with hands' and feet's

positioning it may be considered advantageous deleting them. By applying the same procedure for the estimation of human body's inertial properties to all the 17-point models, which comprehend both the reduced MediaPipe's model and those obtained using the algorithms provided by TensorFlow and YOLOv8, a comparative analysis was carried out. Focusing at first on the reduced model obtained through MediaPipe, in Section 3.2.1 it was examined the variation in the position of the COM and in the inertia tensor of the entire body, firstly for two static poses and then for a moving body. The resulting values were compared to those originally derived using the 33-point model, showing a disparity between them of the order of millimeters for the positioning of the COM and tenths of $kg\ m^2$ for the inertia tensor. Such difference is attributed to the distinct sizing of the head and the absence of hands and feet, whose inertial contribution in the 33-point model, nonetheless, was significantly low, at least an order of magnitude less than those associated with all other body parts.

It was later performed a comparison among the results obtained with the remaining 17-point algorithms, using always the reduced MediaPipe's model as a reference. The analysis conducted in Chapter 3.2.1 revealed that, for what concerns the estimation of the position of the COM, the biggest difference is always detected along the y-axis. Furthermore, taking into consideration the acquisition frequency, it was possible to notice that the transition from 30 to 90 fps is associated with a reduction in the difference between the values identified by the 17-point models and the MediaPipe's reference one. This is reasonable since, despite the reduction in resolution associated with the increase in acquisition frequency may potentially compromise the accuracy of the measurements, updating the estimation more frequently allows the system to respond more promptly to changes in the input data. In this way a larger amount of recent informations is available, enabling an estimation of inertial parameters closer to real values.

The same consideration can be done regarding the difference between the inertia tensor's values. In this case, more specifically, the observed differences are on the order of tenths of $kg\ m^2$ or even hundredths, for the highest frequency of 90 fps. In contrast to the results obtained for MoveNet and PoseNet (i.e. the algorithms provided by TensorFlow), which show that the variation in inertia is almost symmetric between positive and negative values, YOLOv8's algorithms tend to estimate moments of inertia whose numerical values are almost always smaller than those obtained with the reduced MediaPipe's model. Using the findings from this thesis, it is not possible to definitely identify which among the analyzed 17-point models allows for the estimate of closer-to-reality inertial quantities. Nonetheless, considering that YOLOv8's algorithms have been developed more recently and are presumed to be associated with more modern and advanced technologies, they could represent the best choice for the update in real-time of the human body's

inertial properties. For more details, it would be necessary to conduct a more accurate analysis that takes into consideration the precise identification of skeletal models for all algorithms. This involves precisely understanding how each of the 17 points is located in space starting from the images captured by the camera. By doing so, it could be possible to determine the degree of correspondence with reality for each one of the models utilized.

In conclusion, it is possible to announce that the study conducted in this thesis has yielded positive results. It has been successfully developed a geometrical model that, exploiting a depth camera and a deep-learning-based algorithm, is capable of estimating in real-time the human body's inertial properties with a good degree of accuracy. The strength of the developed method lies in its robustness and versatility, allowing for the calculation of inertia tensors associated with bodies of various sizes and proportions during the performance of the most varied physical activities.
The study was conducted considering both 33-point and 17-point models. The resulting values revealed maximum differences of a few centimeters in the positioning of the COM between the two cases, while for the inertia tensor dissimilarities in the order of tenths or hundredths of $kg\ m^2$ were achieved. These variations are not reputed statistically significant, especially considering that the geometrical model adopted is a simplification of human body's real dimensions and sizes, thus leading to already approximated inertial values. Therefore, the best choice is to opt for one among the 17-point models, as they provide significant advantages: the error associated with body extremities' positioning is mitigated and computational time is reduced, enhancing real-time approximation performance. In order to amplify further more this capability, it is necessary to increment the acquisition frequency for the reduction of time between consecutive estimates' update.
Potential avenues for future researches could involve a discretization of the human body that more thoroughly considers the differences among male and female figures, allowing for a more accurate sizing of each body part. Furthermore, a more detailed analysis of the various 17-point models available could effectively individuate the best-performing model, capable of yielding closer-to-reality inertial values.

Bibliography

[1] R Aissaoui, P Allard, A Junqua, L Frossard, and M Duhaime. Internal work estimation in three-dimensional gait analysis. *Medical and Biological Engineering and Computing*, 34:467–471, 1996.

[2] Chuan Zhi Tay, King Hann Lim, and Jonathan Then Sien Phang. Markerless gait estimation and tracking for postural assessment. *Multimedia Tools and Applications*, 81(9):12777–12794, 2022.

[3] M Ganglbauer, M Ikeda, M Plasch, and A Pichler. Human in the loop online estimation of robotic speed limits for safe human robot collaboration. *Procedia Manufacturing*, 51:88–94, 2020.

[4] Claudia Latella, Marta Lorenzini, Maria Lazzaroni, Francesco Romano, Silvio Traversaro, M Ali Akhras, Daniele Pucci, and Francesco Nori. Towards real-time whole-body human dynamics estimation through probabilistic sensor fusion algorithms: A physical human–robot interaction case study. *Autonomous Robots*, 43:1591–1603, 2019.

[5] Ram Vasudevan, Victor Shia, Yiqi Gao, Ricardo Cervera-Navarro, Ruzena Bajcsy, and Francesco Borrelli. Safe semi-autonomous control with enhanced driver modeling. In *2012 American Control Conference (ACC)*, pages 2896–2903. IEEE, 2012.

[6] Victor A Shia, Yiqi Gao, Ramanarayan Vasudevan, Katherine Driggs Campbell, Theresa Lin, Francesco Borrelli, and Ruzena Bajcsy. Semiautonomous vehicular control using driver modeling. *IEEE Transactions on Intelligent Transportation Systems*, 15(6):2696–2709, 2014.

[7] Tony Flenmark. Power production in future european navy and coast/border guard vessels. 2019.

[8] Vladimir Aleshin, Stanislav Klimenko, Mikhail Manuilov, and Leonid Melnikov. Alpine skiing and snowboarding training system using in-duced virtual environment. *Science and Skiing IV*, 4:137–144, 2009.

[9] Katherine Driggs-Campbell, Guillaume Bellegarda, Victor Shia, S Shankar Sastry, and Ruzena Bajcsy. Experimental design for human-in-the-loop driving simulations. *arXiv preprint arXiv:1401.5039*, 2014.

[10] Anup Doshi, Brendan Morris, and Mohan Trivedi. On-road prediction of driver's intent with multimodal sensory cues. *IEEE Pervasive Computing*, 10(3):22–34, 2011.

[11] Sergio Casas, Ricardo Olanda, and Nilanjan Dey. Motion cueing algorithms: a review: algorithms, evaluation and tuning. *International Journal of Virtual and Augmented Reality (IJVAR)*, 1(1):90–106, 2017.

[12] D Basacik and A Stevens. *Scoping study of driver distraction.* Department for Transport London, 2008.

[13] John D Lee, Kristie L Young, and Michael A Regan. Defining driver distraction. *Driver distraction: Theory, effects, and mitigation*, 13(4):31–40, 2008.

[14] Amardeep Sathyanarayana, Sandhya Nageswaren, Hassan Ghasemzadeh, Roozbeh Jafari, and John HL Hansen. Body sensor networks for driver distraction identification. In *2008 IEEE international conference on vehicular electronics and safety*, pages 120–125. IEEE, 2008.

[15] Luis Miguel Bergasa, Jesús Nuevo, Miguel A Sotelo, Rafael Barea, and María Elena Lopez. Real-time system for monitoring driver vigilance. *IEEE Transactions on intelligent transportation systems*, 7(1):63–77, 2006.

[16] Wilfred Taylor Dempster et al. *Space requirements of the seated operator: geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs*, volume 159. Wright Air Development Center Wright-Patterson Air Force Base, Ohio, 1955.

[17] Charles E Clauser, John T Mc Conville, and John W Young. Weight, volume, and center of mass of segments of the human body. Technical report, 1969.

[18] RF Chandler, Charles E Clauser, John T McConville, HM Reynolds, John W Young, et al. *Investigation of inertial properties of the human body*, volume 53. Aerospace Medical Research Laboratory Wright-Patterson Air Force Base, OH, USA, 1975.

[19] Michael Mungiole and Philip E Martin. Estimating segment inertial properties: comparison of magnetic resonance imaging with existing methods. *Journal of biomechanics*, 23(10):1039–1046, 1990.

[20] Chen Wei and Robert K Jensen. The application of segment axial density profiles to a human body inertia model. *Journal of biomechanics*, 28(1):103–108, 1995.

[21] Wlodzimierz S Erdmann. Geometric and inertial data of the trunk in adult males. *Journal of biomechanics*, 30(7):679–688, 1997.

[22] Paolo De Leva. Adjustments to zatsiorsky-seluyanov's segment inertia parameters. *Journal of biomechanics*, 29(9):1223–1230, 1996.

[23] Vladimir M Zatsiorsky. Measuring body segment parameters: X-ray versus gamma scanning. *Journal of biomechanics*, 9(36):1405–1406, 2003.

[24] Jason Wicke and Genevieve A Dumas. Estimating segment inertial parameters using fan-beam dxa. *Journal of applied biomechanics*, 24(2):180–184, 2008.

[25] Jason Wicke, Genevieve A Dumas, and Patrick A Costigan. A comparison between a new model and current models for estimating trunk segment inertial parameters. *Journal of biomechanics*, 42(1):55–60, 2009.

[26] V Zatsiorsky. Estimation of the mass and inertia characteristics of the human body by means of the best predictive regression equations. *Biomechanics; proceedings*, pages 233–239, 1985.

[27] Ernest P Hanavan et al. *A mathematical model of the human body*, volume 32. Aerospace Medical Research Laboratories, Aerospace Medical Division, Air . . . , 1964.

[28] Maurice R Yeadon. The simulation of aerial movement—ii. a mathematical inertia model of the human body. *Journal of biomechanics*, 23(1):67–74, 1990.

[29] Herbert Hatze. A mathematical model for the computational determination of parameter values of anthropomorphic segments. *Journal of biomechanics*, 13(10):833–843, 1980.

[30] William SP Robertsona. A modern take on the theoretical modelling of inertial properties of a human body for biomechanical simulations. In *20th Int. Congress on Modelling and Simulation (ICMS), Adelaide, Australia*, 2013.

[31] Vladimir Konstantinov Kotev, Gergana S Nikolova, and Daniel M Dantchev. Determination of mass-inertial characteristics of the human body in basic body positions: computer and mathematical modelling. In *EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC), Tampere, Finland, June 2017*, pages 579–582. Springer, 2018.

[32] Gergana Nikolova, Vladimir Kotev, Daniel Dantchev, and Mihail Tsveov. Study of mass-inertial characteristics of female human body by walking. In *AIP Conference Proceedings*, volume 2239. AIP Publishing, 2020.

[33] Manisha Jagadale, KN Agrawal, CR Mehta, RR Potdar, and Nandni Thakur. Estimation and validation of body segment parameters using 3d geometric model of human body for female workers of central india. *Agricultural Research*, 11(4):768–780, 2022.

[34] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. Deep 3d human pose estimation: A review. *Computer Vision and Image Understanding*, 210:103225, 2021.

[35] Alexandra Pfister, Alexandre M West, Shaw Bronner, and Jack Adam Noah. Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology*, 38(5):274–280, 2014.

[36] Claudia Latella. Human whole-body dynamics estimation for enhancing physical human-robot interaction. *arXiv preprint arXiv:1912.01136*, 2019.

[37] Timo Von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *Proceedings of the European conference on computer vision (ECCV)*, pages 601–617, 2018.

[38] Thomas Robert, Paul Leborgne, Mahdi Abid, Vincent Bonnet, Gentiane Venture, and Raphaël Dumas. Whole body segment inertia parameters estimation from movement and ground reaction forces: a feasibility study. *Computer methods in BiomeChaniCs and BiomediCal engineering*, 20(sup1):S175–S176, 2017.

[39] Hang Thi Phuong Nguyen, Yeongju Woo, Ngoc Nguyen Huynh, and Hieyong Jeong. Scoring of human body-balance ability on wobble board based on the geometric solution. *Applied Sciences*, 12(12):5967, 2022.

[40] Ci-Jyun Liang, Kurt M Lundeen, Wes McGee, Carol C Menassa, SangHyun Lee, and Vineet R Kamat. A vision-based marker-less pose estimation system for articulated construction robots. *Automation in Construction*, 104:80–94, 2019.

[41] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013.

[42] Leonid Sigal, Alexandru O Balan, and Michael J Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision*, 87(1-2):4–27, 2010.

[43] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014.

[44] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In *2017 international conference on 3D vision (3DV)*, pages 506–516. IEEE, 2017.

[45] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[46] CD Novo, R Boss, P Kyberd, et al. Testing the microsoft kinect skeletal tracking accuracy under varying external factors. *MOJ App Bio Biomech*, 6(1):7–11, 2022.

[47] Abhishek Kar et al. Skeletal tracking using microsoft kinect. *Methodology*, 1(1):11, 2010.

[48] Sauptik Dhar, Junyao Guo, Jiayi Liu, Samarth Tripathi, Unmesh Kurup, and Mohak Shah. A survey of on-device machine learning: An algorithms and learning theory perspective. *ACM Transactions on Internet of Things*, 2(3):1–49, 2021.

[49] Rui Rodrigues, Rui Neves Madeira, Nuno Correia, Carla Fernandes, and Sara Ribeiro. Multimodal web based video annotator with real-time human pose estimation. In *Intelligent Data Engineering and Automated Learning–IDEAL 2019: 20th International Conference, Manchester, UK, November 14–16, 2019, Proceedings, Part II 20*, pages 23–30. Springer, 2019.

[50] A Pourramezan Fard, H Abdollahi, and M Mahoor. Deep active shape model for face alignment and pose estimation. *arXiv e-prints*, 2021.

[51] Ross Girshick, Forrest Iandola, Trevor Darrell, and Jitendra Malik. Deformable part models are convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 437–446, 2015.

[52] Amir Zadeh, Yao Chong Lim, Tadas Baltrusaitis, and Louis-Philippe Morency. Convolutional experts constrained local model for 3d facial landmark detection. In *Pro-*

ceedings of the IEEE International Conference on Computer Vision Workshops, pages 2519–2528, 2017.

[53] Shuxian Wang, Xiaoxun Zhang, Fang Ma, Jiaming Li, and Yuanyou Huang. Single-stage pose estimation and joint angle extraction method for moving human body. *Electronics*, 12(22):4644, 2023.

[54] Erik Murphy-Chutorian and Mohan Manubhai Trivedi. Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 31(4):607–626, 2008.

[55] Andrew Choutka. Using machine learning to model the human run. 2023.

[56] Gergana Stefanova Nikolova and Yuli Emilov Toshev. Estimation of male and female body segment parameters of the bulgarian population using a 16-segmental mathematical model. *Journal of biomechanics*, 40(16):3700–3707, 2007.

[57] Gongbing Shan and Christiane Bohn. Anthropometrical data and coefficients of regression related to gender and race. *Applied ergonomics*, 34(4):327–337, 2003.

[58] J Gavin Reid and Robert K Jensen. Human body segment inertia parameters: a survey and status report. *Exercise and sport sciences reviews*, 18(1):225–242, 1990.

[59] Wonsup Lee, Baekhee Lee, Xiaopeng Yang, Hayoung Jung, Ilgeun Bok, Chulwoo Kim, Ochae Kwon, and Heecheon You. A 3d anthropometric sizing analysis system based on north american caesar 3d scan data for design of head wearable products. *Computers & Industrial Engineering*, 117:121–130, 2018.

[60] Raphaël Dumas, Laurence Cheze, and J-P Verriest. Adjustments to mcconville et al. and young et al. body segment inertial parameters. *Journal of biomechanics*, 40(3):543–553, 2007.

[61] Gergana Nikolova, Vladimir Kotev, and Daniel Marinov Dantchev. Computer and mathematical modelling of the female human body: Determination of mass-inertial characteristics in basic body positions. In *SIMULTECH*, pages 416–421, 2017.

A | Appendix A

Plot of the variability of each segment's dimension over time.

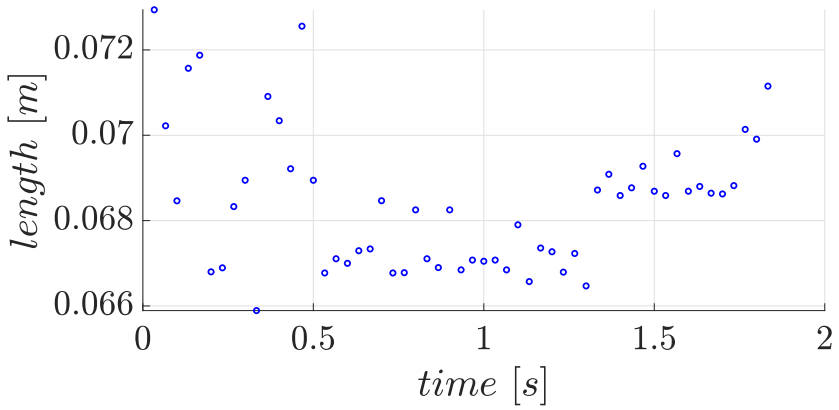


Figure A.1: Length of the minor radius of the head.

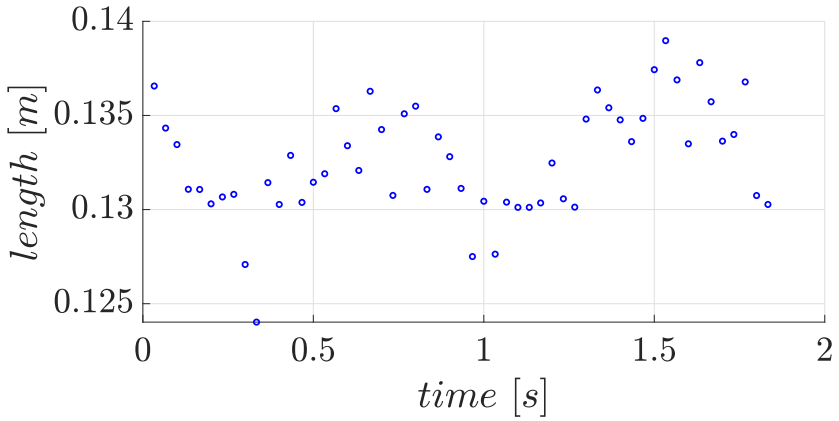


Figure A.2: Length of the major radius of the head.

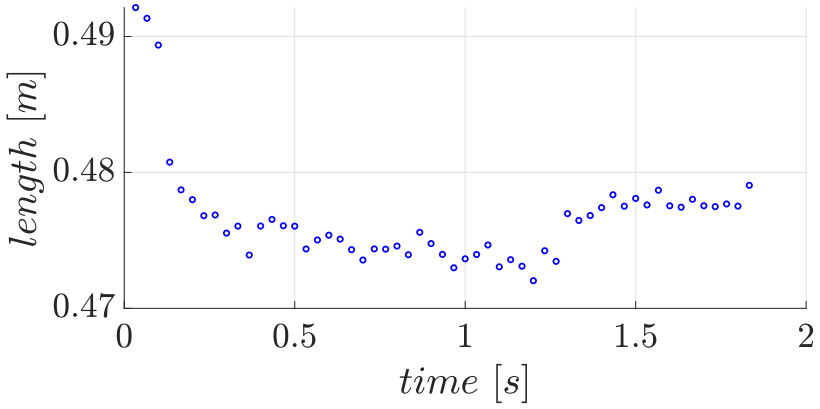


Figure A.3: Length of the torso.

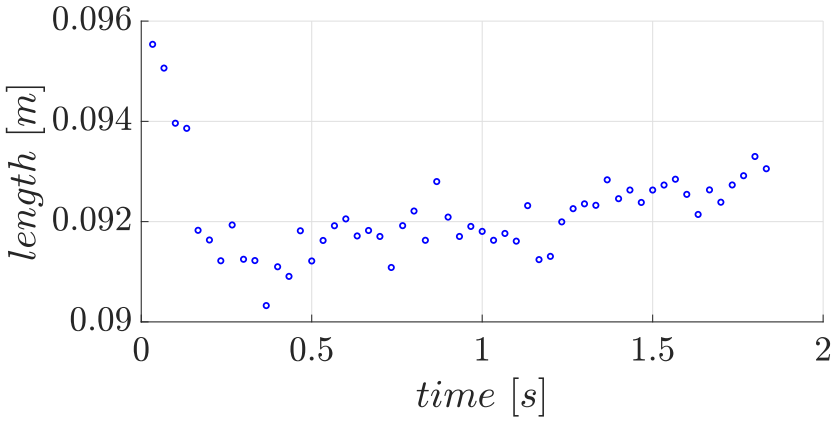


Figure A.4: Length of the minor radius the torso.

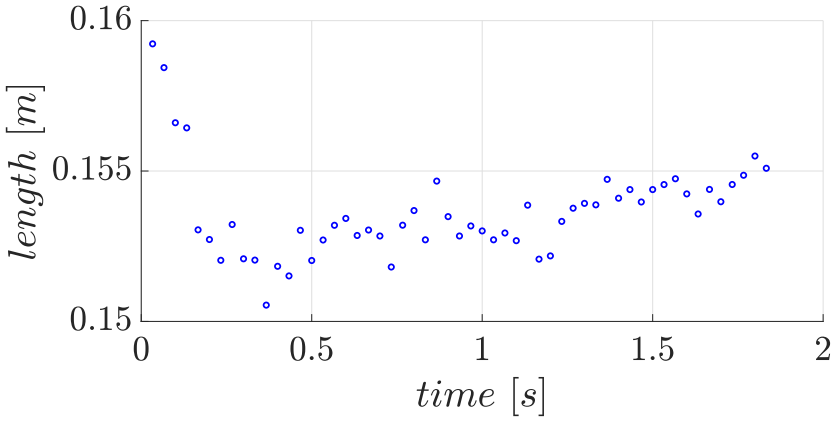


Figure A.5: Length of the major radius the torso.

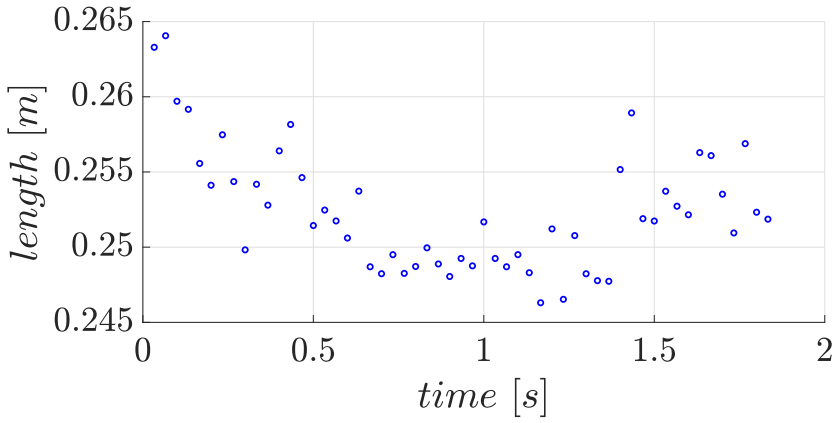


Figure A.6: Length of the upper arm.

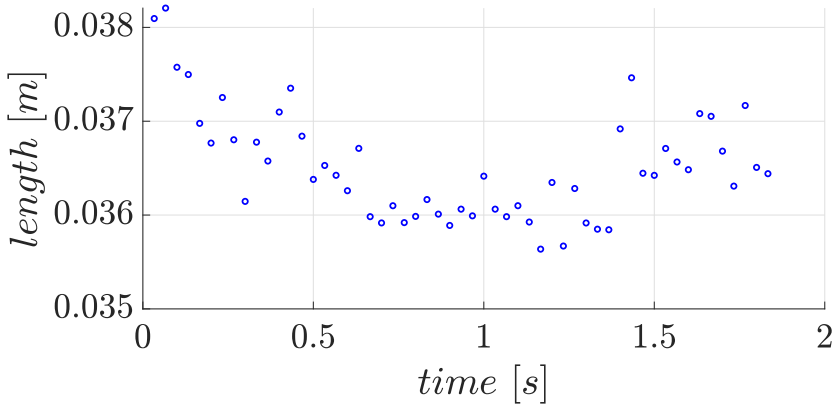


Figure A.7: Radius of the bigger section of the upper arm (shoulder).

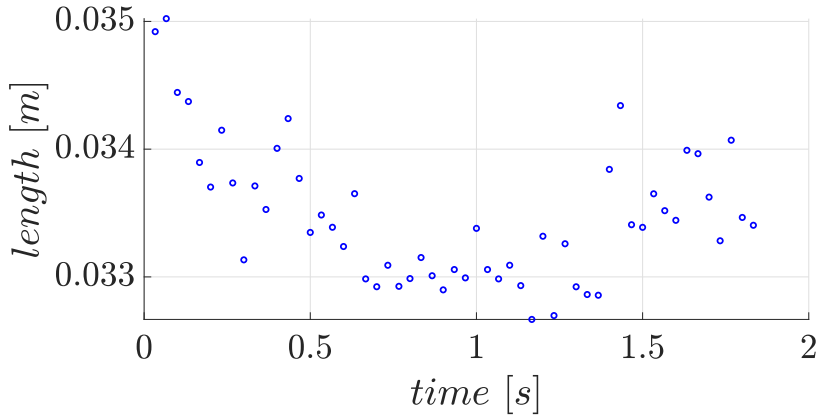


Figure A.8: Radius of the smaller section of the upper arm, which coincides with the radius of the bigger section of the lower arm (elbow).

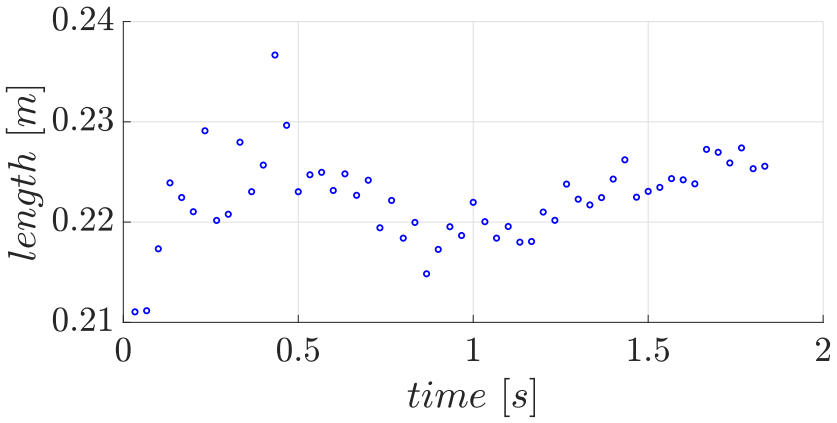


Figure A.9: Length of the lower arm.

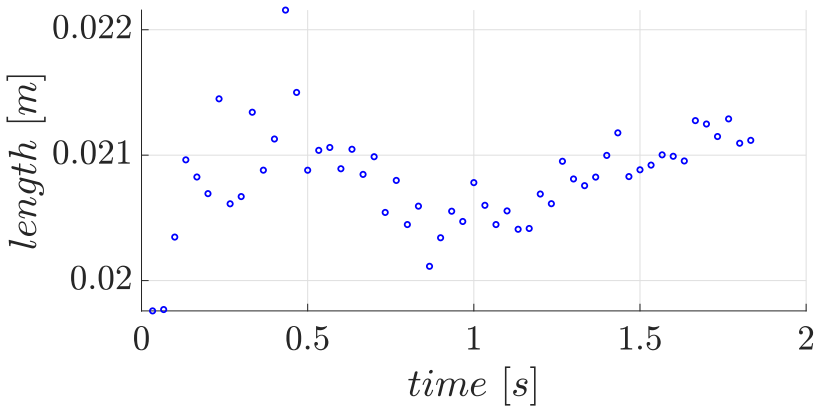


Figure A.10: Radius of the smaller section of the lower arm (wrist).

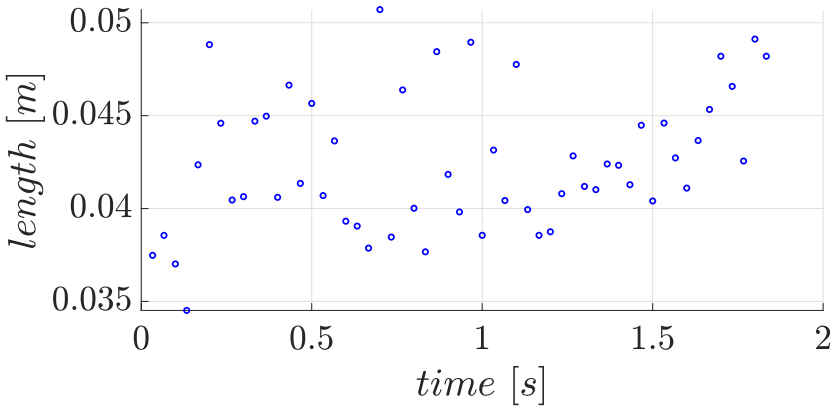


Figure A.11: Radius of the hand.

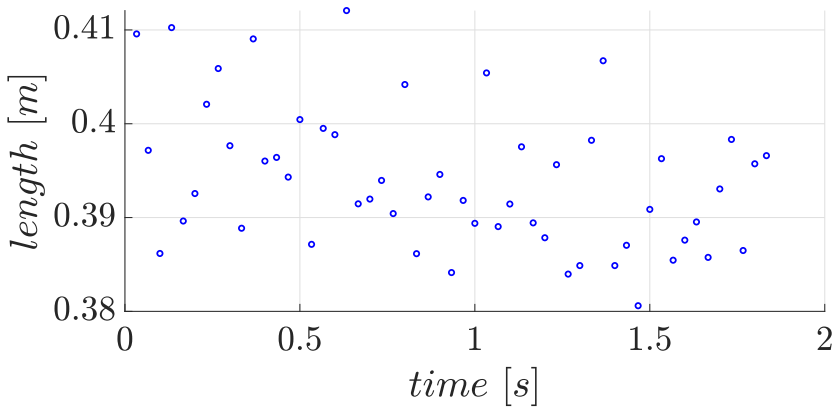


Figure A.12: Length of the thigh.

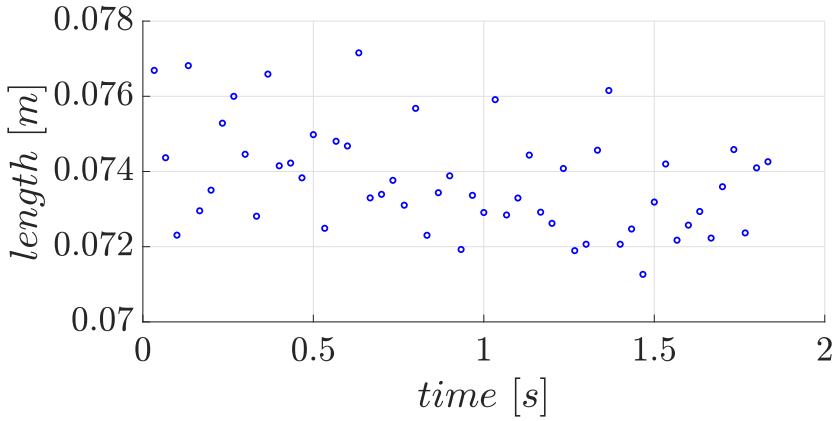


Figure A.13: Radius of the bigger section of the thigh.

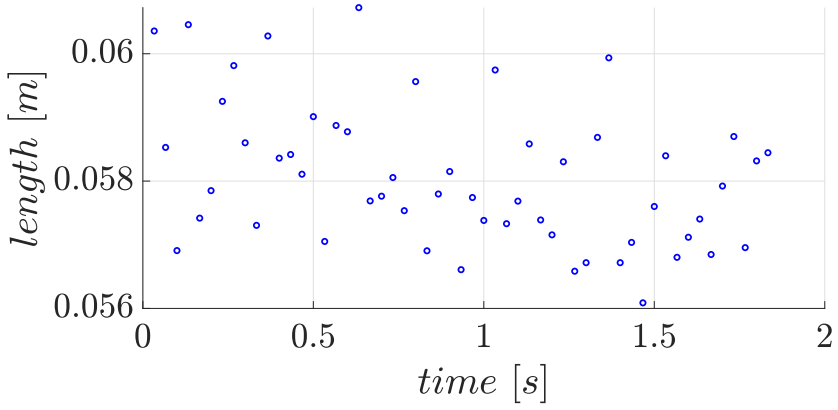


Figure A.14: Radius of the smaller section of the thigh, which coincides with the radius of the bigger section of the shank (knee).

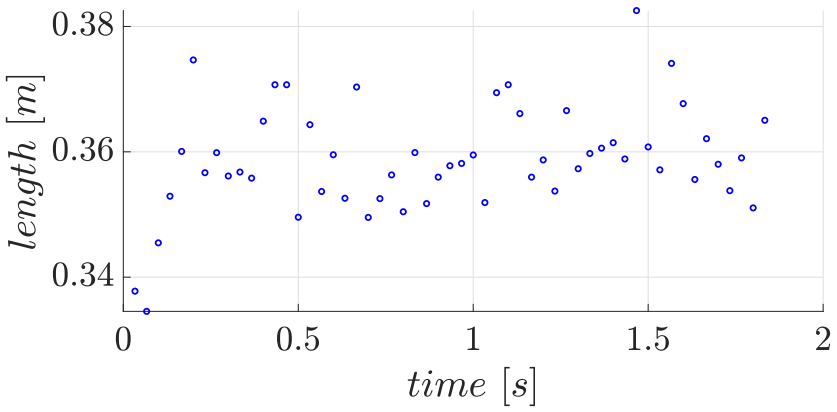


Figure A.15: Length of the shank.

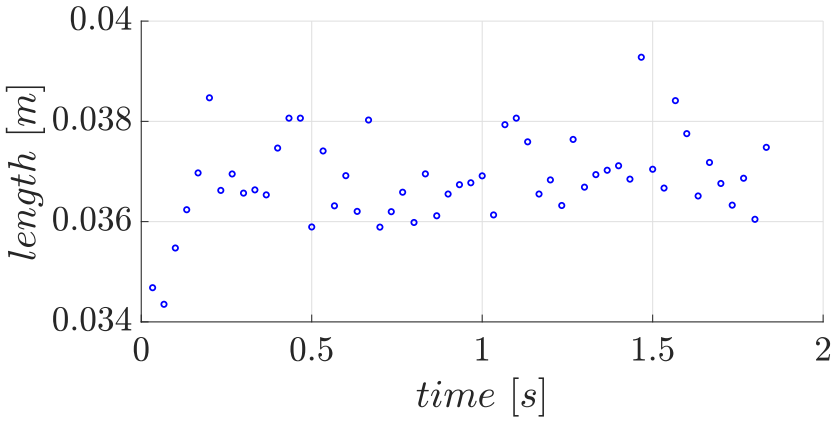


Figure A.16: Radius of the smaller section of the shank.

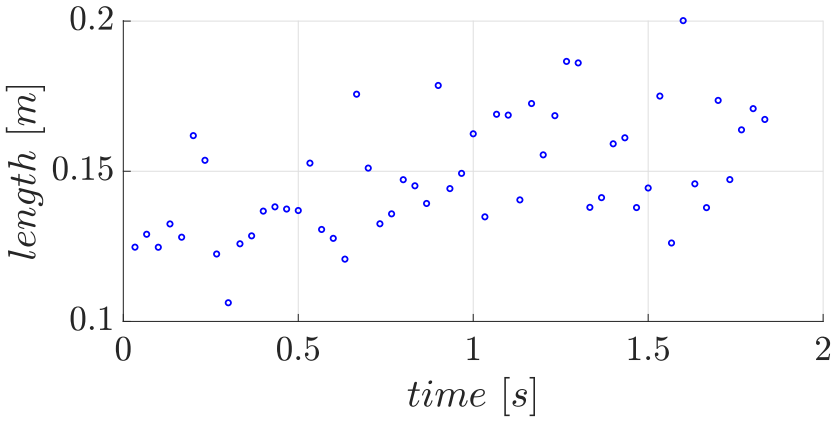


Figure A.17: Length of the foot.

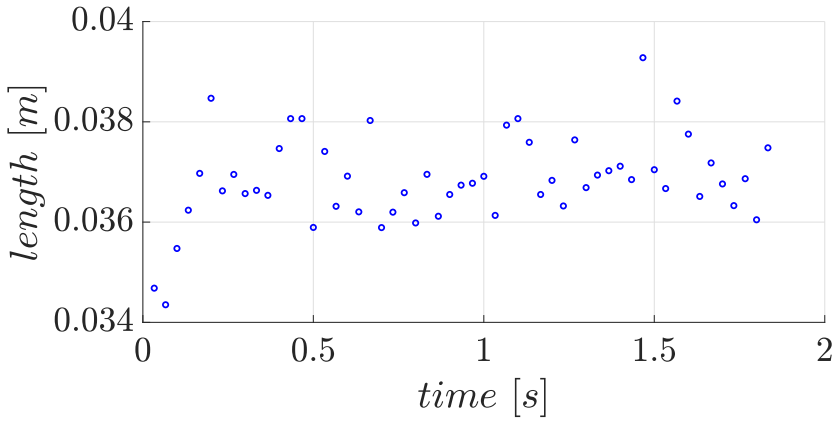


Figure A.18: Radius of the bigger section of the foot.

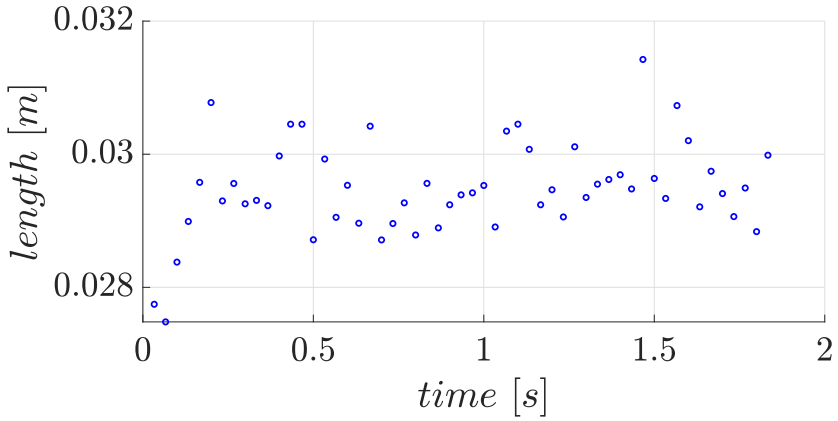


Figure A.19: Radius of the smaller section of the foot.

List of Figures

1.1 Shorter caption 12

2.1 Flow chart showing the sequence of passages followed for the identification of human body's inertial properties instant by instant. 16
2.2 Front picture of the stereo depth camera Intel RealSense D415. 16
2.3 33 points identified by MediaPipe Pose [48]. 18
2.4 17 body landmarks identified by TensorFlow's and YOLOv8's algorithms [55]. 19
2.5 Reproduction of the human body geometrical model in CAD. 21
2.6 Human body global reference frame for a 33-point model. 26
2.7 Human body global reference frame for a 17-point model. 26
2.8 Local reference frame of the frustum of cone. 28
2.9 Local reference frame of the ellipsoid 28
2.10 Shorter caption 28
2.11 Example of rotation of a local reference frame using the direction cosine matrix (DCM). 30

3.1 Position of the points identified by MediaPipe (red circles) and of the COMs of each body segment and of the entire human body (blue circles) for the 'closed arms' pose. 38
3.2 Position of the points identified by MediaPipe (red circles) and of the COMs of each body segment and of the entire human body (blue circles) for the 'open arms' pose. 38
3.3 Position of the points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for an asymmetric body pose. 40
3.4 Shorter caption 41
3.5 Shorter caption 42
3.6 Shorter caption 48
3.7 Shorter caption 51
3.8 Shorter caption 52

3.9 Variability of the length of the torso. 53
3.10 Variability of the length of the upper part of the arm. 53
3.11 Variability of the radius of the bigger section of the upper part of the arm. 54
3.12 Variability of the radius of the hand considering an individual performing broad movements. 55
3.13 Variability of the length of the foot considering an individual performing broad movements. 56
3.14 Shorter caption 59
3.15 Shorter caption 60
3.16 Shorter caption 61
3.17 Shorter caption 64
3.18 Shorter caption 65
3.19 Position of the 17 points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for the 'closed arms' configuration. . 67
3.20 Position of the 17 points identified by MediaPipe (red circles) and of the COM of the entire body (blue circle) for the 'open arms' configuration. . . 67
3.21 Comparison between the position of the COM evaluated using the original skeleton model identified by MediaPipe (blue line) and the reduced 17-point one (red line). 70
3.22 Comparison between the inertia moments evaluated using the original skeleton model identified by MediaPipe (blue line) and the reduced one (red line). 71
3.23 Shorter caption 72
3.24 Shorter caption 76
3.25 Shorter caption 77
3.26 Comparison between the variation of position of the COM in time evaluated using the reduced skeleton model identified by MediaPipe (blue line) and the one identified by MoveNet (red line) for a frequency equal to 90 fps. . . 78
3.27 Difference between the position of the COM evaluated in time using the reduced skeleton model identified by MediaPipe and the one identified by MoveNet for a frequency equal to 90 fps. 79
3.28 Comparison between the variation of the moments of inertia in time using the reduced skeleton model identified by MediaPipe (blue line) and the one identified by MoveNet (red line) for a frequency equal to 90 fps. 80
3.29 Difference between the moments of inertia evaluated in time using the reduced skeleton model identified by MediaPipe and the one identified by MoveNet for a frequency equal to 90 fps. 80

A.1 Length of the minor radius of the head. 95
A.2 Length of the major radius of the head. 95
A.3 Length of the torso. 96
A.4 Length of the minor radius the torso. 96
A.5 Length of the major radius the torso. 96
A.6 Length of the upper arm. 97
A.7 Radius of the bigger section of the upper arm (shoulder). 97
A.8 Radius of the smaller section of the upper arm, which coincides with the radius of the bigger section of the lower arm (elbow). 97
A.9 Length of the lower arm. 98
A.10 Radius of the smaller section of the lower arm (wrist). 98
A.11 Radius of the hand. 98
A.12 Length of the thigh. 99
A.13 Radius of the bigger section of the thigh. 99
A.14 Radius of the smaller section of the thigh, which coincides with the radius of the bigger section of the shank (knee). 99
A.15 Length of the shank. 100
A.16 Radius of the smaller section of the shank. 100
A.17 Length of the foot. 100
A.18 Radius of the bigger section of the foot. 101
A.19 Radius of the smaller section of the foot. 101

List of Tables

2.1 Frames per second recorded by the monochrome sensor of the RealSense D415 stereo-depth system at varying acquisition resolution. 17
2.2 Body segments' mass percentages adopted for the 33-point model. 24

3.1 Dimension, volume, mass, and density of each body segment for the 33-point model identified by MediaPipe. 36
3.2 Position of the COM and principal moments of inertia expressed in the local reference frame of each body segment for the 33-point model. 37
3.3 Comparison of segmental masses between the current study and De Leva's research. 43
3.4 Comparison of the positions of the segmental COMs between the current study and De Leva's research. 44
3.5 Comparison of segmental principal moments of inertia between the current study and De Leva's research. The unit of measurement for all the values is $kg\ m^2 \cdot 10^3$. 45
3.6 Position of the body segments required to run the Matlab code and to reproduce the poses in Inventor. 47
3.7 Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'closed arms' configuration. 49
3.8 Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'open arms' configuration. 49
3.9 Mean value, max deviation from the mean and standard deviation of the variability of the length of the most significant segments' dimensions. . . . 54
3.10 Dimension, volume, mass, and density of each body segment of the 33-point model evaluated for a moving body. 57
3.11 Maximum acceptable variations of the resulting position of the COM and inertia tensor for the 33-point model applied on the 'open arms' body pose. 62

3.12 Resulting maximum acceptable standard deviations associated with the input position data obtained performing the sensitivity analysis for the 33-point model on the 'open arms' body pose. 63
3.13 Position of the COM and principal moments of inertia expressed in the local reference frame of each body segment for the 17-point model. 68
3.14 Comparison of the position of the COM between the 17-point and the 33-point models identified by MediaPipe for 'open arms' and 'closed arms' body poses. 69
3.15 Comparison of the inertia tensor between the 17-point model and the 33-point one for 'open arms' and 'closed arms' body poses. All the elements of the inertia tensor are expressed in $kg\ m^2$. 69
3.16 Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'closed arms' configuration of the 17-point model. 73
3.17 Position of the COM and principal moments of inertia of the entire body obtained with the Matlab code and with the CAD model for the 'open arms' configuration of the 17-point model. 73
3.18 Maximum acceptable variations of the resulting position of the COM and inertia tensor for the 17-point model applied on the 'open arms' body pose. 74
3.19 Resulting maximum acceptable standard deviations associated with the input position data obtained performing the sensitivity analysis for the 17-point model on the 'open arms' body pose. 74
3.20 Comparison between the results obtained with TensorFlow's and YOLOv8's models and the reduced MediaPipe's 17-point model for a frequency of 30 fps. In the table are listed means and standard deviations of the difference between the results obtained with these algorithms, using always as reference the MediaPipe's 17-point model. 81
3.21 Comparison between the results obtained with TensorFlow's and YOLOv8's models and the reduced MediaPipe's 17-point model for a frequency of 90 fps. 82

List of Symbols

Variable	Description	SI unit
V	volume	m^3
R	major radius	m
r	minor radius	m
h	height	m
$\boldsymbol{COM}$	position of the center of mass	m
$\boldsymbol{I}$	inertia tensor	$kg\ m^2$
$\boldsymbol{DCM}$	direction cosine matrix	rad
$\hat{l}$	axis orientation	$-$
$\sum$	summation operator	$-$
σ	standard deviation	m
μ	mean	m