TÉCNICO LISBOA

Kinematic Data Filtering with Unscented Kalman Filter

Application to Senior Fitness Tests Using the Kinect Sensor

Manuel João Duarte Serejo Goulão

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. Alexandre José Malheiro Bernardino
Prof. Maria da Conceição Esperança Amado

Examination Committee

Chairperson: Professor Ana Luísa Nobre Fred
Supervisor: Professor Alexandre José Malheiro Bernardino
Member of the Committee: Professor João Miguel Raposo Sanches

November 2017

Acknowledgements

Agradeço a todos os envolvidos na realização deste trabalho, em especial aos meus orientadores.

Abstract

Sedentary behaviour and the lack of physical exercise are routines existing in a large segment of the population, specially in the elderly. This can lead to health complications and debilitating conditions, reducing the quality of life, as it drives to the inability to perform activities of daily living. The Microsoft Kinect sensor provides an accessible mechanism to implement automatic and assisted training systems, which can then be used, for example, for in-home training, or in care centres, either in a fixed platform or in a mobile assistive robot. However, the need to obtain movement parameters to perform fitness assessment requires data with a large degree of consistency. In this work, a procedure aimed at improving the accuracy of the estimates of the skeleton data from the Kinect sensor is proposed. The method is based on an approach using an unscented Kalman filter applied to a kinematic model, which aspires to reduce the noise and, at the same time, be able to achieve coherence in the estimation, by combining the measurements with previous knowledge on the movements being executed. In the end, it has been possible to extract information, such as angle and angular velocity corresponding to each degree of freedom of each joint, as well as to correct some of the errors present in the data.

Keywords: Biomechanics, Kinect v2 SDK, Motion Capture, UKF.

Resumo

O comportamento sedentário e a falta de exercício físico são rotinas existentes num grande segmento da população, especialmente entre idosos. Isto pode levar a complicações de saúde e condições debilitantes, reduzindo a qualidade de vida visto que leva à incapacidade de levar a cabo actividades da vida quotidiana. O sensor Kinect da Microsoft proporciona um mecanismo acessível que permite implementar sistemas automáticos de treino assistido, que podem ser usados para, por exemplo, sistemas de treino em casa, ou em centros de dia, tanto em plataformas fixas ou em robots assistentes móveis. Não obstante, a necessidade de obter parâmetros do movimento para uma avaliação da aptidão física exige dados com um elevado grau de consistência. Neste trabalho é proposto um procedimento que visa melhorar a precisão das estimativas dos dados do esqueleto obtidas pela Kinect. O método proposto baseia-se num filtro de Kalman *unscented* aplicado a um modelo cinemático, com o qual se pretende diminuir o ruído e simultaneamente conseguir coerência na estimação, combinando as medições com o conhecimento prévio dos movimentos realizados. Por fim, foi possível extrair informação como os ângulos e a velocidade angular correspondentes a cada grau de liberdade de cada articulação, para além da correcção de alguns erros presentes nos dados.

Palavras-chave: Biomecânica, Kinect v2 SDK, Motion Capture, UKF.

Contents

List of Tables

List of Figures

List of Acronyms

3D 3 Dimensional

ADLs Activities of Daily Living

AHA Augmented Human Assistance

API Application Programming Interface

CVD Cardiovascular Disease

DOF Degree(s) Of Freedom

EKF Extended Kalman Filter

FMH Faculty of Human Kinectics, Universidade de Lisboa

KF Kalman Filter

SDK Software Development Kit

SFT Senior Fitness Test

UKF Unscented Kalman Filter

UT Unscented Transform

Chapter 1: Introduction

In contemporary society, the decreased practice of physical exercise and a sedentary lifestyle characterize many people's lives. For various reasons, physical exercise does not play an important role, and sedentary behaviour is common in people from the different social backgrounds and in different ages groups, both in children (Tremblay et al., 2011), adults (Thorp et al., 2011), and specially in the elderly (National Center for Health Statistics (US), 2016).

Physical activity is of major importance in primary and secondary prevention leading to a reduction in the impairment from multiple diseases and disabilities, and also in attaining a better quality of life and a longer life (Biswas et al., 2015; Lee et al., 2012). In fact, physical inactivity and a sedentary lifestyle have many times been reported to have a relation with increased risk of a large variety of conditions, such as, Cardiovascular Disease (CVD) (Ford and Caspersen, 2012; Warren et al., 2010), cardiometabolic disease (Ekelund et al., 2012; Saunders, Chaput and Tremblay, 2014), all-cause mortality (Katzmarzyk et al., 2009; Thorp et al., 2011), type 2 diabetes (Hu, 2003; Wilmot et al., 2012), osteoporosis (Nikander et al., 2010; Warburton, Nicol and Bredin, 2006), and site-specific cancer incidence (Biswas et al., 2015; Lee et al., 2012). Cognitive benefits and an attenuation of cognitive impairment have also been found to be related with the regular practice of physical exercise (Ahlskog et al., 2011; Colcombe and Kramer, 2003).

According to National Center for Health Statistics (US) (2016) latest report's data, physical inactivity, by means of not meeting "neither aerobic activity nor muscle-strengthening guideline" (Department of Health and Human Services (US), 2008), has been found in 47% of adults. Moreover, it has been found that this number rises to 50% when between the ages of 45 and 64 years, and to 59% for people older than 65 years. In Europe, physical inactivity is one of the leading risk factors for health, and is estimated to attribute to one million deaths per year (WHO, 2015). This lifestyle, particularly in the elderly, might lead to debilitating conditions, reducing the person's independence, and preventing them from performing Activities of Daily Living (ADLs). According to Warburton, Nicol and Bredin (2006) "musculoskeletal fitness appears to be particularly important for elderly people and their ability to maintain functional independence". In order to classify the functional status of a person, one's physical condition is a key factor (Sawatzky et al., 2007). Indeed, mobility, muscular strength and balance are major components of someone's basic abilities to perform ADLs.

The Kinect sensor (Microsoft, N.d.) is a system developed by Microsoft, currently in the second version (Kinect v2), which is the version used in the present work. Firstly intended as a gaming device, it has seen a wide adoption in scientific research due to its low cost and easy availability, and specially after the launch of an API for use with Microsoft Windows. The Kinect SDK, made available for both version1 and version2 of the Kinect sensor, provides an easy interface for performing acquisitions with this device. Moreover, the capability to recognize the human body through a "skeleton" structure, providing the 3D positions and orientation quaternions (and more information), provides data in a marker-like system manner, with a certain

degree of accuracy. Actually, this sensor captures both depth and colour images, and through the publicly available SDK, real-time tracking and recognition of the human pose though a 3D virtual skeleton is possible with a sampling rate of 30 Hz (Shotton et al., 2013). Therefore, it provides an easy way to acquire position data that could then be described in a movement analysis application, as it may be used to track and analyse movement, and therefore exploited to extract movement features. Thus, in the recent years, multiple applications have successfully been tested and shown in a large array of scientific works in elderly care, as in fall prevention, and for rehabilitation purposes, as described by Webster and Celik (2014) in his systematic review of these Kinect applications.

Figure 1.1: Kinect v2 sensor (Microsoft, N.d.).

Also recently, Rikli and Jones (2013) developed a set of functional tests, the Senior Fitness Test (SFT), which allows physical fitness assessment in older adults, covering the major components of fitness: strength, endurance, flexibility, agility, and balance, which are required to perform ADLs. Notwithstanding, in terms of physical exercise programs, maintaining a strict and constant plan is usually difficult and often people quit due to reasons such as lack of motivation or support (Schutzer and Graves, 2004; Sinclair, Hingston and Masek, 2007). Nowadays, hospitals and health professionals have been using virtual environments designed with serious games using gamification techniques, *i.e* based on videos games, in treatment and fitness training, and also for a variety of applications like distractions during painful medical procedures, rehabilitation or improvement of motor skills (Michael and Chen, 2005).

However, even with the applications of these serious games in order to fight loss of interest, there are other aspects which must be taken into consideration. Actually, it is essential to consider the adequate level of physical effort to require from the participant, as someone will quickly loose interest if such a hard challenge is presented to her, that she cannot overcome. The same can be said for a challenge that is too easy and which does not require enough effort to be relevant for the purpose, and also to draw interest in playing the game. Therefore, being able to assess the condition of the participant is of relevance, not only in order to have an insight on the actual condition of the person but also for the best functioning and success of the system.

In this work, these factors were taken into consideration, and, since movement parameters take such

an essential part of the fitness assessment procedure of a person, there is the need to enforce integrity on the data from the Kinect in order to have a viable system to use. In fact, a system using state-of-the-art data processing methods and techniques has been created in this work, which makes use of information acquired using the Kinect v2 sensor and its SDK, and enforces a set of kinematic constraints in a kinematic filter method. This way, it should be possible to extract movement features with larger accuracy and biomechanical compatibility, as the data would be subject to multiple constraints, which were not enforced directly by the SDK but represent physiological traits, and so are very relevant in such an important application.

Furthermore, this work was specially designed targeting elderly people fitness care, since this segment of the population is the most vulnerable from a physically inactive lifestyle and has a higher difficulty regarding the practice of physical activity. In fact, it was based on the datasets acquired by Bernardino et al. (2016) in the Augmented Human Assistance (AHA) project, which made available multiple recordings of various sensors, including the Kinect sensor, of young and elderly participants while performing senior fitness tests, acquired at Faculty of Human Kinectics, Universidade de Lisboa (FMH) on 2015.

Therefore, in the development of this work, an innovative kinematic filter model was created, being described together with the context in which it fits, as it is presented an analysis of the ground-breaking models found on the scientific literature published in the last years.

1.1 Motivation

As previously stated, people have become more and more physically inactive, which has lead to the increase in numerous health problems in the population, which is threatening the quality of life of the world population. In fact, in young people, the problems of a sedentary lifestyle may not be quite obvious or immediately visible, but regarding elderly people, the effect of a poor physical condition is quite more serious. A reduction in the functional status, preventing the accomplishment of ADLs and often accompanied by loss of independence and therefore quality of life, is frequently considered to be worse than a reduction in the expectation of years to live. Therefore, trying to improve the capabilities of the older segment of the population is a great concern.

Consequently, it is of major importance to develop a strategy that would allow for the improvement of public health, through the use of accessible devices which allow for large scale applications. Accordingly, this project focuses on aiding on the execution of a system to perform automated assessment of someone's physical condition, and that would be able to provide not only the current fitness level, but also for the choosing of adequate exercises that should reduce the abandonment of the training program due to the feeling of lack of accomplishment and lack of motivation. This is done by improving the acquisition capabilities of the Kinect sensor, and extending the possibilities to create a system which could be used in assisted living, in home environments or elderly care centres. Also, gamification techniques were analysed, as they could incentive both the participation and to improve the results, not only keeping people interested, but also having the goal of improving their health conditions and their quality of life.

1.2 Objectives

This work aims at providing a scientific contribution through the development of a kinematic filter method to use with the Kinect sensor SDK data. In fact, as seen through an analysis on recent scientific research works, the Kinect device has been extensively used in fitness training related applications. Moreover, these works mostly use the data as provided by the Kinect SDK, however, the SDK data does not enforce the anatomical constraints that exist in the human body, and health related applications require data which has a biomechanical cohesion. Therefore, there is the objective to enforce this cohesion in the data though the restriction of certain movements that, for a known exercise being recorded, are not expected to happen. Moreover, a modelling of the human skeleton by the angles performed by each joint, should be achieved, giving the ability to then restrict the Degree(s) Of Freedom (DOF) which each joint has, preventing the data to describe non-physiological movements in directions that are not possible. Furthermore, besides the capability to restrict the data for each joint to the DOF that are expected, the capacity to identify the maximum and minimum angles that each joint may describe, and use this information to restrict the data to the accepted interval would be a major achievement in improving the accuracy of the data. Finally, the fact that there is a variation of the lengths of the bones in the data provided by the Kinect SDK also represents a non-physiological mistake in the data, and addressing its correction is also an end goal of this work.

1.3 Thesis outline

This thesis is divided into 7 chapters. First, the present chapter provides an overview to the subject being studied, as well as the tools used, and the motivation behind this study and its main objectives. Second, a few applications using the Kinect sensor which were found in the literature are presented through an analysis on biomechanical applications and an analysis of the Kinect accuracy, also, other devices used for kinematic applications are depicted. Third, the methods of fitness assessment used, and the methods for the creation of an application of a system for this purpose are described. Fourth, the theoretical principles applied for the creation of the model developed in this work are explained, and the method created is substantiated. Fifth, the results of the application of the developed algorithm to a dataset of senior fitness tests are presented. Sixth, there is the discussion of the advantages and limitations of the proposed method. Finally, seventh, the conclusions drawn from this work are presented, together with an analysis of the accomplishments and contributions of the work, and future developments which would likely further elucidate and improve on the current achievements.

Chapter 2: Literature review

With the advances in technology, and as its price is dropping, there is a large variety of accessible sensors available. The Kinect sensor has recurrently been studied and used in the last few years in scientific research, more specifically in the area of biomechanics. Furthermore, devices based on low-cost and/or wearable sensors are often used in diverse applications, as in rehabilitation, assisted living, and sports. The Kinect SDK allows for an easy method of obtaining kinematic data of the human body through time in a form of a 3D virtual skeleton. This simplifies the use of this device, and is a reason for the increasing number of works using it.

There are plenty of recent works on the general subject of biomechanical analyses using the Kinect sensor. In this chapter, some of the most recent and relevant works found in the literature will be presented, mostly directed towards the main focus of this project: rehabilitation and elderly care, in home environments. First, some works using Kinect applications will be analysed; then, a brief analysis of some studies on the accuracy of the Kinect sensor are presented; finally, some applications of a few devices other than the Kinect on the subject in study are described.

2.1 Kinect applications to biomechanics

A variety of biomechanical applications has been recently developed through the use of the Kinect sensor. These include both physical training, rehabilitation and fall prevention. Gamification techniques are often applied, as games serve as an incentive and motivation, keeping the users entertained and lengthening the period for which a system is used, consequently promoting the goal of regular physical exercise.

Alexiadis et al. (2011) created a motion acquisition system that using a Kinect system, and through the use of the skeleton data provided by the publicly available OpenNI SDK, was able to assess and score a dancer's performance, providing visual feedback in a 3D virtual environment. This was done by comparing the user's performance to a gold-standard, a performance from a professional. The scoring was based on the joint positions, the joint velocities and the 3D flow error, and a weighted mean of the three provided the final value. The system was capable of assigning different scores to different body parts, allowing the user to train specific and more exact movements. In the end, the results regarding the real-time evaluation of dancers were a high score for professional dancers and consistent rankings of amateur dancers, low deviations of ranking when using ground truth, different scores behaving alike, and low scores when synchronization of dance steps was lost.

Gabel et al. (2012) worked on a full body gait analysis using the Kinect sensor to continuously perform gait tracking at home by extracting stride information. In this work, lower-body and upper-body gait parameters were extracted from the 3D virtual skeleton data provided by the available Kinect SDK. Although working on gait analysis, data from the whole body was used to be able to achieve a better accuracy. The system

was concluded to be robust to environmental changes (Kinect being moved) and placement of the sensor (Kinect placed in different positions). The model was tested on 23 subjects with ages from 26 to 56 years old, with the recordings being executed while synchronized with in-shoe pressure sensors and a gyroscope attached to the wrist to provide validation for this model. The gait analysis was done by means of a simple state machine that identified the beginning and the different parts of the strides. A regression model was built, as an ensemble of regression trees was used to predict the desired values from a set of variables (positions and estimations of velocities and accelerations at each frame, centre of mass, and speed and direction of progress). The predictions generated from the model were found to be, in absolute value, within the interval of 35 to 71 ms for the stride durations when compared to the data from the in-shoe sensors, and a 0.9 correlation was found between the angular velocity estimated with data from the Kinect and the angular velocity extracted from the gyroscope.

Gerling et al. (2012) created a set of gestures and design guidelines for full-body motion-controlled games for older adults using the Kinect sensor and the 3D virtual skeleton obtained via the Kinect SDK. They also created a game and presented two studies in which gesture performance was estimated in terms of range of motion, agility and strength, and analysed the response of the participants to the game. Both static and dynamic gestures were created with assistance from a physical therapy expert. First, 15 participants (60 to 90 years old) with movement disabilities described an overall positive experience in questionnaires, with some finding it too simplistic. Although it was possible to calibrate the system for wheelchairs, the wide range of impairments led to a decrease in accuracy with 16.7% of correct gestures not being recognized. In the second study, the questionnaires of the 12 participants (60 to 91 years old) showed an increase in positive affect while no change in negative affect. Here, a calibration was performed to identify the adequate game difficulty, and only static exercises were used, so, completion rates of 90% to 100% were achieved. In the end, the system showed a positive effect on the participants' mood. Moreover, it was found that calibration improved accessibility by adapting the game to the players capabilities, that free play was not suitable without prior gaming experience, and that the lack of hand-held device was an advantage.

Parajuli et al. (2012) proposed a method aimed at senior health monitoring, specifically fall detection, by using the Kinect sensor for gait analysis and to evaluate posture, namely when changing from sitting to standing and vice versa. The Kinect SDK was used to extract the 3D virtual skeleton, and a support vector machine (SVM) using the coordinates of each joint as features was used to process the gait and posture data with the objective of distinguishing normal and abnormal gait. Various corrections to the data were made: the data was transformed to correct distortions due to the viewing angle; the data was cleaned, ignoring anatomical inconsistencies that could exist; finally, different data reduction techniques were tested, such as data saved as the relative change between frames, and ignoring the depth data as it "does not change much while standing or sitting down", or performing a standard gait analysis using pattern recognition techniques for classification between normal and abnormal gait. Support vector machines were chosen due to the high dimensionality of the data and their ability to generate non-linear and high-dimensional classifiers.

Moreover, as various types were tested, C-SVC was chosen with a Radial Basis Function kernel, also the SVM scaling of the data was shown to be critical for the accuracy. Ultimately, this system showed the best results among all tested, regarding posture recognition as in standing and sitting, and gait recognition as normal and abnormal gait, an accuracy of over 90% for sitting and walking tests was obtained.

Chen et al. (2014) used a Senior Fitness Test (SFT) system for a fitness measurement in-home application targeted at the elderly, employing inertial sensors and a Kinect sensor in a synchronized and collaborative way. The Kinect sensor, computing 3D virtual skeleton from the Kinect SDK, was used to monitor the correct pose using relative positions between joints and thresholds on angle values, while the wearable inertial sensors, which measured three axes acceleration and angular velocity, were used to count number of actions performed in the exercise. With this information, visual and audio feedback were used to guide the participants. The SFT movements used in this work were: the *chair-stand test*, the *arm-curl test*, the *step-in-place test* and the *8foot-up-and-go test*. Five subjects participated in these tests, performing the movements 40 times, 20 correct and 20 incorrect performances, such that the accuracy of the system could be estimated. As a matter of fact, no False Positives and 1.5% False Negative rate were obtained. Accordingly, it was concluded that SFT measurements could be obtained by the system under realistic conditions.

Mundher and Zhong (2014) presents a mobile robot employing the Kinect sensor to perform person following, allowing for the monitoring of elderly people, and specifically, to raise an alarm in case of need. More specifically, this work targeted the fall detection problem. The described system was able to interact with the human using gesture and speech recognition methods, which greatly simplified the elderly interacting with the robot. In order to detect and track the user, the system used the skeleton structure from the Kinect SDK, and the fall detection was done by computing the distances from the joints to the floor-plane, and defining a threshold on these values which would detect if a fall had occurred. However, in case the floor-plane was not identified, the falls were detected only based on the skeleton space coordinate system. Finally, in 4 simulated falls, the system was able to identify between 50% and 100% depending on the distance from the person to the sensor.

Stone and Skubic (2015), using a dataset of continuously collected data in 13 apartments of older adults with 454 falls (445 by trained stunt actors and 9 naturally occurring residents) for one year, aimed at performing fall detection using the raw values of the depth stream of the Kinect sensor as already studied and validated for this purpose (Stone and Skubic, 2011). With this goal, an individual was tracked over time and events were identified as the vertical state of an individual was characterized in each frame; median and average filters were used to filter the data. An ensemble of decision trees was used to compute the confidence that a fall had happened using features as minimum vertical velocity, maximum vertical acceleration, mean of the filtered measured vertical state signal, occlusion adjusted change for objects with a height below 38 cm, minimum frame-to-frame vertical velocity. The events were considered as near *vs.* far and occluded *vs.* not occluded, and cross-validation was performed for standing (98% detection), sitting (70% detection) and lying down (71% detection) positions, with one false alarm per month. The system was found to be

subjective to light conditions and the detection dropped significantly when the person was far or occluded.

Wang, Zhang and Leung (2015) show personal robot companion, aimed at encourage active living and fight loneliness. This robot is able to recognize and follow users using the Kinect sensor, while providing also a voice interaction system where the robot was able to talk using text-to-speech methods. This robot could be used both indoors and outdoors, being able to help carrying the user's belongings. The system used the Kinect SDK to process the skeleton data structure. Then, in order to increase robustness of the method, a tracking loss mechanism was implemented by using an 1 second window where the number of joints detected is counted, and using a threshold the tracking is considered to be viable or lost. Moreover, a false detection system was put in place to prevent the recognition of objects, doing so by considering the fact that false detection are often described by large illogical variations, and so, using the variance of position, false detections could be prevented.

Wang et al. (2015*b*) developed a remote health coaching system with the Kinect sensor, using the 3D virtual skeleton data from the Kinect SDK. It aimed at being able to automatically coach elderly users to perform physical therapy at home while tracking their performance throughout the plan, and taking into consideration the motivation factor that could lead them to quit the program. In the system, the performance of the user was evaluated in real-time from repetition counting, most moving joints, reachable range and moving time. As the data from the Kinect was considered to be noisy, an approach using kinematic filtering was used to guarantee the integrity of the kinematic parameters. Thus, an unscented Kalman filter and a kinematic model of the human skeleton were combined for the purpose. Furthermore, outlier removal was performed due to loss tracking, using a model which combined a Gaussian model for the random systematic errors and a uniform model for tracking loss errors. The kinematic model interpreted the human skeleton as a kinematic chain, considering the different articulations as joints, each with 1, 2 or 3 degrees of freedom. The kinematic filtering used the parameters position, segment length and rotation quaternion as input vector and the joint positions as observation vector, the transition was considered as a random-walk. Lastly, it was possible to obtain "smooth kinematic parameters from the noisy data". The same authors have performed a study on the accuracy of the Kinect sensor (Wang et al., 2015*a*), and it will be analysed on Section 2.2.

Kurillo et al. (2015) and Ofli et al. (2016), including some of the authors from (Wang et al., 2015*b*) (preceding paragraph), present an automatic monitoring system using gamification techniques to motivate elderly people to perform regular physical activity in a home setting, using the 3D virtual skeleton from the Kinect SDK. The system aimed at improving flexibility, strength, and balance, using exercises with different levels of demand. To quantify the performance in a score, non-linear combinations of features were used, such as joint angles, angles relative to the vertical or horizontal plane, distances, and absolute positions. The data was filtered using a moving average filter and the correctness of each execution obtained by identifying the multiple stages of an exercise, and computed from pose measurements using fixed and dynamic thresholds. The system was tested in a 18 week study (Kurillo et al., 2015) of 7 individuals (77 to 96 years old), and in a 6 weeks (or 10 sessions) study (Ofli et al., 2016) of 6 individuals (74 to 91 years

old). The general health and physical fitness were measured using the SFT before and during the study, and guided by a health coach. Subsequently, the users gave a positive feedback but some felt the exercises were too easy or too hard. Finally, even though some trends were observed relating the chosen features for the system and the external outcome measures (such as SFT), no significant correlation was found. For a study by some of these authors on the accuracy of the Kinect (Obdržálek et al., 2012), see Section 2.2.

2.2 Kinect accuracy for biomechanical analyses

Diverse biomechanics studies have been performed using the Kinect sensor, thus multiple works on the accuracy of this device for this purpose have also been published in order to validate its use. Therefore, some works regarding the accuracy of the Kinect sensor with different specific aims will be analysed. Works using the Kinect v2 and the 3D virtual skeleton from the Kinect SDK will be preferentially considered as they are the resources used in this study. Notwithstanding, one work on the first version and a comparison between the two will also be presented to substantiate some of the works described in Section 2.1.

Obdržálek et al. (2012) studied the 3D virtual skeleton data from the Kinect v1 sensor and SDK, against a marker-based system (Impulse, PhaseSpace) of 9 infrared cameras at 480 Hz and 43 active LED markers, with sub-millimetre accuracy. After, the pose was derived using PhaseSpace Recap and Autodesk MotionBuilder. The exercises consisted on smaller and larger movements, in different views and using different joints, as pose, distance, and occlusion are said to greatly affect accuracy. An abnormal high placement of the hips was found. Using the LED markers as ground truth for pose estimation, the Recap system presented an error of about 1 cm, and the Kinect and MotionBuilder about 5 cm. Then, considering the Euclidean distance between the Kinect and PhaseSpace, the largest differences (average, frontal view) were in the hips (224 mm, std 25 mm) and ankles (170 mm, std 39 mm). In addition, the number of outliers was larger for joints down the kinematic chain, and a decrease in performance was found when away from a frontal view. A mixture of a Gaussian and a uniform distribution, for systematic and tracking loss errors respectively, modelled the data reasonably well. Finally, in opposition to the Recap and MotionBuilder fixed bone lengths, the variations in the Kinect were significant and claimed that this should be addressed, as from a T-pose to sitting, leg length varied 10 cm. During a controlled body posture, the Kinect is comparable to a motion capture system, however in a general posture, the variability of pose estimation is about 10 cm.

Wang et al. (2015*a*) compared the accuracy of the Kinect v1 and v2 and their SDKs, and a marker-based motion capture system (PhaseSpace Impulse X2) using 8 infrared cameras at 480 Hz and 43 LED markers with sub-millimetre accuracy, computing the skeleton via PhaseSpace Recap 2. Sitting and standing movements were performed at different speeds and orientations. The Recap 2 performs calibration, computing and constraining the bone lengths, in opposition to Kinect, which does not. In terms of results, both the Kinects had a standard deviation of about 10 mm to 50 mm, although significantly larger for moving joints, also, deviations increased with the viewpoint angle. Nevertheless, while moving, the Kinect v2 performed

better than the v1. Furthermore, to model the outliers a mixture of a Gaussian (random errors) and a uniform (tracking loss errors) distributions was used, reducing the standard deviations of joint positions by 30% to 40% and improving the accuracy. From the model, the Kinect v2 was found to have less outliers than v1, hence being more reliable for occlusions and rotations. Finally, the Kinect v2, with an average offset of 50 mm to 100 mm, was found to have better accuracy, as well as lower variations in bone length than v1, except for feet and ankle where an offset of 100 mm or more was found. In fact, the Kinect v2 had an offset (average, frontal view) for the hips of 114 mm with 20 mm std; for the knees 105 mm with 21 mm std; for the ankles 122 mm with 24 mm std; regarding bone length, the average variation was 43 mm with 15 std.

Giblin et al. (2016) analysed the Kinect sensor and SDK, and a state-of-the-art biomechanics software that applies corrections to this data (Capture, Kitman Labs), in comparison to a Vicon Bonito 10 marker-based system, using 6 infrared cameras at 100 Hz with 39 reflective markers, and Nexus 2.1 was used for data processing. The Capture system performs a calibration phase in which limb lengths are computed using the Euclidean distance, performing corrections. These corrections are computed from the ratio between the length computed in calibration and the length in a frame (f), fixing one of the joints and correcting the other in a rigid body, as $\vec{j}_{corr} = f \cdot (\vec{j}_{uncorr} - \vec{j}_{fixed}) + \vec{j}_{fixed}$. The process consists in fixing the hip joint and applying the correction to each limb successively, considering a fixed and a flexible joint, and to all the frames. One subject performed a countermovement jump, 3 repetitions, and the movement phases were identified. Then, the knee angle was computed as a 2 dimension angle between the upper and lower leg. The results demonstrated an improvement from -9.6°$\pm$3.4° to -3.8°$\pm$3.0° in the left, and from -5°$\pm$1.6° to 1.7$\pm$2.8° in the right knee. Moreover, from a precision of 3.0° (left) and 2.8° (right), repeatable results were found. Therefore, bone length calibration significantly improved results.

Otte et al. (2016) analysed the Kinect v2 sensor and SDK, against a Vicon (MX13 and Nexus 2.1) marker-based system, with 16 infrared cameras at 100 Hz with 36 reflecting markers, 2 mm accuracy. The Euclidean distance between the systems was 1 cm to 2 cm, but for feet and ankles exceeded 5 cm. Also, in the anterio-posterior and medio-lateral directions, Pearson's correlation coefficient (r) was larger than 0.7, except for ankles and feet where it was lower. In the vertical direction values were lower than 0.7 for 13 of the 21 joints. The movement type and amplitude also impacted the accuracy. In terms of signal-to-noise ratio, the feet and ankles had again the worst results, under 0 dB, but excluding these, larger values were obtained in the anterio-posterior (over 10 dB) than in medio-lateral (over 5dB) and vertical directions (-10 dB to -8 dB), as well as the latter having more outliers, considered for clinical parameter calculation. For absolute and consistency agreement between systems, 30 of the clinical parameters had an ICC(3,1) $>$ 0.7, and 38 had $r >$ 0.7. However, a systematic measurement bias was found on the range of motion of the knee while walking on the spot; and small range of motion on spine and hips, and poor accuracy for vertical movement were found during a short walk sequence. On repeatability, the standard error of measurement was lower than 20% in 31 parameters and ICC(1,1) $>$ 0.6 in 33. Finally, clinical parameters from axial landmarks and upper body were considered to be valid, while feet and ankle positions showed a considerable instability.

2.3 Kinect alternatives in biomechanical studies

For biomechanical studies, multiple easily accessible devices are also often used besides the Kinect sensor, and numerous works have been reported in the last few years. Considering these different devices, and excluding some of the less accessible and less affordable but also extensively used in scientific research systems (*e.g.* multiple high-precision camera systems using markers), a work using a gaming console, a mobile application and a wearable device are described below.

The Nintendo Wii gaming console uses games controlled by external motion-based devices, such as Wii Fit, a game that promotes exercise and which has been studied by Graves et al. (2010), who found that it provides "light-to-moderate intensity" physical activity, with an energy expenditure and heart rate grater that handheld gaming but lower than treadmill exercise, although an enjoyment rating higher than the latter. Another study using this device, together with the Nintendo Wii Balance Board, is presented by Young et al. (2011). There, a study on the centre of pressure is performed with functional balance training games, with increasing difficulty, associated with prediction of fall risk, maintaining a static position or reaching limits of stability, inside a virtual environment and providing visual feedback of the centre of pressure. 6 older adults (mean 84.1 years old) performed 10 sessions of 20 minutes during 4 weeks, where they were asked to maintain a static position, with the eyes open and with the eyes closed, the variability in the medial–lateral and anterior–posterior planes were studied and a decrease in the mean sway variability was found, but only in the eyes closed condition and in the anterior–posterior plane, the decrease reached significance.

Kranz et al. (2013) report on a pervasive mobile application, using phone integrated sensors, that works in collaboration with a fitness balance board for physical exercise and rehabilitation support, and physical activity assessment. An exercise set was developed with a sports medicine specialist, and 6 users (25 to 33 years old) performed the exercises. In order to evaluate performance, movement angles, smoothness, continuity, regularity, tempo, balance, repetitions and overall correction were analysed. A calibration was performed and in one hand, for local quality assessment, Principal Component Breakdown Analysis (PCBA) was used in a sliding window, through the analysis of eigenvalue spectrum with a predefined threshold for reconstruction quality (95%), and the reconstruction error used as a measure for quality of movement. On the other hand, for global analysis, the axes of largest energy were analysed, a Kaplan–Meier estimation was used to form the empirical distribution function which compared to a template gold standard and a rule-set related to expert knowledge was used on extracted measures to give feedback to the users. When comparing to expert opinions, the error was $<$10% for 77.86% (magnetometer) and 65.89% (accelerometer) of the cases for dynamic exercises, and 89.58% (magnetometer) and 90.63% (accelerometer) for static exercises. The feedback from the users showed acceptance and potential for a long term adoption.

Pierleoni et al. (2015) use a wearable device on the waist, including a 3-axes accelerometer, gyroscope and magnetometer, to provide the orientation and aiming at performing fall detection in real-time. In order to merge the information from the three sensors, a Madgwick orientation filter algorithm was employed. To

identify a fall, the “impact”, “aftermath”, and “posture” phases were computed through threshold comparisons, the first two based on the acceleration and the third one on the orientation. As such, the thresholds were selected by observation of experimental data. The thresholds for the acceleration were computed offline, after accurate training of a SVM method using a one dimension SVM base form with linear threshold functions. The thresholds for the duration and orientation were based on other studies. Machine learning techniques, such as neural networks or Hidden Markov Models were not used as it was infeasible on a real-time application on the used low cost microcontroller. The experimental protocol consisted on backward, forward, lateral left and right falls, as well as syncope, and ADLs (to identify false alarms), where the recovery from a fall was also considered. 10 subjects (22 to 29 years old) performed 18 different states (simulated falls, with and without recovery and ADLs). Finally, the algorithm achieved a 90.37% accuracy, 80.74% sensitivity, and 100% specificity, performing better when comparing to an accelerometer only approach.

In conclusion, with the growth in available technology which allows the development of assisted living applications, many studies bring forth new systems that have the ability to improve the quality of living of many older adults. In fact, the Kinect sensor is being employed in many scientific works in the area of biomechanics, however, some works fall short when describing the details of their systems. Nevertheless, the Kinect has been found to be a better device than various alternatives for application in elderly care, particularly due to the lack of a hand held device. Moreover, the provided free SDK simplifies the know-how required, and the accuracy of the sensor was found to be quite high, although not on par with professional marker-based systems. Therefore, the Kinect presents itself as a premium device for the focus of this work.

Chapter 3: Fitness assessment

The ability to viably identify the physical attributes of a subject takes the consideration of a large number of factors. Therefore, the problem should be reduced to smaller parts beforehand, a set of different categories, each of which contributes on the analysis of the global performance: muscular strength, aerobic endurance, flexibility, balance, and agility. These are important factors when analysing the fitness abilities and functional fitness characteristics of a person, and specially when targeting the abilities to perform ADLs by elderly people (Rikli and Jones, 2013), the main motivation of this investigation work.

Determining the key aspects of each of the fitness fields in a meaningful way requires a valid and reliable standard to be followed. Here, the SFT (Rikli and Jones, 2013) is employed for the task at hand. In such manner, it should be possible to find quantifiable measurements that are defined by the movements of the subject that were captured by the Kinect sensor, and which describe the measurements obtained from the performed SFT exercises performed. By doing so, it is possible to identify what is being exploited in the exercise, and therefore it should be viable to use this information to reduce the burden of some exercises, *e.g.* why do a repetitive test for a long time when there is a test which measures the same factors in half of that time. Or, use the information embedded in other means (such as a game) which are not explicitly fitness assessment physical tests, and therefore are not so cumbersome. Thus granting the ability to maintain the interest of the participants in performing physical exercise by raising the enjoyment in performing those activities.

In the following sections, first, the set of fitness tests being adopted in this work will be individually explained, and their most relevant aspects will be covered. Second, gamification, a technique to turn an activity more appealing by enclosing it in a gaming environment will be analysed. And third, a review of the features often used in the literature regarding fitness assessment will be presented.

3.1 Standard tests for fitness assessment

The SFT (Rikli and Jones, 2013) is a set of fitness assessment tests aimed to assess the physical abilities and functional capabilities of an elderly person, using for that purpose a set of exercises, each targeting a different fitness area or areas. In the following work, not all SFT exercises were employed, as the most relevant for the intended application were selected (the 2-minute step, 30-second chair stand, and 8-foot up-and-go), and upper body exercises were left excluded. Also, in order to study static balance capabilities, another test, the unipedal stance test, which is not included in the SFT, was applied.

The used fitness tests and their areas of application are shown in Table 3.1, while a brief description of the tests and scoring methods may be found in Table 3.2.

Table 3.1: Fitness tests employed in the present work, belonging of tests to the SFT set of exercises, and corresponding fitness areas for each exercise.

Fitness test	**SFT?**	**Main fitness area**
2-minute step	Yes	Aerobic endurance
30-second chair stand	Yes	Lower-body strength
8-foot up-and-go	Yes	Agility and dynamic balance
Unipedal stance	No	Static balance

Table 3.2: Scoring methods for the applied fitness tests.

Fitness test	**Scoring type**	**Scoring method** (Rikli and Jones, 2013)
2-minute step	Count (integer)	Number of full steps, reaching midway between patella and iliac crest, completed in 2 minutes.
30-second chair stand	Count (integer)	Number of full stands from seated position, with arms across chest, in 30 seconds.
8-foot up-and-go	Duration (continuous)	Time from a seated position, walking 8 feet (2.4 meters), turning, and return to seated position.
Unipedal stance	Duration (continuous)	Time from raising the leg, with arms across chest, until putting the leg down or using the arms (for a maximum limit of 30 seconds).

3.1.1 2-minute step test

The 2-minute step test aims at assessing the aerobic endurance of a subject. As such, the person performing the test must raise each knee one after the other, assuring that the knee raises to a point "midway between the patella and iliac crest" (Rikli and Jones, 2013), and trying to do the maximum number of repetitions during the time period of two minutes. The number of repetitions of the exercise is considered as the achieved score.

This exercise is employed as it closely mimics walking and climbing stairs, having multiple ADLs dependent on these actions. *E.g.* going on a walk or shopping, overcoming road obstacles, and other mobility exercises that take a certain duration and therefore require considerable physical endurance.

3.1.2 30-second chair stand test

The 30-second chair stand test is employed in order to determine the lower-body strength of the person performing the exercise. It takes the action of standing up and sitting down on a chair, successively, for a period of thirty seconds, with the subject trying to achieve the maximum number of repetitions in this period

of time that he or she is able to. The number of repetitions of standing from and sitting on the chair that are accomplished are considered the score obtained on the exercise.

The chair stand test verifies the abilities to perform a large variety of ADLs, as climbing stairs, walking, getting out of a car or tub, all dependent on lower-body strength, moreover, this exercise is associated with reduced chance of a fall (Rikli and Jones, 2013).

3.1.3 8-foot up-and-go test

The 8-foot up-and-go test targets the fitness areas of agility and dynamic balance of the subject. This test starts with a person sitting down on a chair, and requires her to stand up, walk 8 feet (2.4 meters), turn around, and return to the original seated position, with the pace at which the exercise is performed determining the score. The performance of the exercise is measured as the time, in seconds, that the person takes from standing up to returning to the seated position which she started on.

The test attempts to quantify the abilities of the subject to execute ADLs that require the ability to perform faster movements, such as getting off the bus, going to the bathroom, cooking, or answering the phone (Rikli and Jones, 2013).

3.1.4 Unipedal stance test

The final test is the unipedal stance test. This test is not in the SFT, but was created as a measure of the static balance of a subject, which is considered of interest. To perform this exercise, the subject crosses her arms across the chest, while raising a leg from the ground. The performance is measured as the time that the person takes, from raising the leg until putting it down again, without loosing balance. There is a maximum limit of thirty seconds on the exercise, after which the test is considered to be complete.

The analysis of static balance allows for the assessment of the condition to perform a large variety of ADLs, as most daily activities require standing, and is a large factor in fall prevention. Riding the bus, and walking in a narrow line or sidewalk are examples of such activities.

3.2 Gamification process

The process of gamification implies the exploration of human-computer interactions through the use of game design concepts (Deterding et al., 2011) with the purpose of, in the case of this study, motivate regular physical exercise in the elderly population, while promoting enjoyment of the people performing it by using game characteristics. Furthermore, the concept of games that incorporate physical activity on their play are often referenced as exergames (Staiano and Calvert, 2011).

In the literature, the concept of gamification has been largely studied and exploited in a variety of areas. Hamari, Koivisto and Sarsa (2014) did a the literature review on 24 studies which use gamification techniques in multiple areas, as: education and learning, health and exercise, or intra-organizational systems,

among others. It is shown that most works have positive effects on all or part of the tests employed.

In addition, Larsen et al. (2013) conducted a systematic review in a scope more specific for this work, as it targets randomized controlled trial studies based on the effects of the employment gaming systems that promote physical activity in the elderly population. It was found that these games have the potential to improve the health of the elderly, nevertheless, the methodology of the studies in question was found to be of low-to-moderate quality. Therefore, there is the requirement for, not only longer and better designed studies to assess the effectiveness and long-term adherence, but also studies comparing the exergames with the best available alternative, and also exergames personalised for each individual, *i.e.* exergames focused on the needs of each user (content, design, and demands) were found to be desirable.

Finally, in terms of the present work, the integration of the employed tests (Table 3.1) on a videogame environment requires the adaptation of the defined physical exercises into a set of movements or movement characteristics that may be incorporated in a gaming environment. Consequently, since the scores that the tests provide are descriptive of the physical abilities of the subject who performed them, there is the need of finding such scores by means of some other activity or movement that the subject performs while playing the game. As walking for 2.4 meters or to march for 2 minutes would not make a very interesting and captivating game, some other way of finding the scores is required that may be used to create an appealing activity. Accordingly, some rudimentary measures extracted from the exercise may be related with the scoring of the test. If that would be the case, these features may be extracted from the activity performed in a game designed for the purpose, and used to obtain a related score to one that would be found by performing the actual test. This way, games that allow the measurement of a set of features that describe the scoring from the tests, may be used to measure the fitness characteristics of an individual, as well as be used to promote the practice and development of the physical attributes in question, and moreover, being an incentive to the accomplishment of regular physical exercise.

3.3 Features for fitness assessment

There is no unique way to assess the physical capabilities of a subject, as such, a diversity of methods are found throughout the literature. One way to perform fitness assessment, which was already depicted above is the SFT. However, works on the subject present more ways of accomplishing the desired physical evaluation, by doing so using more rudimentary measurements and features that, in opposition to the SFT exercises, could be swimmingly integrated in a desired gaming environment.

As a matter of fact, actual works on the relevance of features of each of the movements are not common in the literature, therefore, there is the need to search other works on the subject of fitness assessment in order to analyse possible relevant features. In the end, joint angles and positions, and velocities, or measures directly related to these are the ones that are most often used, as well as absolute, or relative displacement of the joints or other body segments.

For instance, Gabel et al. (2012) use measures like the position and direction of progress of the centre of mass, and the speed of walking. Also, they compute the positions, velocities, and accelerations for each of the joints considered in their biomechanical model, in order to be able to achieve the goal of extracting the needed gait parameters used for a full body gait analysis, in which their work is focused on.

Gauthier and Cretu (2014) propose a monitoring system for physical exercise based on the Kinect sensor. They use features to quantify the performed movements such as the positions of the joints and their velocities, the angles at the joints (using the scalar product considering the corresponding segments connected at the joint). Also, they compute the functional working envelope, corresponding to the volume generated considering every possible point touched by a given body, and the rate of fatigue, as the rate of decay of the velocity of the movements being performed, considering the decline in velocity as being associated with the fatigue.

Another study using interesting and relevant movement features would be Moniz-Pereira et al. (2014), who analyse the influence of different pose estimation models, using a marker-based system, and found, on one hand, that if the segments are considered with six DOF, nonanatomical dislocations from soft tissue artefacts may occur. On the other hand, that restricting the model to three DOF on the hip, one on the knee and two on the ankle may not be appropriate as it may hide certain movements that happened in reality. Finally, the features of joint angles are described to be more sensitive than joint moments.

Ofli et al. (2016), use joint angles, distances, and absolute positions as their model's features. However, in this work, the angles of each of the joints are computed relative to the vertical and horizontal plane. These features are used in a model that is intended to quantify the performance of a set of physical exercises which are used in this work's coaching system for older adults.

In conclusion, there are no standard features that are universally used, or even consistently useful. In fact, measurements such as joint angles, positions and velocities, are the ones more often found, as depicted in the works described above. Therefore, the more basic measures are also the most commonly found in the literature, alternatively to more complex and intricate features. And so, systems are usually composed of multiple features, whose best combinations are frequently found to be also from the more rudimentary ones. Furthermore, the capability to automate the measurement of the features are a large advantage in relation to usual methods of measuring them manually, which requires more time and physical resources, as well as introducing a bias depending on which person is performing the measurement.

Chapter 4: Human kinematic motion analysis

The Kinect sensor and the data obtained using the Kinect SDK, as a set of Cartesian coordinates for multiple joints, do not guarantee kinematic and anatomic coherence. Thus, it is important to understand the data behaviour, and the anomalies that should be addressed. In fact, the SDK provides a skeleton model described as 3D coordinates for each of the joints. This data was modelled by a kinematic chain using the Unscented Kalman Filter (UKF) algorithm to enforce multiple constraints, which target the correction of different types of errors present in the data. In reality, the data from the Kinect sensor and Kinect SDK had multiple faults as the algorithm which is implemented by this library does not impose any kinematic or physiological constraints. Indeed, there is no guarantee that the links connecting the joints have constant length throughout the acquisition. This is described as being an important factor respecting the accuracy of the Kinect by Obdržálek et al. (2012), who were able to achieve better accuracy by enforcing fixed length constraints on the data. Other issues could be seen as non-physiological positions being present, *i.e.* positions that are not possible to be performed by a human being, and therefore must be erroneous.

A search on the recent literature of data processing on biomechanical studies using a skeleton model from the Kinect sensor's data revealed that many works do not describe or perform any processing on the data before the analysis or before the implementation of a method. Nevertheless, there are some exceptions which do apply processing methods, the most frequently employed were: moving average filter (Andersson, Dutra and Araújo, 2014; Kurillo et al., 2015; Stone and Skubic, 2015), low-pass Butterworth filter (Bonnechere et al., 2012; González, Hayashibe and Fraisse, 2013; Nixon, Howard and Chen, 2013), Kalman Filter for a linear systems (Machida et al., 2012; Pastor, Hayes and Bamberg, 2012; Bostanci, Kanwal and Clark, 2015). Then, not so frequent were the extensions of the Kalman Filter (KF) for non-linear systems: Extended Kalman Filter (Ratsamee et al., 2012; Shu, Hamano and Angus, 2014) and Unscented Kalman Filter (Wang et al., 2015*b*). Furthermore, as seen in Chapter 2, some researches addressed the outlier problem specifically, by trying to fit probability distributions to the data (Obdržálek et al., 2012) or defining limits on the velocity or amplitude difference (Otte et al., 2016).

This chapter goes into the important concepts aiming at building the background foundations required for the implementation of the system of the present work. In fact, first there will be the explanation of a human kinematic model as a kinematic chain, and also the notation and conventions used. Second, an introduction to the Kalman filter will be explained, as it serves as the basis for the actual algorithm that is going to be used in this work. The unscented Kalman filter, which will be analysed next, uses the same general concepts of the Kalman filter, but circumvents its the main drawback, being devised for non-linear applications. Finally, the actual methods applied to support the processing of the human skeleton data structure will be explained in detail.

4.1 Kinematic model

In order to tackle the problems related to the anatomical structure of the human model, a kinematic filtering approach has been employed. In fact, the skeleton data was modelled as a kinematic chain, with the root in the abdomen, and described by kinematic joints of one, two, or three DOF, corresponding to representations of the anatomical joints and their DOF. Therefore, the human body may be described as,

- the driver (root joint, abdomen) spacial coordinates in a Cartesian coordinate system, the universe coordinate system (in this work, the same as the camera or Kinect coordinate system),
- the orientation of the driver joint,
- the angles formed by each of the bones in the three anatomical planes (frontal/coronal, median/sagittal and transverse/horizontal), according to the their DOF,
- the lengths of each of the bones that connected the joints.

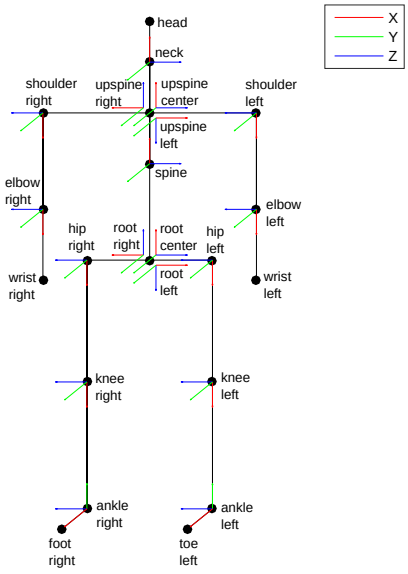


Figure 4.1: Kinematic model with joint descriptions and frames relative to each of the joints according to the implemented kinematic chain model of the human body.

These factors depict the location of the subject in an Euclidean space, considering the position and orientation of the root joint and transformations from the reference frame of one joint to the next, allowing the description of the locations of all the joints. Moreover, by defining the rotations according to the anatomical planes, it was possible to identify rotations in terms of the anatomical movements of adduction and abduction, flexion and extension, and rotation along the axis of the bone: pronation and supination, and medial and lateral rotation.

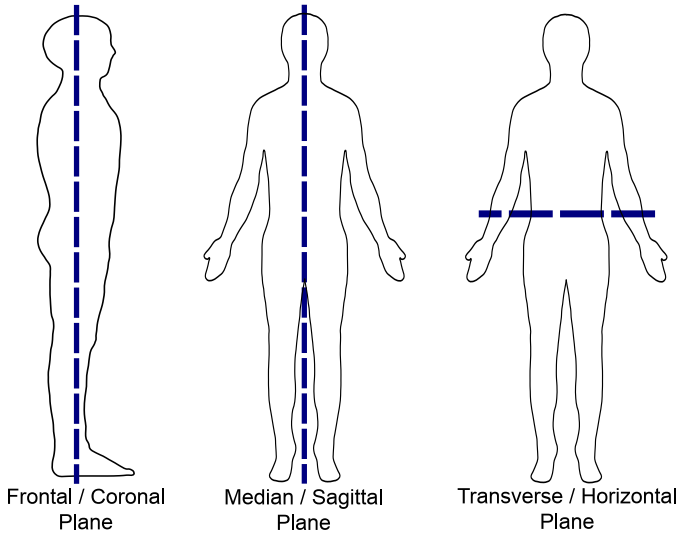


Figure 4.2: Anatomical planes.

Accordingly, the approach of a kinematic filter, taking into consideration the anatomical conditions of the human body, aimed at improving the accuracy of the noisy data from the Kinect sensor by enforcing constraints and correcting “impossible” positions that are observed in the raw data, thus achieving a more robust and accurate model.

In practice, the body of the human subject was modelled as 19 joints and 18 rigid bodies, as seen in Figure 4.1, taking each joint with a frame associated with the next joint in the chain, such that the kinematic chain may be computed iteratively from the root to the extremities. This was done by means of a non-linear filter based on the Kalman Filter (Section 4.2), the Unscented Kalman Filter, taking into account the system as described by the driver, angles, and length parameters, and it will be presented in Section 4.3. Further details about the methodology and implementation of the system will be analysed in Chapter 5.

4.1.1 Euler angles

A rotation matrix is a way of describing a rotation of a frame in the Euclidean space, however, this 3 by 3 matrix is not only quite large and complex to write, but also quite unintuitive. Therefore, there are conventions that use angles which represent rotations along predefined axes. These angles may be used interchangeably with the rotation matrix, resulting in a simplification in representing a rotation, as well as giving some meaning to three values (Craig, 2005). In fact, three angles may be used to describe a rotation of a frame, and, in the present work, the Z-Y-X Euler angle convention was employed due to being one of the most commonly found in works on the subject, as well as for having a special meaning in the studied biomechanical context, which will be described ahead.

Accordingly to the Z-Y-X Euler angle convention, the rotation of a frame occurs successively in the three axes (in the following paragraph the symbol $'$ represents an axis after being rotated), *i.e.* the first rotation of an angle α rotates the frame around the Z axis, then the second rotation takes the Y$'$ axis of the new frame

and performs a rotation around it of magnitude β, and finally the third rotation considers the X$''$ axis of the now current frame and performs a rotation of γ around it. This process may be better understood by looking at the figure below (Figure 4.3).

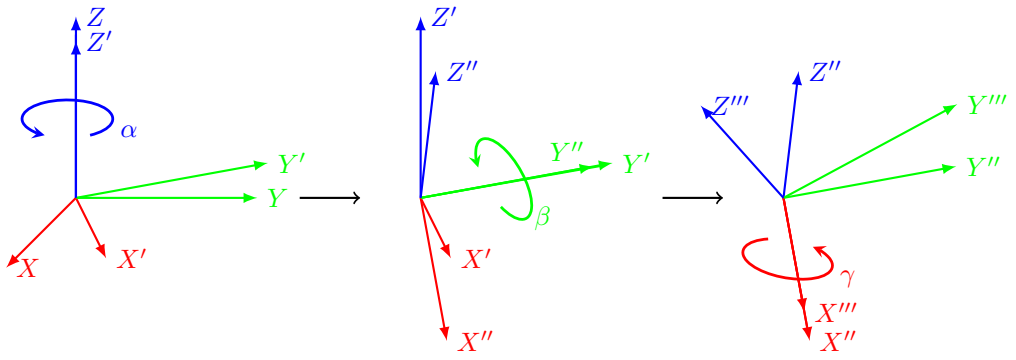


Figure 4.3: Rotations according to the convention of Z-Y-X Euler angles.

Therefore, when using this convention, and taking s and c as abbreviations for the $sine$ and $cosine$ functions, the rotations according to each axis are given by,

$$R_Z(\alpha) = \begin{bmatrix} c\alpha & -s\alpha & 0 \\ s\alpha & c\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.1}$$

$$R_Y(\beta) = \begin{bmatrix} c\beta & 0 & s\beta \\ 0 & 1 & 0 \\ -s\beta & 0 & c\beta \end{bmatrix} \tag{4.2}$$

$$R_X(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\gamma & -s\gamma \\ 0 & s\gamma & c\gamma \end{bmatrix} \tag{4.3}$$

Hence the combination as successive rotations (Equations, 4.1, 4.2, and 4.3) giving the final orientation,

$$R_{Z'Y'X'}(\alpha,\beta,\gamma) = R_Z(\alpha)R_Y(\beta)R_X(\gamma) = \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma \\ -s\beta & c\beta s\gamma & c\beta c\gamma \end{bmatrix} \tag{4.4}$$

One important factor and a motivation for having used this convention (and taking into account that the frame would be aligned with the anatomical planes), was the meaning of the axes and rotations being considered according to the movements in the anatomical planes: the rotation of the Z axis (α) mapped to flexion and extension movements, the Y axis (β) to abduction and adduction movements, and the X axis (γ) to rotations along the axis that has the direction of the bone.

4.1.2 Transformations

The model was defined as a kinematic chain, therefore, from the driver joint, the joints' rotations, and the lengths of the bones, the entire system was characterized. Consequently, to define the positions of each joint in the universe reference frame, a set of transformations was applied. These convert from a Cartesian coordinate system corresponding to a joint frame, successively down the kinematic chain, done from the parent to the next joint, considering its rotation and Euclidean distance to the parent as a translation.

Considering the notation ${}^{A}P_i$ as the position of a point in space P_i represented as a vector relative to the reference frame A. Then, a transformation from the representation in one frame (A) to another (B) may consist in a translation (Figure 4.4) by means of a vector ${}^{A}P_B$. Thus, a vector written in B may be written in A, and *vice versa*, by applying such transformation.

$$ {}^{A}P_i = {}^{B}P_i + {}^{A}P_B \tag{4.5} $$

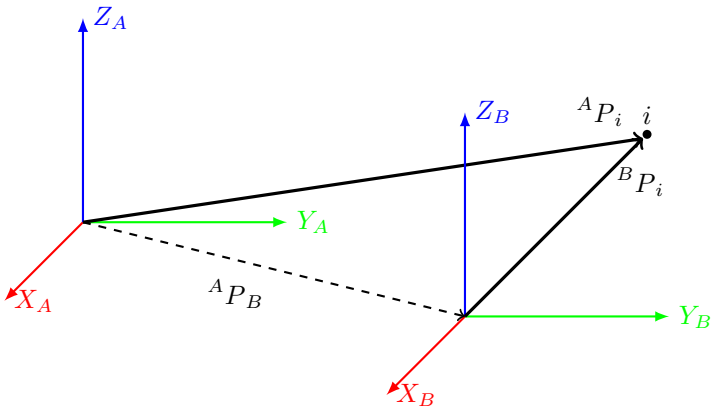


Figure 4.4: Translation transformation.

Nevertheless, the frame in which one desires to describe a vector does not need to have its axes in the same orientation as the other frame. Therefore, a transformation may be described by a rotation (Figure 4.5) ${}^{A}_{B}R$, which describes frame B relative to frame A, *i.e.* consists in the basis of B written in A.

$$ {}^{A}P_i = {}^{A}_{B}R \ {}^{B}P_i \tag{4.6} $$

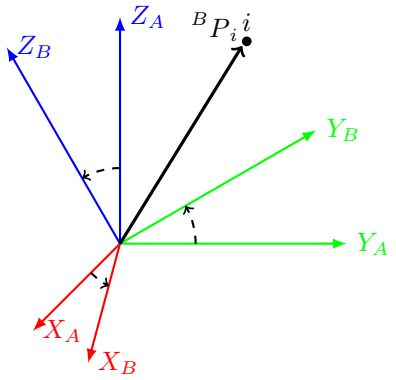


Figure 4.5: Rotation transformation.

Then, a transformation may consist in both a rotation and a translation, as seen in Figure 4.6. So, in order to identify a vector in a desired frame ${}^{A}P_i$, first there should be a rotation of a known frame (B) to the orientation of the desired frame (A) ${}^{A}_{B}R$, followed by a translation corresponding to the vector of the origin of the second frame (B) expressed in relation of the first (A), ${}^{A}P_B$.

$$ {}^{A}P_i = {}^{A}_{B}R \ {}^{B}P_i + {}^{A}P_B \tag{4.7} $$

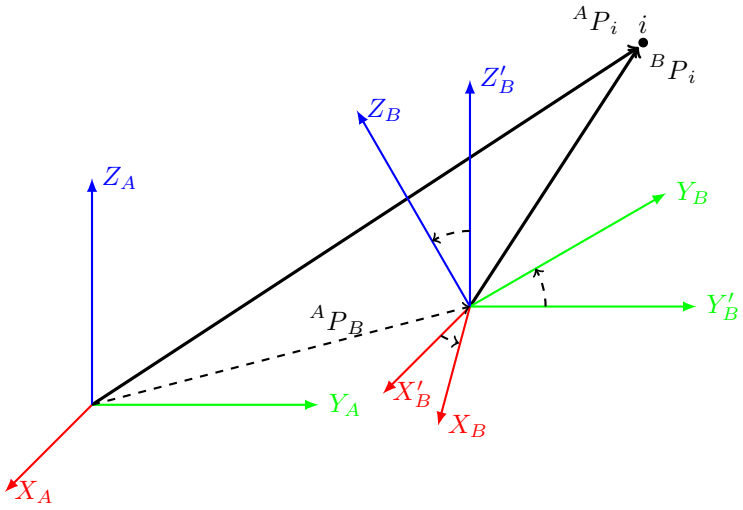


Figure 4.6: Rotation and translation transformation.

Furthermore, a transformation from frame A to frame B of Equation 4.7 may be also described, in a more compact and clear notation as ${}^{A}_{B}T$, the homogeneous transform from frame B to frame A (Craig, 2005), defined as,

$$ {}^{A}_{B}T = \left[\begin{array}{c|c} {}^{A}_{B}R & {}^{A}P_B \\ \hline 0\ 0\ 0 & 1 \end{array}\right] \tag{4.8} $$

And therefore,

$$ {}^{A}P_i = {}^{A}_{B}T\,{}^{B}P_i \Leftrightarrow \tag{4.9} $$

$$ \begin{bmatrix} {}^{A}P_i \\ 1 \end{bmatrix} = \left[\begin{array}{c|c} {}^{A}_{B}R & {}^{A}P_B \\ \hline 0\,0\,0 & 1 \end{array}\right] \begin{bmatrix} {}^{B}P_i \\ 1 \end{bmatrix} \tag{4.10} $$

In addition, transformations may be combined together if a frame is known in relation to another that itself is known in relation to a third (Figure 4.7). Respectively, if ${}^{C}P_i$ is known, as well as ${}^{B}_{C}T$ and ${}^{A}_{B}T$, it is possible to compute the desired ${}^{A}P_i$ as,

$$ {}^{A}P_i = {}^{A}_{B}T\ {}^{B}_{C}T\ {}^{C}P_i \Leftrightarrow \tag{4.11} $$

$$ \begin{bmatrix} {}^{A}P_i \\ 1 \end{bmatrix} = \left[\begin{array}{c|c} {}^{A}_{B}R\ {}^{B}_{C}R & {}^{A}_{B}R\ {}^{B}P_C + {}^{A}P_B \\ \hline 0\,0\,0 & 1 \end{array}\right] \begin{bmatrix} {}^{C}P_i \\ 1 \end{bmatrix} \tag{4.12} $$

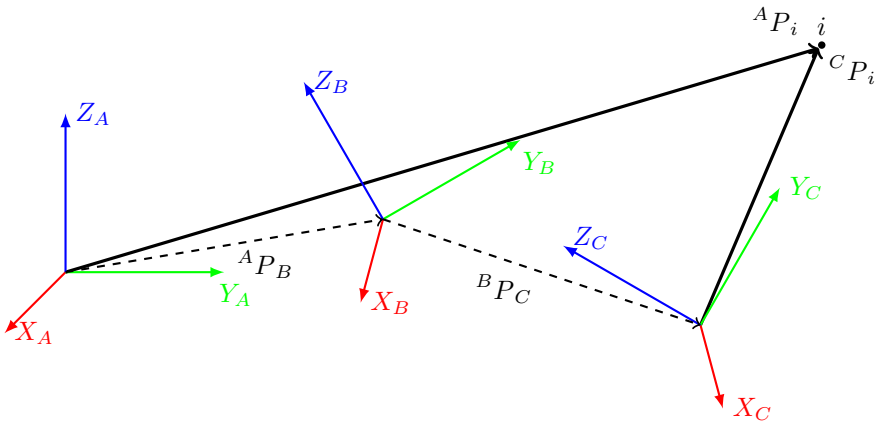


Figure 4.7: Composed transformation.

Finally, as the root is considered the driver, the transformations may be applied in turn, starting from the root to the extremities, considering the rotations and translations when obtaining the coordinates of the joint which is next in the kinematic chain, until an extremity is reached. Accordingly, considering only the coordinates of the driver and the lengths and angles of rotation between each of the joints, it is possible to obtain any joint's coordinates in the universe reference frame. The notation that has been used in this section might also be interpreted as the transformation of the formulation of the joint i in frame B to expressing it in frame A.

4.2 Kalman filter

The data observed from the Kinect sensor, may be considered as the true positions of the joints of the subject in the universe reference frame, plus noise. Therefore, due to the structure of the problem, a KF approach was experimented since the movements of the model may be modelled by known expressions, supporting the ability to approximate the true position by filtering the noise using this algorithmic approach.

The Kalman Filter (Kalman et al., 1960) is a method which, based on taking observed measurements in time and considering that the observations have noise, aims at generating estimates that approximate the real, unknown, values. This method is an optimal estimator for linear systems, as it minimizes the mean squared error of the estimated values (Simon, 2006). In addition, it works by producing an estimate of the state of a linear system, *i.e.* internal values in accordance to a linear dynamic system model (the state) and its uncertainty (covariance matrix). Then, by comparison with the measured values, adjusts the prediction with a weighted average, *i.e.* according to weights calculated from the covariance of the state - the Kalman gain (Kalman et al., 1960) - which is larger for values with less uncertainty and smaller for values that are more uncertain (Grewal, 2011). The process works recursively, depending only on the previous state and on its covariance, being repeated at each time step by taking a new estimation of the state and covariance and updating them according to the weighed average with the noisy measurements, thus it is possible to be run in real-time applications. The Kalman gain, computed from the covariance, tunes the weights that are given to the predictions and measurements. As a result, a lower gain implies that the system will trust its predictions more than the noisy measurements, hence having smoother changes despite large variations in the measured values. In opposition, for a higher gain, the system will follow the measured values closer, being more responsive to the observed quantities.

4.2.1 Kalman filter formulation

The KF consists mainly of two steps, first the prediction step, where a prediction by a linear transformation is made, based on the state and its covariance at the previous instant, and taking into consideration the noise model of the state-transition model. Second, the update step, where the Kalman gain is computed. In this step, the prediction is corrected in agreement with its magnitude and an observation model, taking into account the observed values and their noise model. At last, the final corrected values for the state and covariance for the current time step are worked out. The complete and detailed formulation for the KF may be found in (Kalman et al., 1960; Bishop and Welch, 2001).

In short, based on the notation of the cited works above, the KF aims at estimating a state $\mathbf{x}_i$, therefore, it assumes that the observed data is generated over time (observed at each time step i) from the immediately preceding instant, by means of a linear state-transition model $\mathbf{F}_i$, contaminated with process noise $\mathbf{w}_i$. This is thus described by,

$$\mathbf{x}_i = \mathbf{F}_i \mathbf{x}_{i-1} + \mathbf{w}_i \tag{4.13}$$

where the process noise is assumed to be a multivariate Gaussian (of dimension L, the dimension of the state) with mean $\mathbf{0}$ and covariance $\mathbf{Q}_i$,

$$\mathbf{w}_i \sim \mathcal{N}_L(\mathbf{0},\ \mathbf{Q}_i) \tag{4.14}$$

Regarding the measurements $\mathbf{z}_i$, the KF assumes that they come from the state $\mathbf{x}_i$ through an observation model $\mathbf{H}_i$, also tainted by observation noise $\mathbf{v}_i$. Hence,

$$\mathbf{z}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{v}_i \tag{4.15}$$

where the observation noise is also considered to be multivariate Gaussian with mean $\mathbf{0}$ but covariance $\mathbf{R}_i$, of dimension M (of the observation model),

$$\mathbf{v}_i \sim \mathcal{N}_M(\mathbf{0},\ \mathbf{R}_i) \tag{4.16}$$

Then, using these assumptions, it is possible to create an algorithm that performs the desired filtering operation. Hence, the state of the filter is defined by a state estimate (estimates were represented by a hat $\hat{}$) $\hat{\mathbf{x}}$ and a state covariance representing its uncertainty $\mathbf{P}$. These follow the prediction-correction cycle (Figure 4.8) stating with a given estimate ($\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$). As the predictions (represented by a superscript minus $^-$, as the *a priori* estimate) are given by, with ^{T} representing the transpose of a matrix,

$$\hat{\mathbf{x}}_i^- = \mathbf{F}_i\hat{\mathbf{x}}_{i-1} \tag{4.17}$$

$$\mathbf{P}_i^- = \mathbf{F}_i\mathbf{P}_{i-1}\mathbf{F}_i^T + \mathbf{Q}_i \tag{4.18}$$

Then, and taking the following expression (Equation 4.19) for the Kalman gain (the parameter that denotes how much weight is given to the state prediction and how much is given to the observations) as being the optimal Kalman gain (Kalman et al., 1960), the correction follows as, with $\mathbb{I}$ representing the identity matrix,

$$\mathbf{K}_i = \mathbf{P}_i^-\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T + \mathbf{R}_i)^{-1} \tag{4.19}$$

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i(\mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_i^-) \tag{4.20}$$

$$\mathbf{P}_i = (\mathbb{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i^- \tag{4.21}$$

A graphical scheme of the program is presented in Figure 4.8 and the pseudocode for implementing the KF may be seen in Program 4.1.

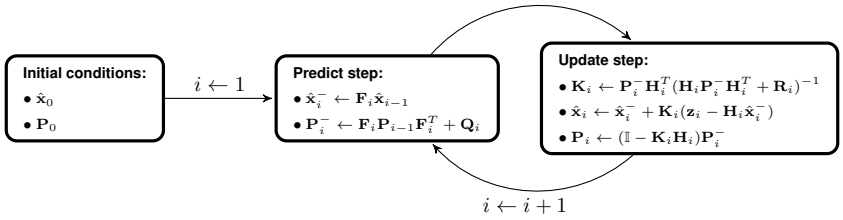


Figure 4.8: Kalman filter steps and equations.

Table 4.1: Variables of Kalman filter program (Program 4.1).

SCOPE	NAME	TYPE	DESCRIPTION
input	$data$	double matrix	input data
	$\mathbf{F}$	double matrix	state-transition model*
	$\mathbf{H}$	double matrix	observation model*
	$\mathbf{Q}$	double matrix	covariance of state-transition model*
	$\mathbf{R}$	double matrix	covariance of observation model*
	$\hat{\mathbf{x}}_0$	double array	initial state
	$\mathbf{P}_0$	double matrix	initial covariance of state
output	$filt$	double matrix	filtered data
internal	$\hat{\mathbf{x}}^-$	double array	predicted state
	$\mathbf{P}^-$	double matrix	predicted covariance matrix
	$\hat{\mathbf{x}}$	double array	updated state
	$\mathbf{P}$	double matrix	updated covariance matrix
	$\mathbf{K}$	double matrix	Kalman gain

*considered constant through time in the program

Program 4.1: Kalman filter.

```
1  input: double[][] data, F, H, Q, R
2  double[] x̂_0, P_0
3  output: double[][] filt
4  begin
5    foreach i in 1...length(data)
6      // prediction step:
7      x̂_i^- ← F x̂_{i-1}
8      P_i^- ← F P_{i-1} F^T + Q
9      // update step:
10     K_i ← P_i^- H^T (H P_i^- H^T + R)^{-1}
11     x̂_i ← x̂_i^- + K_i (data_i − H x̂_i^-)
12     P_i ← (I − K_i H) P_i^-
13
14     filt_i ← x̂_i
15   end
16   return filt
17 end
```

4.3 Unscented Kalman filter

One big drawback of the Kalman Filter is the linear assumption. It does only work if the transition of the system in time is linear, as well as the observational model. In the scope of this project, the created model is non-linear, so, an alternative to the KF must be applied to circumvent this limitation. There are several methods that approximate the KF in a non-linear application, the most well known and used is the Extended Kalman Filter (EKF). This method considers a linearisation of the system by a Taylor series expansion (often to the first order) around the considered point and substituting the Jacobian matrices in place of the linear ones from KF (Ljung, 1979). However, this approach requires the computation of the derivatives of the state-transition and observation models' non-linear functions, which reveal themselves as being quite complex to implement for large systems, as well as being only reliable for systems that are almost linear, with violations of the local linearity producing an unstable filter (Julier and Uhlmann, 2004).

Therefore, another useful approach for this work is seized, the Unscented Kalman Filter (UKF) (Julier and Uhlmann, 1997, 2004). Indeed, the UKF has in its basis the Unscented Transform (UT), which stipulates a mechanism of transforming mean and covariance information by virtue of an approximation of a probability distribution instead of an approximation of a non-linear function. Hence, a set of "sigma points" is chosen specifically for their mean and covariance to be $\bar{\mathbf{x}}$ and $\mathbf{P}$, and the non-linear function is evaluated at these points. The statistics from the transformed points may then be recovered, forming an estimate of the non-linearly transformed mean and covariance. This results in a method which yields the ability to propagate second-order properties of a distribution (due to the deterministically chosen sigma points), in addition to creating an interpretation that is more general than a probability distribution (Julier and Uhlmann, 2004). In the following paragraphs, the practical component is described in summary, yet a more complete and detailed formulation and implementation may be seen in (Julier and Uhlmann, 2004; Wan and Van Der Merwe, 2000; Simon, 2006).

First, the UKF introduces some parameters that must be taken into consideration, these are related with the computation of the sigma points, controlling their spread around the mean and the prior knowledge about the distribution. In fact, the UKF takes three parameters that did not exist in the KF:

- α controls how spread the sigma points are around the mean of the state $\bar{\mathbf{x}}$,
- β carries information and allows to introduce prior knowledge of the distribution of the state $\mathbf{x}$,
- κ is a secondary scaling parameter, for fine tuning.

The usual values for these parameters may be seen in Table 4.2. Besides these three parameters, there is also the parameter λ, which depends on α, κ and L, the dimension of the state.

$$\lambda = \alpha^2(L + \kappa) - L \tag{4.22}$$

Table 4.2: Unscented Kalman filter sigma points' usual values (Wan and Van Der Merwe, 2000).

PARAMETER	USUAL VALUE
α	$1 \cdot 10^{-3}$
β	$2^{\star}$
κ	0

*optimal value for Gaussian distributions

With regard to the computation of the sigma points $\mathcal{X}$, these consist on $2L + 1$ points, where L is the dimension of the state $\mathbf{x}$, considering the mean of the state $\bar{\mathbf{x}}$ and its covariance $\mathbf{P}$, as well as the algorithm parameters described above. In this context, the index $_i$ represents the i-th entry in the associated vector, and $(\sqrt{(L+\lambda)\mathbf{P}})_i$ is the i-th row of the matrix. Then,

$$\mathcal{X}_0 = \bar{\mathbf{x}} \tag{4.23}$$

$$\mathcal{X}_i = \bar{\mathbf{x}} + (\sqrt{(L+\lambda)\mathbf{P}})_i, \quad i = 1, \ldots, L \tag{4.24}$$

$$\mathcal{X}_i = \bar{\mathbf{x}} - (\sqrt{(L+\lambda)\mathbf{P}})_{i-L}, \quad i = L+1, \ldots, 2L \tag{4.25}$$

where the square root of the positive definite matrix ($\mathbf{P}$, a covariance matrix) is defined as,

$$\mathbf{A} = \mathbf{P}^{1/2} \Leftrightarrow \tag{4.26}$$

$$\mathbf{P} = \mathbf{A}\mathbf{A}^T \tag{4.27}$$

The sigma points have a weight associated with them, which is used to reconstruct the statistics from the transformed points. $\mathbf{W}^{(m)}$ are the weights used to reconstruct the transformed mean, and $\mathbf{W}^{(c)}$ are the weights used to reconstruct the transformed covariance of the system. Their values are given by,

$$\mathbf{W}_0^{(m)} = \frac{\lambda}{L+\lambda} \tag{4.28}$$

$$\mathbf{W}_0^{(c)} = \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta) \tag{4.29}$$

$$\mathbf{W}_i^{(m)} = \mathbf{W}_i^{(c)} = \frac{1}{2(L+\lambda)}, \quad i = 1, \ldots, 2L \tag{4.30}$$

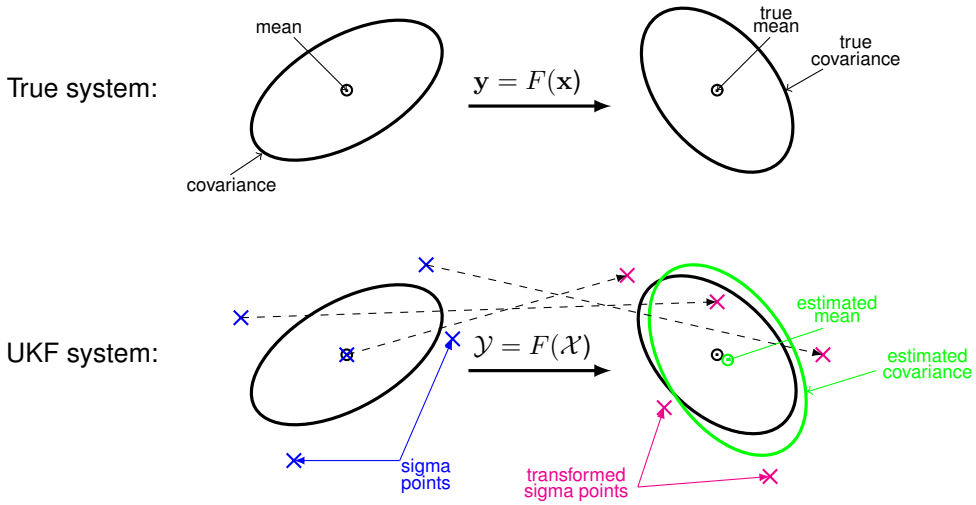


Figure 4.9: True and UKF systems' probability density functions, as schematic demonstration of the transformation of the sigma points and estimation of the new mean and covariance.

With respect to the iterative component, the UKF works mostly in the same way as the normal KF, taking the unscented transform for approximating the non-linear functions. Therefore, the same system is reasoned (with the non-linearities), and considering that the noise is additive (Simon, 2006),

$$\mathbf{x}_i = F(\mathbf{x}_{i-1}) + \mathbf{w}_i \tag{4.31}$$

$$\mathbf{y}_i = H(\mathbf{x}_i) + \mathbf{v}_i \tag{4.32}$$

where,

$$\mathbf{w}_i \sim \mathcal{N}_L(\mathbf{0}, \mathbf{Q}_i) \tag{4.33}$$

$$\mathbf{v}_i \sim \mathcal{N}_M(\mathbf{0}, \mathbf{R}_i) \tag{4.34}$$

Therefore, the explained parameters of the algorithm $(\alpha, \beta, \kappa, \lambda, \mathbf{W})$ are considered, and, at each time step $(_i)$, the sigma points are computed (Equations 4.23 to 4.25) and are propagated through the non-linear state-transition function,

$$\mathcal{X}_i^* = F(\mathcal{X}_i) \tag{4.35}$$

Next, the predicted state and predicted system covariance (again indicated with a superscript minus $^-$) is computed taking the transformed sigma points and the weights associated to each, as well as the noise model for this transition,

$$\hat{\mathbf{x}}_i^- = \sum_{j=0}^{2L} \mathbf{W}_j^{(m)} \mathcal{X}_{i,j}^* \tag{4.36}$$

$$\mathbf{P}_i^- = \sum_{j=0}^{2L} \mathbf{W}_j^{(c)} \left(\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-\right) \left(\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-\right)^T + \mathbf{Q}_i \tag{4.37}$$

Also, the transformed sigma points are propagated through the non-linear observation model function, and the average is reconstructed considering the sigma points' weights as,

$$\mathcal{Y}_i = H(\mathcal{X}_i^*) \tag{4.38}$$

$$\hat{\mathbf{y}}_i = \sum_{j=0}^{2L} \mathbf{W}_j^{(c)} \mathcal{Y}_{i,j} \tag{4.39}$$

Regarding the measurements, two covariance matrices are computed, one considering only the values obtained through the observation model function and considering the observation noise model, and another, a cross-covariance between the sigma points from the state-transition function and the observations function,

$$\mathbf{P}_{y_i} = \sum_{j=0}^{2L} W_j^{(c)} \left(\mathcal{Y}_{i,j} - \hat{\mathbf{y}}_i\right) \left(\mathcal{Y}_{i,j} - \hat{\mathbf{y}}_i\right)^T + \mathbf{R}_i \tag{4.40}$$

$$\mathbf{P}_{(x,y)_i} = \sum_{j=0}^{2L} W_j^{(c)} \left(\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-\right) \left(\mathcal{Y}_{i,j} - \hat{\mathbf{y}}_i\right)^T \tag{4.41}$$

The Kalman gain is then obtained from these two covariances as,

$$\mathbf{K}_i = \mathbf{P}_{(x,y)_i} \mathbf{P}_{y_i}^{-1} \tag{4.42}$$

Ultimately, the final state for the current time step is corrected according to the Kalman gain and the error between the observed data and the one predicted, and the final covariance from the gain and $\mathbf{P}_{y_i}$,

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i(\mathbf{y}_i - \hat{\mathbf{y}}_i) \tag{4.43}$$

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{K}_i \mathbf{P}_{y_i} \mathbf{K}_i^T \tag{4.44}$$

At last, the UKF presents various advantages in relation to the EKF, some of the most relevant being (Wan and Van Der Merwe, 2000; Julier and Uhlmann, 2004):

- The EKF expects the Jacobian matrix of the system-transition and observation function, however, for a large system the computation of the Jacobian is quite painful and error-prone on the algebraic process, being difficult to implement and tune.
- EKF's linearisation operation is only valid if the system is approximately linear at the point being anal-

ysed, not being reliable otherwise.

- In opposition to the EKF first-order accuracy, the UKF's sample points describe the true mean and true covariance of the Gaussian random variable, and after propagated through the non-linear system, these capture the statistics to the third-order Taylor series expansion.
- The computation complexity of the UKF is of the same order of magnitude of the EKF ($O(N^3)$), accomplishing better accuracy for a comparable complexity, and being sufficiently accurate for highly non-linear applications.
- Wan and Van Der Merwe (2000) evidence a consistent improved accuracy for the UKF in relation to the EKF in multiple domains, such as state-estimation, dual estimation, and parameter estimation.

In program 4.2, the pseudocode for the implementation of the UKF is presented, with the description of each of the variables defined in Tables 4.1 and 4.3

Table 4.3: Variables of unscented Kalman filter program (Program 4.2), complementary of Table 4.1.

SCOPE	NAME	TYPE	DESCRIPTION
	$F()$	function handler	state-transition model
input	$H()$	function handler	observation model
	α, β, κ	double	parameters of the UKF
	L	integer	dimension of the state
	λ	double	parameter of UKF
	$W^{(x)}$	double matrix	sigma point weights (x: mean (m), covariance (c))
	$\mathcal{X}$	double matrix	sigma points
internal	$\mathcal{X}^*$	double matrix	sigma points propagaded through nonlinear function $F()$
	$\mathcal{Y}$	double matrix	sigma points propagaded through nonlinear function $H()$
	$\hat{\mathbf{y}}$	double array	predicted observation
	$\mathbf{P}_y$	double matrix	measurement covariance after nonlinear functions
	$\mathbf{P}_{(x,y)}$	double matrix	cross-covariance of state and measurement

Program 4.2: Unscented Kalman filter.

```
1  input: double[][] data, Q, R
2         function F(), H()
3         double[] x̂_0, P_0
4         double α, β, κ
5  output: double[][] filt
6  begin
7    // compute weights for sigma points:
8    L ← length(x̂_0)
9    λ ← α²(L + κ) − L
10   W^(m) ← [ λ/(L+λ)   (1/(2(L+λ)))_i ], i ∈ {1, 2, ..., 2L}
11   W^(c) ← [ λ/(L+λ) + 1 − α² + β   (1/(2(L+λ)))_i ], i ∈ {1, 2, ..., 2L}
12   foreach i in 1 ... length(data)
13     // compute sigma points:
14     X_i ← [ x̂_{i−1}   x̂_{i−1} + (√((L + λ)P_{i−1}))_j   x̂_i − (√((L + λ)P_{i−1}))_j ], j ∈ {1, 2, ..., L}
15     // time update:
16     X*_i ← F(X_i)
17     x̂⁻_i ← Σ_{j=0}^{2L} W_j^(m) X*_{i,j}
18     P⁻_i ← Σ_{j=0}^{2L} W_j^(c) (X*_{i,j} − x̂⁻_i)(X*_{i,j} − x̂⁻_i)^T + Q
19     Y_i ← H(X*_i)
20     ŷ_i ← Σ_{j=0}^{2L} W_j^(m) Y_{i,j}
21     // measurement update:
22     P_{y_i} ← Σ_{j=0}^{2L} W_j^(c) (Y_{i,j} − ŷ_i)(Y_{i,j} − ŷ_i)^T + R
23     P_{(x,y)_i} ← Σ_{j=0}^{2L} W_j^(c) (X*_{i,j} − x̂⁻_i)(Y_{i,j} − ŷ_i)^T
24     K_i ← P_{(x,y)_i} P^{−1}_{y_i}
25     x̂_i ← x̂⁻_i + K_i(data_i − ŷ_i)
26     P_i ← P⁻_i − K_i P_{y_i} K_i^T
27
28     filt_i ← x̂_i
29   end
30   return filt
31 end
```

4.4 The Human Kinematic UKF

The processing of the data retrieved from the Kinect sensor was focused on achieving better accuracy by means of attempting to automatically correct the errors which expose themselves as non-physiological dispositions of the human model, *i.a.* variable lengths of the bones, or joint constraints violated. In order to do this, an approach using an UKF was employed on a modelling of the human body as a kinematic chain, which will be illustrated in this section.

In the followed approach, two kinematic chains were considered, one modelling the upper half of the body, and the other modelling the lower half, as seen in Figure 4.10. In the end, both chains may be grouped again, by considering the filtered positions of both models. In fact, by considering these two systems as being independent, the complexity of the model would be reduced, improving the computation time, and parallel processing of both would be possible as there are no interdependencies. This consideration is of

special relevance if the system ought to be employed in a real-time application, as the latency of processing one frame would be substantially lower. However, in the end, an error would be introduced, as the positions of the roots of both models might not match, but this should not be a problem, as upper or lower movement features usually do not require this aspect, and so it was not addressed. Notwithstanding, computing the mean between both positions would be an easy way to correct this issue, if required.

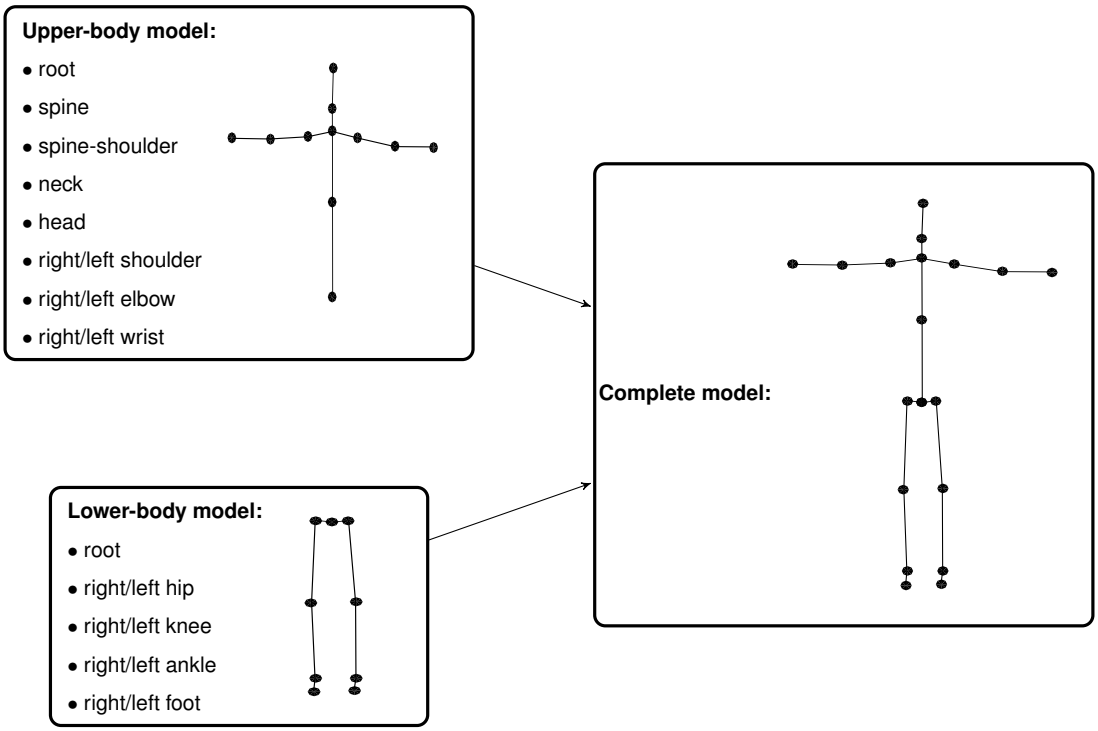


Figure 4.10: Joints considered in the upper, and lower-body kinematic chains.

Furthermore, the kinematic chains were modelled as going from the root to the extremities. Thus, from the position of the root, successive transformations were applied using rotations, considering one (elbow, knee), two (ankle), or three (spine, spine-shoulder, neck, shoulder, hip) angles depending on the DOF attributed to each joint, and considering the lengths of the links as the translation factors. Therefore, there was the need to compute a set of angles for each joint, so the Euler angle convention has been considered, with each angle aligned to a corresponding anatomical plane, as seen in Figure 4.1, so meaningful information on the angular movements being performed could be extracted. As not all joints in the body allow movement in all directions, depending on which rotations are possible by a certain joint, a number of DOF were provided for it, and the same number of angles was computed.

In fact, a transformation works from the parent joint to the following one, by considering the joint's position as $(x, y, z) = (L, 0, 0)$ in the parent's reference frame, where L is the length of the bone connecting the considered joint and its parent. Hence, the reference frame of the parent joint is considered always to have its X axis aligned with the bone connecting it to the successive joint in the kinematic chain. Therefore, the joint's coordinates in the parent's reference frame is a translation in the X axis of value given by the parameter

L. These link length parameters should be computed right before the capturing taking place, with a person in a static position, and during a period of time that would allow for the averaging throughout that interval, reducing the possibility of obtaining wrong measurements in this step. Nevertheless, it is clear that the joints are not all aligned in the same direction, therefore, rotations of each of the reference frames must be taken into account. These rotations were represented in the state as Euler angles, and represented the rotation of the parent's frame, such that the X axis of the parent was given the direction of the joint in consideration. So, the description of a joint in its parent reference frame was set as the rotation of the parent's reference frame such that the X axis aligned with the link direction, and a following translation of norm equal to the link length. Moreover, these transformation could be composed (Equation 4.11), *i.e.* chained together, in order to obtain the description of joints deeper in the kinematic chain in reference frames corresponding to previous joints, and in the end, in the Kinect's reference frame.

These models were implemented using a kinematic filter based on an UKF, which worked by making a prediction on the state for the following frame, taking into account the previous state and a given state-transition function which acted on that state, and combining it with the observations obtained through the Kinect using a given observation model, according to a scaling factor, the Kalman gain (Equation 4.42), as conceptually described on Section 4.3. This was done using a constant velocity model, meaning that the variations were introduced through the acceleration component. On one hand, the state considered by the filter was composed by the position and the angles for each joint in relation to the direction of the next, and the link lengths, as described by the vector in Subsection 4.4.1. On the other hand, the observation vector corresponded to the joints, as 3D coordinates given by the Kinect SDK, as seen in Subsection 4.4.2. Likewise, the state-transition function, which made predictions based on the previous state, and the observation function, which transformed the state into Cartesian coordinates into the same coordinate system as the Kinect's, are detailed in Subsections 4.4.3 and 4.4.4, respectively. Also, the square root of the covariance matrix was computed using the Cholesky decomposition, a stable and efficient method often employed in the UKF (Julier and Uhlmann, 2004).

In addition, the SDK algorithm did not enforce any physiological constraint on the skeleton model, and it was asserted that there were variations of the lengths of the links, which was definitely not a physiological phenomenon. Hence, its acquisition was clearly wrong, and the filter targeted the issue, trying to maintain the lengths constant throughout the acquisition, as seen in the section relative to the results, Section 5.3.

In the following subsections, the specific components of the implemented UKF are described.

4.4.1 State vector

The state vector $\mathbf{x}$ described the system as the two kinematic chains, modelling each one by the positions and velocities of the root in the universe reference frame, and the angles and angular velocities, and link lengths needed for the description of the skeleton analysed above.

Also, as explained, a model for the lower-body (Equation 4.45), and a model for the upper-body (Equation

4.46) were implemented.

Lower-body model:

$$\begin{aligned}\mathbf{x} = [&x_{root}, y_{root}, z_{root}, vx_{root}, vy_{root}, vz_{root}, \theta x_{root}, \theta y_{root}, \theta z_{root}, \omega x_{root}, \omega y_{root}, \omega z_{root}, \theta x_{r-hip}, \dots \\ &\theta y_{r-hip}, \theta z_{r-hip}, \omega x_{r-hip}, \omega y_{r-hip}, \omega z_{r-hip}, \theta z_{r-knee}, \omega z_{r-knee}, \theta y_{r-ankle}, \theta z_{r-ankle}, \omega y_{r-ankle}, \dots \\ &\omega z_{r-ankle}, \theta x_{l-hip}, \theta y_{l-hip}, \theta z_{l-hip}, \omega x_{l-hip}, \omega y_{l-hip}, \omega z_{l-hip}, \theta z_{l-knee}, \omega z_{l-knee}, \theta y_{l-ankle}, \dots \\ &\theta z_{l-ankle}, \omega y_{l-ankle}, \omega z_{l-ankle}, L_{r-root-hip}, L_{r-hip-knee}, L_{r-knee-ankle}, L_{r-ankle-foot}, \dots \\ &L_{l-root-hip}, L_{l-hip-knee}, L_{l-knee-ankle}, L_{l-ankle-foot}]^T\end{aligned} \tag{4.45}$$

Upper-body model:

$$\begin{aligned}\mathbf{x} = [&x_{root}, y_{root}, z_{root}, vx_{root}, vy_{root} vz_{root}, \theta x_{root}, \theta y_{root}, \theta z_{root}, \omega x_{root}, \omega y_{root}, \omega z_{root}, \theta x_{spine}, \dots \\ &\theta y_{spine}, \theta z_{spine}, \omega x_{spine}, \omega y_{spine}, \omega z_{spine}, \theta x_{l-shoulder-c}, \theta y_{l-shoulder-c}, \theta z_{l-shoulder-c}, \dots \\ &\omega x_{l-shoulder-c}, \omega y_{l-shoulder-c}, \omega z_{l-shoulder-c}, \theta x_{l-shoulder}, \theta y_{l-shoulder}, \theta z_{l-shoulder}, \dots \\ &\omega x_{l-shoulder}, \omega y_{l-shoulder}, \omega z_{l-shoulder}, \theta z_{l-elbow}, \omega z_{l-elbow}, \theta x_{r-shoulder-c}, \theta y_{r-shoulder-c}, \dots \\ &\theta z_{r-shoulder-c}, \omega x_{r-shoulder-c}, \omega y_{r-shoulder-c}, \omega z_{r-shoulder-c}, \theta x_{r-shoulder}, \theta y_{r-shoulder}, \dots \\ &\theta z_{r-shoulder}, \omega x_{r-shoulder}, \omega y_{r-shoulder}, \omega z_{r-shoulder}, \theta z_{r-elbow}, \omega z_{r-elbow}, \theta x_{up-shoulder-c}, \dots \\ &\theta y_{up-shoulder-c}, \theta z_{up-shoulder-c}, \omega x_{up-shoulder-c}, \omega y_{up-shoulder-c}, \omega z_{up-shoulder-c}, \dots \\ &\theta x_{neck}, \theta y_{neck}, \theta z_{neck}, \omega x_{neck}, \omega y_{neck}, \omega z_{neck}, L_{root-spine}, L_{spine-shouldercenter}, \dots \\ &L_{l-shouldercenter-shoulder}, L_{l-shoulder-elbow}, L_{l-elbow-wrist}, L_{r-shouldercenter-shoulder}, \dots \\ &L_{r-shoulder-elbow}, L_{r-elbow-wrist}, L_{shouldercenter-neck}, L_{neck-head}]^T\end{aligned} \tag{4.46}$$

4.4.2 Observation vector

The data provided by the Kinect SDK described the human model as a set of joints in a skeleton structure, which included the position of each joint as a set of coordinates in a reference frame, the camera reference frame (which in the implementation was also the universe reference frame). Thus, each joint was described by a vector of three coordinates, and again, a model for the lower-body (Equation 4.47), and a model for the upper-body (Equation 4.48) were defined.

It should be noted, by comparison with Figure 5.1 (presented ahead in the section related to the data acquisition, Section 5.1), that the positions for the hand tip and thumb were discarded, as the study of the fitness tests would not require gesture detection or opening/closing of the hand information, and besides this, another reason for not being included was the reduction of the complexity of the computations.

Finally, regarding the notation used in this subsection, $\mathbf{y} = [\mathbf{x}]$ must be seen as describing the three components of $\mathbf{x}$ as being components of $\mathbf{y}$, *i.e.* $\mathbf{y} = [x_1, x_2, x_3]$.

Lower-body model:

$$\mathbf{y} = [\mathbf{x}_{root}, \mathbf{x}_{l-hip}, \mathbf{x}_{l-knee}, \mathbf{x}_{l-ankle}, \mathbf{x}_{l-foot}\mathbf{x}_{r-hip}, \mathbf{x}_{r-knee}, \mathbf{x}_{r-ankle}, \mathbf{x}_{r-foot}]^T \tag{4.47}$$

Upper-body model:

$$\mathbf{y} = [\mathbf{x}_{root}, \mathbf{x}_{spine}, \mathbf{x}_{c-shoulder}, \mathbf{x}_{neck}, \mathbf{x}_{head}, \mathbf{x}_{l-shoulder}, \mathbf{x}_{l-elbow}, \mathbf{x}_{l-wrist}, \mathbf{x}_{r-shoulder}, \mathbf{x}_{r-elbow}, \mathbf{x}_{r-wrist}]^T \tag{4.48}$$

4.4.3 State-transition function

The state-transition function F (Equation 4.31) delineated the evolution of the state $\mathbf{x}$ (Equations 4.45 and 4.46). Hence, this evolution is represented by a linear system of the physical equations of motion considering each time step to be $\Delta t = \frac{1}{f_{samp}}$ (again, $f_{samp} = 30$ Hz), which depending on the number of DOF of each joint would be computed as follows:

- For joints with 1 DOF,

$$\begin{bmatrix} x_i \\ vx_i \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{i-1} \\ vx_{i-1} \end{bmatrix} \tag{4.49}$$

- For joints with 2 DOF,

$$\begin{bmatrix} x_i \\ y_i \\ vx_i \\ vy_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \\ vx_{i-1} \\ vy_{i-1} \end{bmatrix} \tag{4.50}$$

- For joints with 3 DOF,

$$\begin{bmatrix} x_i \\ y_i \\ z_i \\ vx_i \\ vy_i \\ vz_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \\ vx_{i-1} \\ vy_{i-1} \\ vz_{i-1} \end{bmatrix} \tag{4.51}$$

Then, for the link lengths, these were defined as being constant throughout the entire acquisition, therefore, they were described by,

$$L_i = L_{i-1} \tag{4.52}$$

4.4.4 Observation function

The observation function H (Equation 4.32) was set using an accumulation of composed transformations (Equation 4.11), which were applied starting at the root and evolving outward, applying successive rotations and translations according to the angles predicted in the state $\mathbf{x}$, to then compute the values of the positions of each joint in the universe reference frame, described in the observation vector (Equations 4.47 and 4.48). The resulting values would then be able to be compared to the ones obtained directly though the Kinect sensor, as both would be 3D coordinates of the joints in the universe reference frame.

Since describing the process for the entire system would be too cumbersome due to the large size, an example using the model's leg is presented below in order to provide a more intuitive view of the process of computing the chain of composed transformations:

1. The lengths of each link were parameters known from the calibration phase: distance from root to hip, from hip to knee, from knee to ankle, and from ankle to foot.
2. The root (r) was considered the driver, and was modelled by its Cartesian coordinates (x, y, z) and its orientation in Euler angles (α, β, γ), both in relation to the universe reference frame $({}^{W})$.

$$ {}^{W}P_r = (x, y, z) \tag{4.53} $$

$$ {}^{W}_{r}R = R_{Z'Y'X'}(\alpha, \beta, \gamma) \tag{4.54} $$

3. The hip frame in the universe reference frame (${}^{W}P_h$) was given by the transformation ${}^{W}_{r}T$ applied to ${}^{r}P_h$ of the known vector of the hip in the root frame,

$$ {}^{W}P_h = {}^{W}_{r}T \ {}^{r}P_h = {}^{W}_{r}R \ {}^{r}P_h + {}^{W}P_r \tag{4.55} $$

4. The knee frame in the universe reference frame (${}^{W}P_k$) was given by composition from the transformations of known vector of the knee in the hip frame, from the hip frame to the root frame, and then from the root frame to the universe reference frame.

$$ {}^{W}P_k = {}^{W}_{h}T \ {}^{h}P_k = {}^{W}_{r}T \ {}^{r}_{h}T \ {}^{h}P_k \tag{4.56} $$

5. To obtain the ankle frame (${}^{W}P_a$, Equation 4.57) and the foot frame in relation to the universe reference frame (${}^{W}P_f$, Equation 4.58), the same rules as before were applied,

$$ {}^{W}P_a = {}^{W}_{k}T \ {}^{k}P_a = {}^{W}_{r}T \ {}^{r}_{h}T \ {}^{h}_{k}T \ {}^{k}P_a \tag{4.57} $$

$$ {}^{W}P_f = {}^{W}_{a}T \ {}^{a}P_f = {}^{W}_{r}T \ {}^{r}_{h}T \ {}^{h}_{k}T \ {}^{k}_{a}T \ {}^{a}P_f \tag{4.58} $$

Chapter 5: Experimental results

In this chapter the methods described in Chapter 4 are portrayed in a practical implementation: the pre-processing of data for fitness assessment. The Kinect v2 datasets acquired by Bernardino et al. (2016) consist on functional fitness tests performed by young and elderly participants. The filter parameters were selected in order to improve the accuracy and biomechanical compatibility of the skeleton data provided by the Kinect SDK.

The treatment of the data was performed using the MATLAB (version R2015b) software suite, with the skeleton data being made available in the MATLAB MAT-file format, and the implemented methods programmed in this language, which took advantage of its matrix computation capabilities. The visualization, including both plots and the model representations, was also accomplished resorting to MATLAB.

In what follows, the experimental methods employed in the present work will be explained. First, the relevant processes performed during the data acquisition will be presented. Second, the practical implementations will be portrayed. Third, an extensive analysis on the results obtained through the implemented model will be substantiated accordingly.

5.1 Data acquisition

Bernardino et al. (2016) present multiple datasets of different sensors (Kinect skeleton, Kinect RGB, electrocardiogram, respiratory rate, blood volume pressure, accelerometer data) of functional fitness tests being performed by young and elderly people. In this work, only data from the Kinect skeleton was used, hence, from now on, when referring to the datasets, only these are being considered. The tests consist on some exercises from the Senior Fitness Test (SFT): *2-minute step*, *30-second chair stand*, and *8-foot up-and-go*, and another test not present on the SFT, the *unipedal stance*. The details of these fitness assessment exercises, both their description and scoring methods, are explained in Section 3.1.

The acquisitions were performed at Faculty of Human Kinectics, Universidade de Lisboa (FMH) in 2015. The four exercises were performed by 11 young adults (7 female and 4 male), and by 10 senior adults (9 female and 1 male, with an average age of 61 years) (Bernardino et al., 2016). In total, the datasets consist of 21 people performing the four physical exercises stated above, with sometimes more than one repetition being recorded.

Looking at the organisation of the data, the Kinect SDK provides a skeleton structure with a 30 Hz sampling frequency, which includes:

- the timestamp for each frame,
- the orientation of the floor, in the Hessian normal form,
- the 3D positions of each of the 23 joints in a Cartesian coordinate system,
- the orientation of each joint as an orientation quaternion,

- and a ternary indicator of tracking, not-tracking, or inference on the position of each joint.

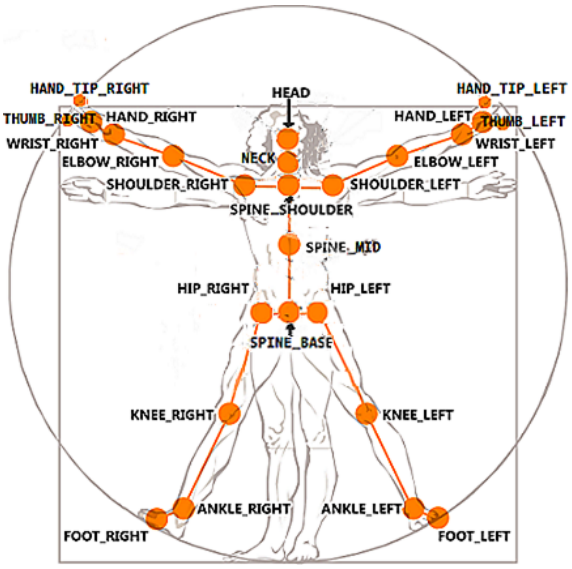


Figure 5.1: Set of skeleton joints provided by the the Kinect SDK. (Robotics, N.d.).

When acquiring the data, the person performing the exercise would be in front of the Kinect sensor, which was still and able to capture the full body of the participant for the entire duration of the exercise. The capture was done either in the sagittal plane, or in the coronal plane, depending on the movement being performed. However, there were some exceptions, as for some subjects the same exercise was recorded in different angles, specifically in the *2-minute step* test.

5.2 Data processing

The practical implementation of the kinematic filter use multiple techniques combined, some of which required precise tuning of various parameters. The processing of the data took multiple steps, from the simple removal of gross errors by hand, to the actual kinematic filter and processing by the UKF. In this section, the steps performed in the experiment, starting from the raw data until the collection of the final results are delineated, providing information on the methods used in the creation of the kinematic filter.

First, before the actual filter were to take place, there were some cases where multiple people were recognized by the Kinect, yet only the subject in question would be performing the exercise. Therefore, the data from the other people who were present was discarded manually. Also, some datasets included periods of missing data, where nothing had been captured for some consecutive frames. The analysis and treatment of these incomplete measurements is beyond the scope of this work and these were, in general skipped. However, when occurring in the middle of the exercise, the last frame and the first frame after the blank period were concatenated by hand.

Then, the filter is applied. Its goal is to reduce noise, either originating directly from the data or generated by the Kinect SDK low precision real-time methods. The filter takes advantage of *a priori* knowledge on the maximum velocities and accelerations of the movements being performed by the participants, in each game. With this knowledge, it is possible to attenuate movements which are not likely to be performed, such as rapid movements by a joint which is not expected to move that much. This knowledge is incorporated into the state-prediction step that determines in which directions and with which variations each joint moves, through the covariance of the noise of the state-transition model.

For the values used on the UKF, the sigma points' parameters selected were the common ones described in the literature, seen in Table 4.2, while the computation of the values that were set as the fixed link lengths were selected in a calibration phase. This period would ideally be performed before the execution of the exercise and in controlled circumstances, however, in the case of this work, as the acquisitions had already been performed, it was defined as the average value of the lengths acquired during the first second of each dataset, as the person would often remain still during this period, and so lower variations were experienced. The initial positions were set as generic standing positions for the *2-minute step* and *unipedal stance* test, and generic seated positions for the *30-second chair stand* and *8-foot up-and-go* exercises, approximately the same for each dataset, with initial velocities set to zero. Meanwhile, the values for the variations, *i.e.* the covariances, took quite some more tuning, and since there was no standard to compare the result to, the values were chosen by experimentation, trying to improve smoothness, while still maintaining a recognizable characteristic of the movement and a similarity to the original features visible in the data. This was done by permitting the system's variations to be high enough to cover the possible movements being performed, but constricting the ones which would be too high that would lead to wrong positions.

5.2.1 Covariance values

The selection of the filter parameters, such as the initial values for the positions and covariance, and for the covariances of the state-transition and observation model, were performed using multiple instances of each exercise. In fact, the use of these multiple datasets aimed at avoiding a bias towards a specific one, while intentionally leaving out others, thus separating the datasets into two sets. This way, there would be some unbiased datasets to analyse, as well as allowing for a comparison between both sets, so the generalization ability of the model could be assessed.

In this subsection, the best values found for the observation and state uncertainties, according to the explained criteria, will be depicted.

In this work, the covariances of the state-transition and observation processes were considered fixed through time, and are denoted by $\mathbf{Q}$ and $\mathbf{R}$. The tuning of these values was the most time-expensive step in the process. In fact, as there was no standard which could be used to perform any type of validation to set these values, they were set by experimentation according to the previously explained criteria.

First, in Table 5.1, the values for the observations' uncertainty were defined. These were all set to the

same value, as there was no performed analysis on the distribution of the noise of the data from the Kinect sensor. However, it was noticeable that in the exercise where the participant would be actually walking (*8-foot up-and-go* exercise), the noise would be slightly larger, thus a larger value was set in comparison to the remaining exercises.

Second, in Tables 5.3 and 5.2, the covariances of the uncertainty of the state transition are presented, for the upper-body and for the lower-body. The tuning of these values was aided by the existing knowledge on each of the exercises. In fact, a joint which has a larger and faster amplitude of movement in a certain exercise is expected to have larger covariances associated with it. The covariance of the link length was set 10^{-3} smaller than the smallest covariance of the model in question. Ultimately, the values selected were found to be a useful trade-off between the smoothness and noise cancelling, and the ability to describe the expected variations performed in the exercise.

Table 5.1: Observation covariance values for the upper-body and lower-body models (per frame). ($\mathbb{I}$ represents the identity matrix.)

2-minute step	*30-second chair stand*	*8-foot up-and-go*	*Unipedal stance*
$1 \cdot 10^{-4}\,\mathbb{I}$	$1 \cdot 10^{-4}\,\mathbb{I}$	$1 \cdot 10^{-3}\,\mathbb{I}$	$1 \cdot 10^{-4}\,\mathbb{I}$

Table 5.2: State covariance values for the upper-body model (per frame). The vector entries of the table all have the same specified value in their components.

Variable	*2-minute step*	*30-second chair stand*	*8-foot up-and-go*	*Unipedal stance*
$\mathbf{x}_{root}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\mathbf{v}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{root}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{spine}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{spine}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{neck}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{neck}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{shoulder-c}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{shoulder-c}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{shoulder-l/r}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{shoulder-l/r}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-5}$
θ_{elbow}	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
ω_{elbow}	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$
link length	$1 \cdot 10^{-11}$	$1 \cdot 10^{-10}$	$1 \cdot 10^{-10}$	$1 \cdot 10^{-11}$

Table 5.3: State covariance values for the lower-body model (per frame). The vector entries of the table all have the same specified value in their components.

Variable	*2-minute step*	*30-second chair stand*	*8-foot up-and-go*	*Unipedal stance*
$\mathbf{x}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\mathbf{v}_{root}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{root}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{hip}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{hip}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
θ_{knee}	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
ω_{knee}	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-5}$
$\boldsymbol{\theta}_{ankle}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{ankle}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-5}$
link length	$1 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-11}$

Finally, the initial value for the covariance of the state $\mathbf{P}$ was set as being $\mathbf{P}_0 = 100 \cdot \mathbf{Q}$. In fact, this allows for a faster convergence phase since there is no guarantee that, at the start of the exercise, the initial values set for the filter are close to the true ones.

5.3 Data analysis

The results obtained after applying the filter to the data were analysed by comparison to the original data from the Kinect sensor. The fact that there was no ground truth to compare the obtained results against, limited the possibility of a quantitative evaluation. This may be performed in future work for more meaningful results on the accuracy and precision of the proposed system.

In this section, several cases of the performed exercises are presented, illustrating the most interesting features of the filter. In the presented plots the colour blue represents the filtered data, and the colour red represents the original data, unless otherwise stated.

The datasets used for this work (Bernardino et al., 2016) included data from 21 participants performing the 4 senior fitness tests explained in Chapter 3. Thus, by segmenting the datasets into two sets, the analysis of the performance of the filter could be accomplished using datasets which were not used on the selection of the parameters. The fact that similar behaviour was obtained using the two sets provided an indication of low variability of the filter with respect to the selected parameters.

5.3.1 Exercise time-lapses

Figures 5.2 to 5.5 show time-lapses of an elderly female participant performing the senior fitness tests, both the original data obtained from the Kinect SDK (red), and the data provided by the kinematic filter (blue) proposed in this work. In these figures, the circles represent the joints' positions, and the lines depict the links connecting them. These figures show that, some times, the filter introduces a small delay on the movement, with the filter replaying the original movement with some time lag. However, as will be seen in Subsection 5.3.5 various anatomical errors are also corrected.

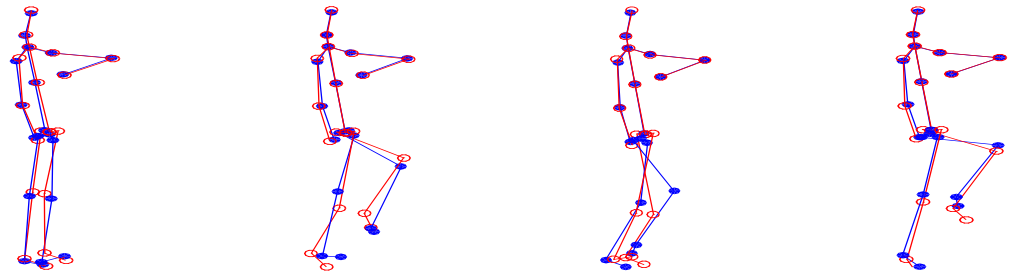


Figure 5.2: 2-minute step test time-lapse.

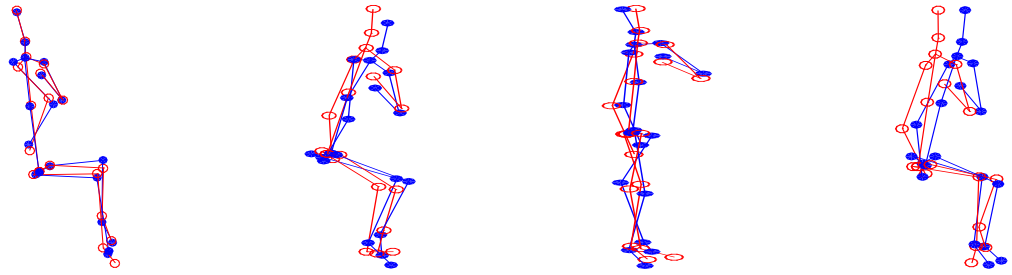


Figure 5.3: 30-second chair stand test time-lapse.

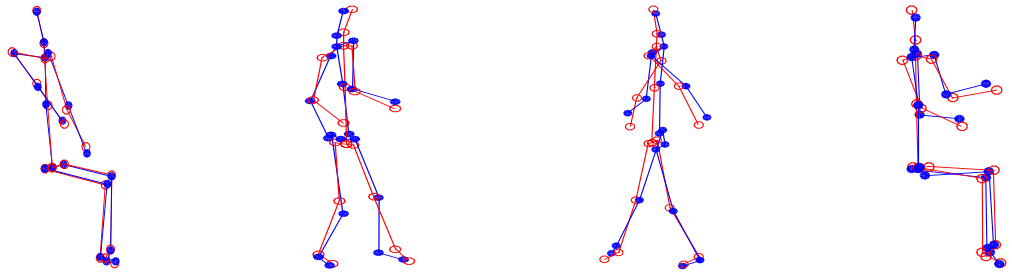


Figure 5.4: 8-foot up-and-go test time-lapse.

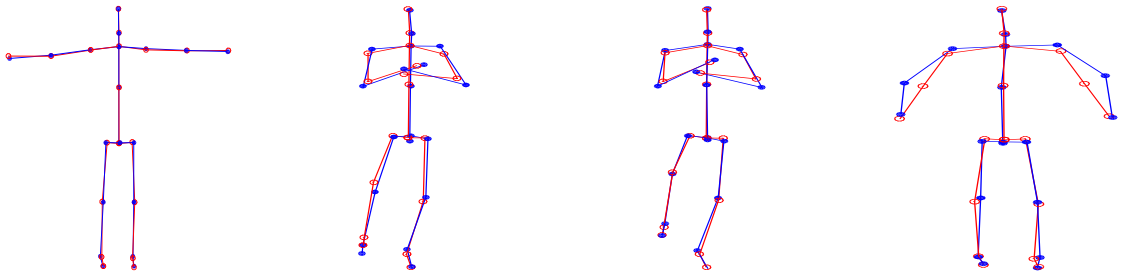
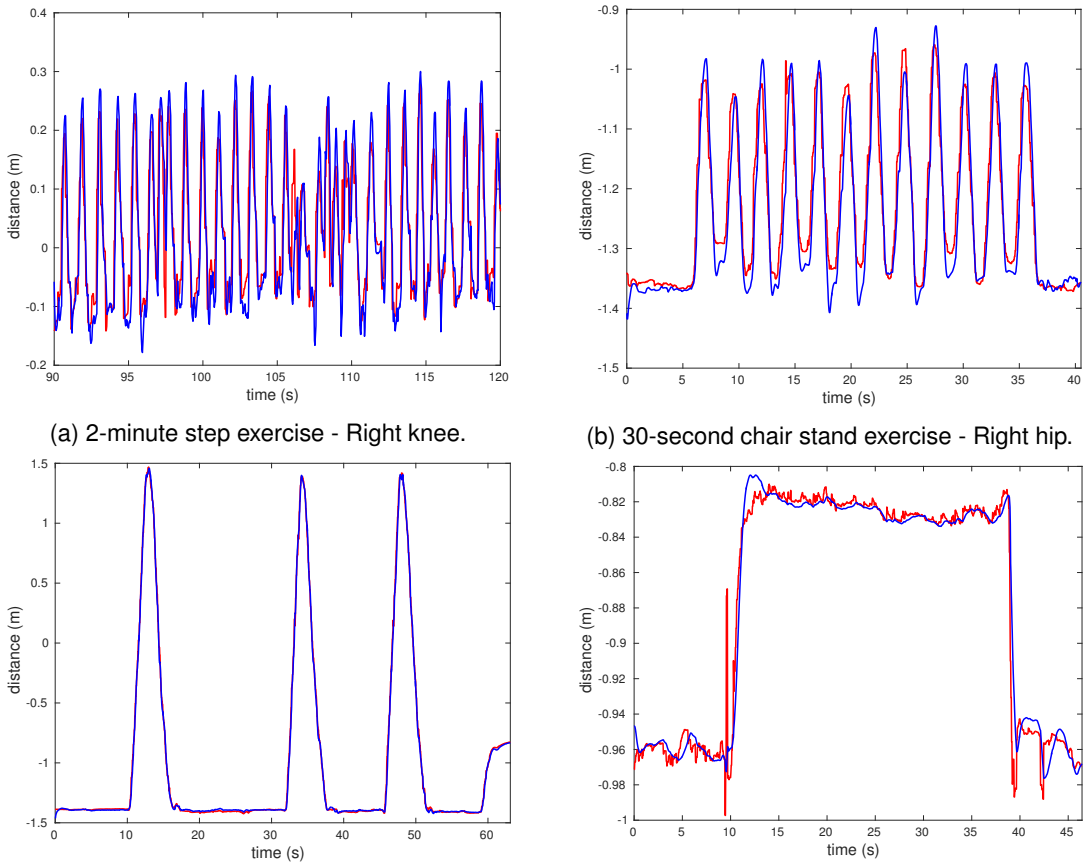


Figure 5.5: Unipedal stance test time-lapse.

5.3.2 SFT positions representation

In Figure 5.6 it is possible to observe the time plot of the position of selected joints of the four fitness tests, before and after applying the filter. The knee was chosen to represent the *2-minute step* test (Figure 5.6a) where the knee is periodically raised and lowered, while the hip was picked to describe the *30-second chair stand* test (Figure 5.6b) whose position increases for the standing up and decreases for the sitting down parts of the cycle. The hip is also the representative of the *8-foot up-and-go* test (Figure 5.6c) where the walking away and returning to the original position is seen thrice, and for the *unipedal stance* test (Figure 5.6d), the foot was selected since its raising off the ground is the most notable fact of this exercise. These joints were chosen as they were the ones that best described the exercise in question. In the *2-minute step* test the entire exercise is not presented due to its length, that would render the plot too dense, thus only 30 seconds from the middle of the exercise are shown. All the other plots show the entire acquisition, and all four were recorded from the same elderly female participant performing the exercises. The dataset for *8-foot up-and-go* test had 3 repetitions of the exercise, which are clearly seen as the movement of walking and returning to the original position.



(a) 2-minute step exercise - Right knee.

(b) 30-second chair stand exercise - Right hip.

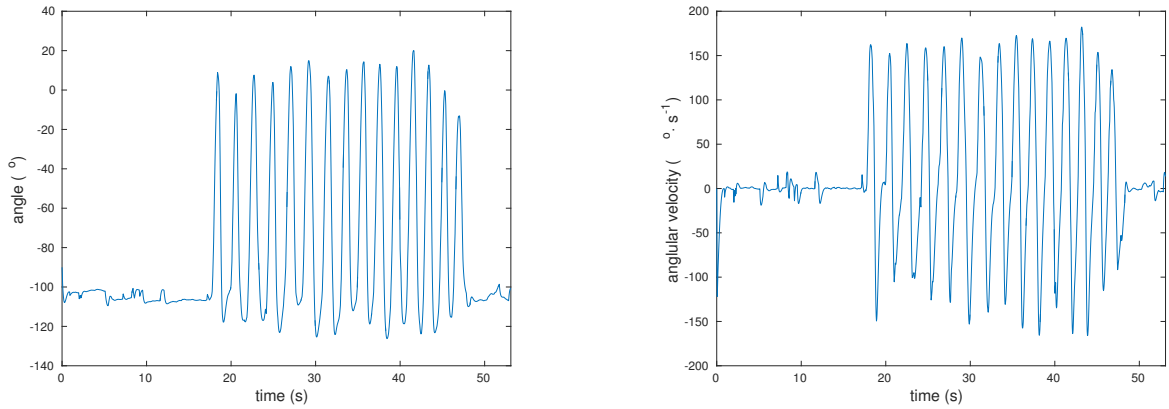
(c) 8-foot up-and-go exercise - Right hip.

(d) Unipedal stance exercise - Right foot.

Figure 5.6: Fitness exercises performed by an elderly female participant. In red, the original data is displayed, while in blue there is the processed data. (The distance in meters is relative to the Kinect sensor.)

5.3.3 Angular information

The kinematic filter, implemented using the UKF, used the angular information of the joints in the state parameter of the filter to model the skeleton structure using a kinematic chain design. Consequently, the filter processed the angles and angular velocities for each plane of the movement, describing movements of adduction and abduction, extension and flexion, and rotation. As such, Figure 5.7 demonstrate the capability of the filter to compute the knee angle (Figure 5.7a) and the corresponding angular velocity (Figure 5.7b). Here, as the knee only has one DOF, it matched the extension and flexion of the knee. In this plot, one may see the standing up and sitting down of the participant performing the chair stand for 15 times during the 30 seconds. When directly comparing the angle and the angular velocity, it is possible to see the zero-crossing of the velocity corresponding to the maximum and minimum of the angle, and the maximum velocity for halfway though the standing up or sitting down instants.



(a) Right knee angle.

(b) Right knee angular velocity.

Figure 5.7: Angle and angular velocity, from the kinematic filter, for the execution of the 30-second chair stand exercise performed by an elderly male participant.

5.3.4 Systematic errors

The data acquired using the Kinect sensor and present in the treated datasets displayed a variety of errors. These could be directly related to the acquisition method, which hindered the correct acquisition of data and led to some wrongfully detected movements. In fact, Figure 5.8 show the data from an elderly female participant executing the *2-minute step* test, where the acquisition was performed in the sagittal plane, and the sensor only detected one leg moving, *i.e.* the raising of each leg was seen as the same leg being raised twice the correct number of times. Nevertheless, other acquisitions in the same plane led to the correct detection of the exercise, which indicated that the acquisition angle was not the unique factor which caused the issue, suggesting a larger set of different, or unidentified, factors being responsible for this phenomenon. Here, the filter did not perform as intended, as the data did not agree with the values set for exercise.

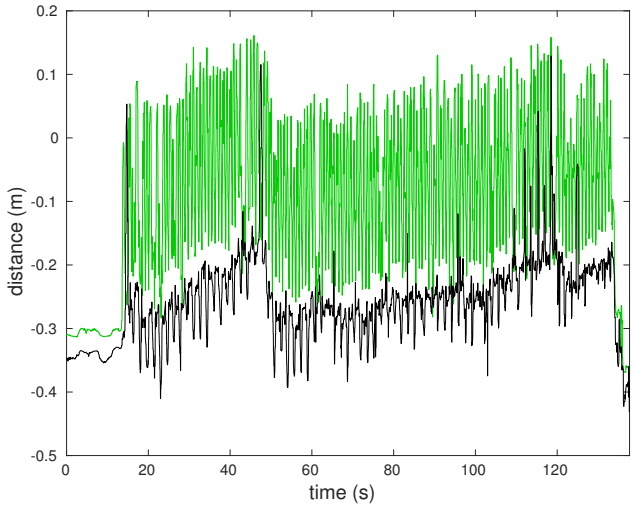


Figure 5.8: 2-minute step exercise performed by an elderly female, showing an error in the acquisition where only the left leg (green) is registered, instead of alternation with the right leg (black).

5.3.5 Anatomical errors

Errors related to the physical properties of the human body were also present in the dataset. Indeed, although each joint has a certain number of DOF, these may be constrained to a limited range, *e.g.* the knee or elbow can only be extended until the arm is completely straight and can only be flexed until it is fully bent, thus their angles have a range of rotation of approximately $[0, \pi]$. The distances between joints are also a physiological constraint of the human body, as the lengths of the bones are not variable. However, the data from the Kinect does not impose this constraint, and the lengths between joints, computed as the Euclidean distance between the positions, are variable through time. The implemented kinematic filter reduced this error, maintaining this distance close to the value set during the preceding calibration phasee. This can be seen in Figure 5.9 where the raw lengths of the lower leg keep oscillating according to the movement that the participant is performing (the standing up and sitting down movements of the *30-second chair stand* test) but the filtered data is almost constant.

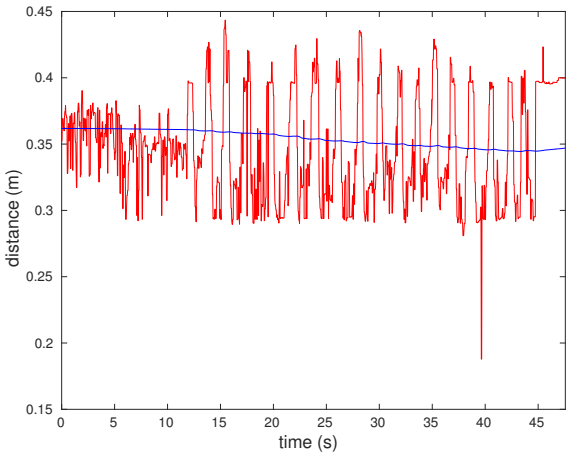
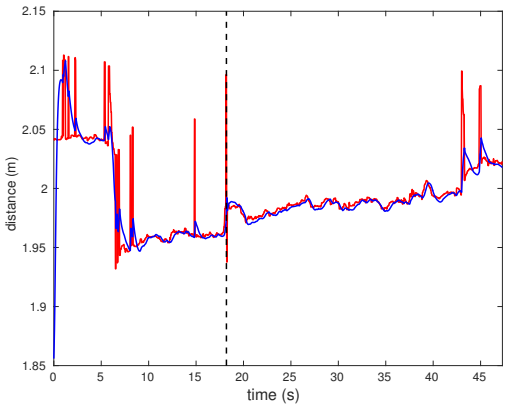
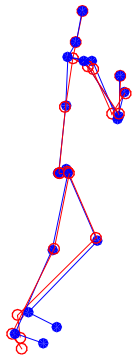


Figure 5.9: Lower leg length of an elderly female participant performing the 30-second chair stand exercise. In red the original lengths from the Kinect, and in blue the lengths while using the kinematic filter.

Another type of errors found in the data were large and fast variations that cannot describe the actual motion of the person being recorded. An example may be seen in Figure 5.10a showing multiple mistakes during an execution of the *unipedal stance* test, as large movements of the foot were described by spikes in the original data (red), and were highly attenuated by the filter (blue). A different visualization may be observed in the representation of the skeleton model (Figure 5.10b), at the instant marked in Figure 5.10a by a dashed black line (at around 18 seconds). Here, it is visible the error present in the raised foot, with the original data showing the foot fully extended (in the vertical position) whereas the filtered data maintained the foot in the expected raised position, which it held from before.



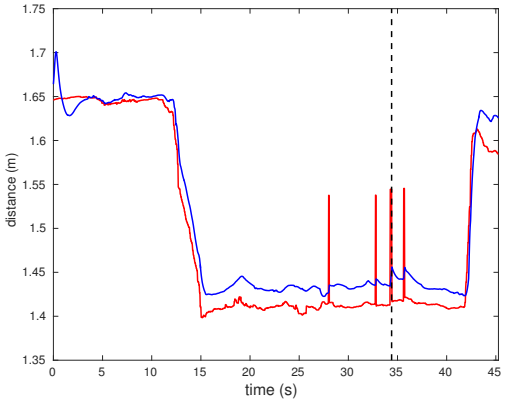
(a) Left foot plot.



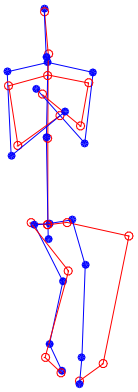
(b) Lateral view of the model.

Figure 5.10: Unipedal stance exercise by a young female participant. An error in the original data (dashed line) is represented in a model showing the difference between the original data (red) and filtered (blue).

A similar case may be seen in Figure 5.11, and also in the *unipedal stance* exercise, where, in this case, the knee of the weight supporting leg showed multiple repetitions of large deviations away from the body to its side, illustrated in Figure 5.11a by the large spikes in the position. As in the previous representation (Figure 5.10), the skeleton model was shown (Figure 5.11b) matching the time instant represented by the black dashed line in Figure 5.11a, and the mistake from the capturing of the knee is clearly seen (in red), while the filtered model kept the correct position.



(a) Left knee plot.



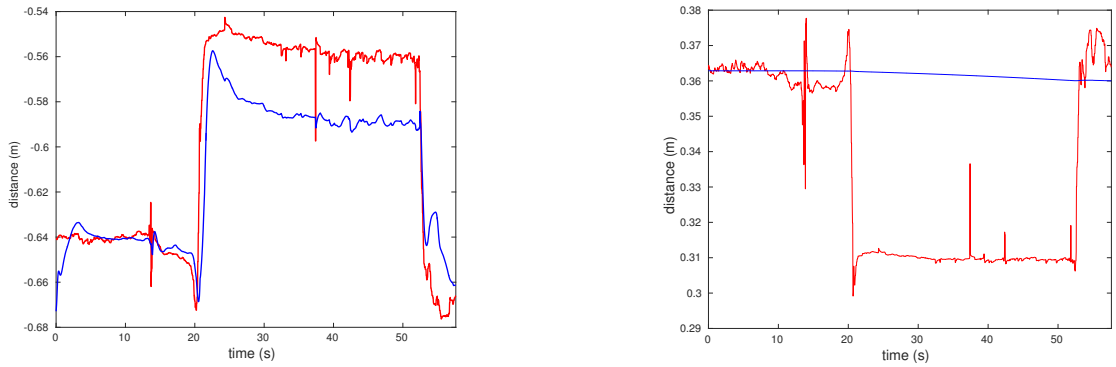
(b) Frontal view of the model.

Figure 5.11: Unipedal stance exercise by a young female participant. An error in the original data (dashed line) is represented in a model showing the difference between the original data (red) and filtered (blue).

5.3.6 Comparison between positions and link lengths

The correction of the variation of the link lengths present in the raw data sometimes had a significant influence in the joint positions. In fact, the proposed kinematic filter tries to correct these variations by using a calibration phase which computes the lengths, and then tries to maintain them constant, as seen in Figure 5.9. However, this process had significant influence on the position data of the joints of a few datasets, cre-

ating an offset between the filtered and original data. An example of this process may be seen in the raised knee position of the *unipedal stance* exercise in Figure 5.12. Here, the offset in the position of the knee (Figure 5.12a), and matching variation of the length of the upper leg (Figure 5.12b) can be recognisable.



(a) Right knee plot.

(b) Upper leg length.

Figure 5.12: Unipedal stance exercise performed by an elderly female showing an offset between 20 and 50 s (leg lifted), comparison between knee height and upper leg length, before (red) and after the kinematic filter (blue).

The final relevant feature identified in the results obtained was the convergence phase at the beginning of some exercises, a small period at the start of an exercise where the positions computed by the filter go from the initial values, closer to the ones provided by the Kinect sensor, following this data thereafter. Figure 5.10 show this process to a considerable extent, while Figures 5.11a and 5.12a evidence the same convergence phase, but to a smaller degree.

Chapter 6: Discussion

The work developed in the present project demonstrated the capabilities of the created kinematic filter to regularize the model of the human body through the skeleton structure provided by the Kinect SDK. The Kinect sensor use in scientific research has been of considerable importance, as it provides an easily accessible and affordable device, specially due to the availability of the Kinect SDK which makes the interaction with the device particularly easy. However, it does have many flaws. These flaws may not be much relevant for some applications, such as games or gesture recognition, but when considering health related applications, special care should be taken. The developed filter introduced conditions that establish biomechanical compatibility in the structure of the data, enhancing the skeleton structure provided by the SDK. The presented model has been shown to have various advantages over the raw measurements from the Kinect, but some drawbacks were also detected. In this chapter, these advantages and disadvantages will be highlighted.

6.1 Reasoning for the algorithmic approach

An UKF model was implemented to filter the data, as explained in Chapter 4. The structure of a KF was a natural fit to the problem at hand, but because the human kinematic model is non-linear, it was used one of the non-linear approximations described in the literature. The UKF was selected as it is a nicely suited approach for non-linear systems (Julier and Uhlmann, 2004). Nevertheless, the attempt to use other algorithms for the kinematic modelling would be an interesting development on this project, and would allow a comparison between algorithms. Indeed, the EKF is the most common KF alternative for non-linear state-prediction and/or observation functions. However, its approximation using a linearisation around a given point by taking the multivariate Taylor series expansion of the function (usually only the first-order term) may not be the best fit for functions presenting large degrees of non-linearity (Julier and Uhlmann, 2004). Another possibility might be the use of a particle filter, as the unscented particle filter (Van Der Merwe et al., 2001), which instead of using the few and predefined sigma points approach of the UKF, uses a large sample of points, applying the function at each one, and reconstructing the result of the non-linear function. However, this is a more computationally expensive system, which may not be adequate for some purposes, *e.g* a robot with limited computational capabilities.

Despite the dimension and complexity of the developed filter, its computation is fast enough to apply the proposed model in a real-time environment. However, real-time was not the focus of this work since all the performed tests were done using the datasets already acquired and published by Bernardino et al. (2016), so the implemented algorithm was oriented to off-line processing. Notwithstanding, minor modifications to the core of the algorithm would allow its execution in real-time.

6.2 Evaluation of the system

Regarding the evaluation of the results, a quantitative analysis of the filter capabilities was not possible, as there was no ground truth data in the used datasets. Future work should consider capturing a set of movements with both the Kinect sensor and with a millimetre or sub-millimetre precision marker-based system. This would gather information for the assessment of the proposed system's accuracy and precision. Moreover, since SFTs are widely established and used in the literature, this would also enable the direct comparison to systems developed by other authors. Apart from the technical issues, the *unipedal stance* exercise had a design flaw, as the 20 participants of the total of 21 participants, achieved maximum score, so a reformulation of the exercise should be performed.

Given the fact that no quantification of the performance was possible, a qualitative analysis of the results was presented in Chapter 5. With respect to the filter's performance, it would be interesting to define a set of motion parameters typically used by health professionals. Computing these motion parameters on both the filtered and original data, against the judgement from a health professional. This would contribute with more insight on the performance of the kinematic filter, producing results more relevant to its actual purpose.

The implemented algorithm required the tuning of parameters by experimentation, as covariance values. These values, if too high would lead to high frequency oscillations in the data, while too low covariance values would slow down the variations required for the correct modelling of the exercise. To prevent overfitting, the tuning of the filter parameters was done in only a fraction of the dataset. A comparison between the datasets which were used to pick the parameters and the ones which were not, had similar performance, so similar behaviour ought to be expected in possible future acquisitions.

6.3 Advantages of using the filter

The kinematic filter used angular information to model the joints of the skeleton. This information is of great importance to the analysis of human movement features, as the angles and angular velocities are major factors in the kinematic study of the joints of the human body, being used on a large portion of the literature on human kinematic analysis. Moreover, the information computed by the filter takes into account the three directions of joint movement, *i.e.* adduction and abductions, flexion and extension, and rotation. In fact, this is substantial, as it would otherwise require the computation of all these directions by the defining the frames for each joint at each time step. In opposition, the angle between two links, which may be computed simply by using the scalar product on the directions of the links, provides considerably less information.

Another achieved result was the correction of wrong information exhibited in the Kinect data. As a matter of fact, the kinematic filter defined various constraints on the movement, based on limitations of the changes possible for the features described in the state. These constraints limited the variations possible for each joint, by using *a priori* knowledge on the movements being performed, in this case the senior

fitness tests. Then, only variations which seemed reasonable would be allowed, while still enabling enough variation for the correct modelling of the exercise. In order to do this, the system was enforced to follow the physical equations of motion, for the time step relative to the sampling frequency of the Kinect. Therefore, the fluctuations in position from frame to frame cannot suddenly change, being restricted to the possible variations given by the current velocity at each time step. In fact, the implemented model was able to correct instances of these incorrect variations which were present in the data, as seen in Figures 5.10, 5.11 and also Figure 5.6d, where there were large fast variations in the data. Then, since the filter expected low variation in this exercise, as the *unipedal stance* exercise is mostly static and does not take executing rapid movements, the variations were mostly attenuated.

In addition, the filter was able to correct another physiological mistake existing in the data from the Kinect SDK. Indeed, the variation of the lengths of the links was clearly an inaccurate feature acquired by the Kinect. As previously stated, this was an important factor for the accuracy of the data, due to the fact that it represented a wrongful description of the human anatomy, being responsible for non-anthropometric modelling of the human body (Obdržálek et al., 2012). As seen in Figures 5.9 and 5.12b, the filter was able to attenuate this feature, using a calibration phase at the beginning of the acquisition, and strongly restricting the variation of this value. Nevertheless, if this calibration phase fails to capture the correct values for the lengths, it could be possible that an amplification of this error could happen. However, this did not happen for any dataset studied. Thus, the calibration should be performed with care, considering a large number of samples, ideally with the subject in a static position and confirming that such a position would allow for a correct acquisition. Finally, it was found that the length variations were highly dependent on the movement, *e.g.* Figure 5.9 shows the leg lengths oscillating as the person is standing up or sitting down but this is corrected by the filter, with the length staying stable at about half way. Also, Figure 5.12b shows a drop of the upper leg length right when the hip and knee are flexed.

6.4 Limitations of using the filter

In the implementation of the filter, no information on the rotation of a link along its axis was introduced on the system from the observations. Yet, the state had this information encoded in the θ_x and ω_x components, and the state-transition function made attempts to perform these predictions. Thus, as the system took into account the uncertainty associated with the rotation, but there was no information observed about this direction, random variations were observed regarding these angles and velocities, as garbage data was returned by the filter for these components. So, the inclusion of orientation data would be an important improvement for this filter. This information is also provided by the skeleton structure by the Kinect SDK in the form of orientation quaternions. Nevertheless, the interpretation of these values was not easily accomplished, as there was very few information provided on this data structure, on its meaning, and also on its application methods, so further study and experimentation is mandatory.

There were two strong assumptions about the behaviour of the noise in the acquisition process. The first one was that the noise was Gaussian. However, this was clearly not true, but the statistical characterization of the noise would be a time demanding process which was not possible to do, so this simpler and standard assumption about an unknown noise process was used. The second one was that the noise did not vary depending on the position and movement throughout the exercise. Also, this was not correct as corroborated by analysing the data and verifying the dependence of the noise amplitude on the speed of the movement magnitude, and the direction in which the recording was performed in relation to the positioning of the participant. In principle, the filter could be extended to include a non-stationary noise model depending on the position or velocity on the previous iteration, dynamically updating the covariance matrices. Also, a study on the dependency of the noise on the directions of capturing could also be carried out and used to optimize the process of recording.

One large difference between the original data and the filtered one was an offset found in some instances of the exercises, as seen in Figure 5.12a. This phenomenon was identified in some plots of positions of the lower limbs of participants performing the *unipedal stance* test. Through a comparative analysis with the lengths of the upper and lower leg, a matching offset was encountered, as seen in Figure 5.12b. This might be a reason for this effect, as the links had their length variation constrained while large variations were actually happening in the data. Indeed, the covariance for the positions was set to be larger than for the link lengths, than first are expected to change more than the latter.

A significant mistake which was identified by the representation and visualization of the human skeleton was the delay introduced in some exercises. This may be an indication that the values set for the allowed variations for that exercise were set too low and higher ones should be investigated.

Also, this representation showed that sometimes there was a significant offset between the processed roots of the upper-body and lower-body models which are computed independently. Thus, a technique for computing the root's position should be employed by either combining the two in some way, *e.g.* taking the average of the two, or extending the model such that the entire skeleton is processed by the filter at the same time.

Another possible issue which should be taken into account was the non-constant convergence times for the exercises. Although taking mostly short periods, the convergence from the initializations to the true positions represent wrong positions. The initialization was set for the four tests separately, but uniformly for all the executions of the same exercise. Consequently, the initialization did not correspond to the actual initial positions of the participant, possibly even being quite far from them. So, some exercises show longer convergence phases than others depending on how close they are to the initial positions. Indeed, in one hand, Figure 5.10a shows a large variation in the beginning, yet in a short period of time, while on the other hand, Figures 5.11a and 5.12a show a smaller but longer convergence phase. Both outcomes have clear negative implications, as for a long period the data did not correspond to the truth for a longer time, while fast and large variations could lead to large velocities being attained, leading to illegitimate interpretations of the

movements. In order to prevent such errors, the recording could include a convergence phase, discarded in the computation of any parameters, which would allow the filter to converge to a reasonable initial condition and then acquire the actual, correct data. In this work, previously acquired datasets were used, so one alternative would be to compute the parameters ignoring a certain time period *e.g.* the first 30 of 60 frames (equivalent to 1 or 2 second) of data, as there the filter would most likely be in the convergence phase. Nevertheless, this might not be a good solution, since for some datasets the participant starts the exercise right away, and so the initial frames cannot be discarded as they already described the execution of the exercise. So, new acquisitions could be done taking this facts in consideration, and explicitly record a few seconds of data preceding the exercise to be used as calibration of the filter.

Moreover, the filter could not identify errors which did not describe non-anatomical limitations, as the data anatomically seemed like a possibly correct movement but does not correspond to the real action being performed, and only an individual analysis could have allowed this detection. For example, Figure 5.8 shows an acquisition of an instance of the *2-minute step* exercise, where the information provided by the SDK showed the participant raising the left leg twice the times she had actually done, while not raising the right leg, which does not describe the real movement. So, the Kinect recorded only one leg moving with twice the frequency as it failed to identify which of the legs was in reality moving. Therefore, the covariance set for the exercise was about half the velocity which was recorded, and the filter did not perform as expected. Although sequences like this could only be identified for sure by a direct analysis, a prior study on the best angles of acquisition for each exercise could bring forward conclusions about the best methods to perform the recordings, decreasing the possibilities of occurrences like the one described in this paragraph.

The last points refers to the final limitation of the proposed model: the requirement for knowledge on the movements being performed. Consequently, there is the need to create “profiles” for each exercise, as each different exercise has its own movement requirements, and so, its own covariance matrix. This implies that for a new exercise, requiring the performance of a different set of movements, a new profile needs to be created. In order to attack this limitation, the natural mechanism would be an adaptive noise model, as this limitation is basically setting different covariance matrices, *i.e.* noise models, depending on the movement being performed, and the inability of the filter to adapt. Then, this would be introduced either by dynamically changing the covariance model throughout the exercise, extracting information from the movements. Or, it would require a calibration such as, for example, one repetition of the movement to capture the variations and create the model, and a second one to actually process the exercise.

6.5 Relevant implementation details

The way the filter computes its main feature, the Kalman gain (Equation 4.42), is also its main drawback, both in terms of computation cost, as inverting such a large matrix (observation model covariance matrix) takes a long time to compute for each time step. In addition, the factorization performed to calculate the square root

of the state covariance matrix, implemented using the Cholesky decomposition, in the computation of the sigma points for each iteration (Equations 4.23 to 4.25), was also a large computation bottleneck. However, these two steps of the process constrict even more the possible computations which might be tried, as the matrices must be non-singular and positive-definite, respectively. These details introduce some limitations on the range of possible operations which might had been performed. Indeed, this is more easily seen by considering the experiment where one wants to remove the information with respect to the observation data (or from the state prediction) from one of the iterations. In fact, this would make sense if this information to be removed represented outliers or any clear mistake on the data. If that was the case, it would be logical to set the Kalman gain to attribute zero weight to the observation (or prediction) information, while keeping only the other one for the filtered data, as the Kalman gain controls how much weight is given to which scope of the filter. This could be even further improved by the incorporation of the intrinsic stochastic nature of the UKF to find out if the observations fall in an acceptable range of the predictions according to, for example, a likelihood function, and accept or reject based on this information. Nevertheless, this is not easily feasible to be accomplished many repeated times (*e.g.* many outliers in consecutive frames), as it would increase the possibility that the resulting computed covariance matrix would turn out to be singular. Still, this approach is given as a promising method which requires additional study and experimentation, being an important process to be surveyed in the future.

6.6 Important remark

It should be noted that the implemented system was only realized as a proof of concept, which was not fit or programmed to be employed in an actual system, being only able to process previously acquired data, and was only able to handle one person recordings. In a real world application, restricting the capture to only one subject would be infeasible, as even though one could restrict the physical space to only one person, the Kinect seldom wrongfully identifies objects as people, crashing the system as it cannot handle more than one person at a time. Therefore, before an actual practical use of the system, further studies must be carried out to establish the usefulness of the filter. Then, if it turns out that positive results are attained, the system should be ported, integrating the provided API for the interaction with the Kinect, and taking care of its practical limitations, so that a useful application could, in the end, be completed.

Chapter 7: Conclusions

In conclusion, a method has been developed aiming at the improvement of the accuracy of the data supplied by the Kinect SDK. This method, a kinematic filter based on a UKF model, can be used in many different applications devised to promote physical activity on its users, such as assistant robots or entertainment systems employed on day care centres. It was created focused on data from senior fitness tests, however it could easily be generalized for other cases, as long as there is a previous knowledge on the movements being performed. Moreover, it allows for the prevention of certain wrong features of the skeleton data from the Kinect, but also provides information on the angles of each joint in either of the anatomical planes set to be movable by that joint, and also the corresponding angular velocity for the same directions. Therefore, the accuracy and extraction of kinematic parameters from Kinect skeleton data would be highly enhanced by the proposed biomechanical model which improves the anatomical coherence of the subject participating in the physical training.

7.1 Accomplishments and further developments

In this work, most of the proposed objectives presented at the start were accomplished. Indeed, it was possible to achieve the creation of a kinematic filter model which has the capability to significantly improve the accuracy of the data provided by the Kinect SDK. In fact, the filter was able to eliminate multiple wrong features existing in the data, having the capability to highly attenuate the variations in link lengths present in the original data, and also to remove erroneous large amplitude changes that do randomly appear in the skeleton data. Therefore, this method provides a valuable pre-processing method intended for applications using this device for biomechanical analysis. Nevertheless, some limitations still prevailed, as it was not possible to limit the the movements by establishing limits on the angles of each joint according to their physiological limit.

Ultimately, there are multiple points that would still be of great value if were to be addressed. In fact, one important thing missing from this study was the comparison of the obtained kinematic filter results against a standard, which would provide a more clear insight on the actual capabilities of the created model, as it would give the means to perform a quantitative assessment of the performance, which was not possible during the development of this work. Moreover, the extraction of movement parameters from the filtered data would also be an interesting development on this work, establishing a comparison with the results from other methods found in the literature, and even more interestingly, a comparison with the *de facto* method of evaluating movement features, the opinion of a health specialist. Regarding the practical implementation, there are a few points which are worth mentioning, first, the unaccomplished ability to be able to predict, and prevent the joint's rotations outside their true anatomical capabilities. Second, the introduction of observations containing orientation information, *i.e* the orientation quaternions provided by the SDK, which could

then improve accuracy of the system and enhancing the filter with the capability to compute the rotation DOF of some joints. Third and finally, reduce the number of assumptions made, such as the noise to be Gaussian and additive, as well as allow for variations of the noise models depending on the movement being performed. Also, a further fine tuning of all the filter's parameters would certainly improve the performance. The last group of relevant developments which would be interesting would be to design the kinematic filter using other methods suited for the purpose, such as using the standard EKF instead of the less common UKF employed, or using a particle filter instead, and finally comparing the results of all algorithms. In the end, porting the system in a way that it would be able to be employed in real-time applications, and in real systems.

References

Ahlskog, J. Eric, Yonas E. Geda, Neill R. Graff-Radford and Ronald C. Petersen. 2011. Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. In *Mayo Clinic Proceedings*. Vol. 86 pp. 876–884.

Alexiadis, Dimitrios S., Philip Kelly, Petros Daras, Noel E. O'Connor, Tamy Boubekeur and Maher Ben Moussa. 2011. Evaluating a dancer's performance using Kinect-based skeleton tracking. In *Proceedings of the 19th ACM international conference on Multimedia*. pp. 659–662.

Andersson, Virginia, Rafael Dutra and Ricardo Araújo. 2014. Anthropometric and human gait identification using skeleton data from Kinect sensor. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. pp. 60–61.

Bernardino, Alexandre, Christian Vismara, Sergi Bermudez i Badia, Élvio Gouveia, Fátima Baptista, Filomena Carnide, Simão Oom and Hugo Gamboa. 2016. A dataset for the automatic assessment of functional senior fitness tests using kinect and physiological sensors. In *International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*. pp. 1–6.

Bishop, Gary and Greg Welch. 2001. "An introduction to the kalman filter." *Proc of SIGGRAPH, Course* 8(27599-23175):41.

Biswas, Aviroop, Paul I. Oh, Guy E. Faulkner, Ravi R. Bajaj, Michael A. Silver, Marc S. Mitchell and David A. Alter. 2015. "Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults a systematic review and meta-analysissedentary time and disease incidence, mortality, and hospitalization." *Annals of internal medicine* 162(2):123–132.

Bonnechere, Bruno, Bart Jansen, P. Salvia, H. Bouzahouene, L. Omelina, J. Cornelis, M. Rooze and S. Van Sint Jan. 2012. What are the current limits of the Kinect sensor. In *Ninth International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*. pp. 287–294.

Bostanci, Erkan, Nadia Kanwal and Adrian F. Clark. 2015. "Augmented reality applications for cultural heritage using Kinect." *Human-centric Computing and Information Sciences* 5(1):1–18.

Chen, Chen, Kui Liu, Roozbeh Jafari and Nasser Kehtarnavaz. 2014. Home-based senior fitness test measurement system using collaborative inertial and depth sensors. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. pp. 4135–4138.

Colcombe, Stanley and Arthur F. Kramer. 2003. "Fitness effects on the cognitive function of older adults: a meta-analytic study." *Psychological science* 14(2):125–130.

Craig, John J. 2005. *Introduction to robotics: mechanics and control*. Vol. 3 Pearson Prentice Hall Upper Saddle River.

Department of Health and Human Services (US). 2008. *2008 Physical Activity Guidelines for Americans*. Department of Health and Human Services (US).

Deterding, Sebastian, Dan Dixon, Rilla Khaled and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. pp. 9–15.

Ekelund, Ulf, Jian'an Luan, Lauren B. Sherar, Dale W. Esliger, Pippa Griew, Ashley Cooper, International Children's Accelerometry Database (ICAD) Collaborators et al. 2012. "Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents." *Jama* 307(7):704–712.

Ford, Earl S. and Carl J. Caspersen. 2012. "Sedentary behaviour and cardiovascular disease: a review of prospective studies." *International journal of epidemiology* pp. –078.

Gabel, Moshe, Ran Gilad-Bachrach, Erin Renshaw and Assaf Schuster. 2012. Full body gait analysis with Kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. pp. 1964–1967.

Gauthier, Stevens and Ana-Maria Cretu. 2014. Human movement quantification using Kinect for in-home physical exercise monitoring. In *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2014 IEEE International Conference on*. pp. 6–11.

Gerling, Kathrin, Ian Livingston, Lennart Nacke and Regan Mandryk. 2012. Full-body motion-based game interaction for older adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1873–1882.

Giblin, Susan, Dara Meldrum, Mark McGroarty, Stuart O'Brien, Simon Rand, Stephen Smith, Karl Grogan and Friedrich Wetterling. 2016. Bone length calibration can significantly improve the measurement accuracy of knee flexion angle when using a marker-less system to capture the motion of countermovement jump. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. pp. 392–397.

González, Alejandro, Mitsuhiro Hayashibe and Philippe Fraisse. 2013. Subject-Specific Center of Mass Estimation for In-Home Rehabilitation–Kinect-Wii Board vs. Vicon-Force Plate. In *Converging Clinical and Engineering Research on Neurorehabilitation*. Springer pp. 705–709.

Graves, Lee EF, Nicola D. Ridgers, Karen Williams, Gareth Stratton, Greg Atkinson and Nigel T. Cable. 2010. "The physiological cost and enjoyment of Wii Fit in adolescents, young adults, and older adults." *Journal of Physical Activity and Health* 7(3):393–401.

Grewal, Mohinder S. 2011. *Kalman filtering*. Springer.

Hamari, Juho, Jonna Koivisto and Harri Sarsa. 2014. Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences (HICSS)*. pp. 3025–3034.

Hu, Frank B. 2003. "Sedentary lifestyle and risk of obesity and type 2 diabetes." *Lipids* 38(2):103–108.

Julier, Simon J. and Jeffrey K. Uhlmann. 1997. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97*. pp. 182–193.

Julier, Simon J. and Jeffrey K. Uhlmann. 2004. "Unscented filtering and nonlinear estimation." *Proceedings of the IEEE* 92(3):401–422.

Kalman, Rudolph Emil et al. 1960. "A new approach to linear filtering and prediction problems." *Journal of basic Engineering* 82(1):35–45.

Katzmarzyk, Peter T., Timothy S. Church, Cora L. Craig and Claude Bouchard. 2009. "Sitting time and mortality from all causes, cardiovascular disease, and cancer." *Med Sci Sports Exerc* 41(5):998–1005.

Kranz, Matthias, Andreas MöLler, Nils Hammerla, Stefan Diewald, Thomas PlöTz, Patrick Olivier and Luis Roalter. 2013. "The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices." *Pervasive and Mobile Computing* 9(2):203–215.

Kurillo, Gregorij, Ferda Ofli, Jennifer Marcoe, Paul Gorman, Holly Jimison, Misha Pavel and Ruzena Bajcsy. 2015. Multi-disciplinary design and in-home evaluation of kinect-based exercise coaching system for elderly. In *International Conference on Human Aspects of IT for the Aged Population*. pp. 101–113.

Larsen, Lisbeth H., Lone Schou, Henrik Hautop Lund and Henning Langberg. 2013. "The physical effect of exergames in healthy elderly—a systematic review." *GAMES FOR HEALTH: Research, Development, and Clinical Applications* 2(4):205–212.

Lee, I-Min, Eric J. Shiroma, Felipe Lobelo, Pekka Puska, Steven N. Blair, Peter T. Katzmarzyk, Lancet Physical Activity Series Working Group et al. 2012. "Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy." *The lancet* 380(9838):219–229.

Ljung, Lennart. 1979. "Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems." *IEEE Transactions on Automatic Control* 24(1):36–50.

Machida, Eiji, Meifen Cao, Toshiyuki Murao and Hiroshi Hashimoto. 2012. Human motion tracking of mobile robot with Kinect 3D sensor. In *2012 Proceedings of SICE Annual Conference (SICE)*. pp. 2207–2211.

Michael, David R. and Sandra L. Chen. 2005. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade.

Microsoft. N.d. "Kinect Sensor for Xbox One." https://www.microsoft.com/en-us/store/d/Kinect-Sensor-for-Xbox-One/91HQ5578VKSC [Accessed: 10 June 2017].

Moniz-Pereira, Vera, Silvia Cabral, Filomena Carnide and António P. Veloso. 2014. "Sensitivity of joint kinematics and kinetics to different pose estimation algorithms and joint constraints in the elderly." *Journal of applied biomechanics* 30(3):446–460.

Mundher, Zaid A. and Jiaofei Zhong. 2014. "A real-time fall detection system in elderly care using mobile robot and kinect sensor." *International Journal of Materials, Mechanics and Manufacturing* 2(2):133–138.

National Center for Health Statistics (US). 2016. *Health, United States, 2015: with special feature on racial and ethnic health disparities*. National Center for Health Statistics (US).

Nikander, Riku, Harri Sievänen, Ari Heinonen, Robin M. Daly, Kirsti Uusi-Rasi and Pekka Kannus. 2010. "Targeted exercise against osteoporosis: a systematic review and meta-analysis for optimising bone strength throughout life." *BMC medicine* 8(1):47.

Nixon, Mason E., Ayanna M. Howard and Yu-Ping Chen. 2013. Quantitative evaluation of the Microsoft Kinect TM for use in an upper extremity virtual rehabilitation environment. In *2013 International Conference on Virtual Rehabilitation (ICVR)*. pp. 222–228.

Obdržálek, Štěpán, Gregorij Kurillo, Ferda Ofli, Ruzena Bajcsy, Edmund Seto, Holly Jimison and Michael Pavel. 2012. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In *2012 annual international conference of the IEEE Engineering in medicine and biology society (EMBC)*. pp. 1188–1193.

Ofli, Ferda, Gregorij Kurillo, Štěpán Obdržálek, Ruzena Bajcsy, Holly Brugge Jimison and Misha Pavel. 2016. "Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned." *IEEE Journal of Biomedical and Health informatics* 20(1):201–212.

Otte, Karen, Bastian Kayser, Sebastian Mansow-Model, Julius Verrel, Friedemann Paul, Alexander U. Brandt and Tanja Schmitz-Hübsch. 2016. "Accuracy and reliability of the kinect version 2 for clinical measurement of motor function." *PloS one* 11(11):–0166532.

Parajuli, Monish, Dat Tran, Wanli Ma and Dharmendra Sharma. 2012. Senior health monitoring using Kinect. In *2012 Fourth International Conference on Communications and Electronics (ICCE)*. pp. 309–312.

Pastor, Isaac, Heather A. Hayes and Stacy JM Bamberg. 2012. A feasibility study of an upper limb rehabilitation system using kinect and computer games. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 1286–1289.

Pierleoni, Paola, Alberto Belli, Lorenzo Palma, Marco Pellegrini, Luca Pernini and Simone Valenti. 2015. "A high reliability wearable device for elderly fall detection." *IEEE Sensors Journal* 15(8):4544–4553.

Ratsamee, Photchara, Yasushi Mae, Kenichi Ohara, Tomohito Takubo and Tatsuo Arai. 2012. People tracking with body pose estimation for human path prediction. In *2012 International Conference on Mechatronics and Automation (ICMA)*. pp. 1915–1920.

Rikli, Roberta E. and C. Jessie Jones. 2013. *Senior fitness test manual*. Human Kinetics.

Robotics, Microsoft. N.d. "Kinect Sensor." `https://msdn.microsoft.com/en-us/library/hh438998.aspx` [Accessed: 10 June 2017].

Saunders, Travis J., Jean-Philippe Chaput and Mark S. Tremblay. 2014. "Sedentary behaviour as an emerging risk factor for cardiometabolic diseases in children and youth." *Canadian journal of diabetes* 38(1):53–61.

Sawatzky, Richard, Teresa Liu-Ambrose, William C. Miller and Carlo A. Marra. 2007. "Physical activity as a mediator of the impact of chronic conditions on quality of life in older adults." *Health and quality of life outcomes* 5(1):68.

Schutzer, Karen A. and B. Sue Graves. 2004. "Barriers and motivations to exercise in older adults." *Preventive medicine* 39(5):1056–1061.

Shotton, Jamie, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook and Richard Moore. 2013. "Real-time human pose recognition in parts from single depth images." *Communications of the ACM* 56(1):116–124.

Shu, Jody, Fumio Hamano and John Angus. 2014. "Application of extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates." *Journal of Engineering Mathematics* 88(1):161–175.

Simon, Dan. 2006. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.

Sinclair, Jeff, Philip Hingston and Martin Masek. 2007. Considerations for the design of exergames. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*. pp. 289–295.

Staiano, Amanda E. and Sandra L. Calvert. 2011. "Exergames for physical education courses: Physical, social, and cognitive benefits." *Child development perspectives* 5(2):93–98.

Stone, Erik E. and Marjorie Skubic. 2011. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. pp. 71–77.

Stone, Erik E. and Marjorie Skubic. 2015. "Fall detection in homes of older adults using the Microsoft Kinect." *IEEE journal of biomedical and health informatics* 19(1):290–301.

Thorp, Alicia A., Neville Owen, Maike Neuhaus and David W. Dunstan. 2011. "Sedentary behaviors and subsequent health outcomes in adults: a systematic review of longitudinal studies, 1996–2011." *American journal of preventive medicine* 41(2):207–215.

Tremblay, Mark S., Allana G. LeBlanc, Michelle E. Kho, Travis J. Saunders, Richard Larouche, Rachel C. Colley, Gary Goldfield and Sarah Connor Gorber. 2011. "Systematic review of sedentary behaviour and health indicators in school-aged children and youth." *International Journal of Behavioral Nutrition and Physical Activity* 8(1):98.

Van Der Merwe, Rudolph, Arnaud Doucet, Nando De Freitas and Eric A. Wan. 2001. The unscented particle filter. In *Advances in neural information processing systems*. pp. 584–590.

Wan, Eric A. and Rudolph Van Der Merwe. 2000. The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*. pp. 153–158.

Wang, Qifei, Gregorij Kurillo, Ferda Ofli and Ruzena Bajcsy. 2015*a*. Evaluation of pose tracking accuracy in the first and second generations of Microsoft Kinect. In *2015 International Conference on Healthcare Informatics (ICHI)*. pp. 380–389.

Wang, Qifei, Gregorij Kurillo, Ferda Ofli and Ruzena Bajcsy. 2015*b*. "Remote health coaching system and human motion data analysis for physical therapy with Microsoft Kinect." *arXiv preprint arXiv:1512.06492* .

Wang, Rong, Huiguo Zhang and Cyril Leung. 2015. "Follow me: a personal robotic companion system for the elderly." *Int. J. Inf. Technol* 21(1).

Warburton, Darren ER, Crystal Whitney Nicol and Shannon SD Bredin. 2006. "Health benefits of physical activity: the evidence." *Canadian medical association journal* 174(6):801–809.

Warren, Tatiana Y., Vaughn Barry, Steven P. Hooker, Xuemei Sui, Timothy S. Church and Steven N. Blair. 2010. "Sedentary behaviors increase risk of cardiovascular disease mortality in men." *Medicine and science in sports and exercise* 42(5):879.

Webster, David and Ozkan Celik. 2014. "Systematic review of Kinect applications in elderly care and stroke rehabilitation." *Journal of neuroengineering and rehabilitation* 11(1):108.

WHO, World Health Organization. 2015. "Factsheets on health-enhancing physical activity in the 28 EU Member States of the WHO European Region, 2015.". `http://www.euro.who.int/en/health-topics/disease-prevention/physical-activity/country-work/factsheets-on-health-enhancing-physical-activity-in-the-28-eu-member-states-of-the-who-european-region` [Accessed: 06 June 2017].

Wilmot, E. G., C. L. Edwardson, F. A. Achana, M. J. Davies, T. Gorely, L. J. Gray, K. Khunti, T. Yates and S. J. H. Biddle. 2012. “Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis.” *Diabetologia* 55:2895–2905.

Young, William, Stuart Ferguson, Sebastien Brault and Cathy Craig. 2011. “Assessing and training standing balance in older adults: a novel approach using the ‘Nintendo Wii’Balance Board.” *Gait & posture* 33(2):303–305.