Human movements evaluation using depth sensors

Carlos Palma

March 26, 2017

Human Movements Evaluation Using Depth Sensors

Carlos Fernando Palma Roldan

In partial Fullfillment of the requirements for the degree of
Masters in Engineering

Advisors: Prof. Jesus Francisco Vargas Bonilla, PhD
Prof. Augusto Salazar, PhD

Field: Signal Processing Research Group: Sistemas Embebidos e Inteligencia Computacional -SISTEMIC-

Universidad de Antioquia
Facultad de Ingenieria
Medellín,Colombia
2017

This work would have never been possible without funding provided by Sistema General de Regalías de la Nación through the project Plataforma tecnológica para los servicios de teleasistencia, emergencias médicas,seguimiento y monitoreo permanente a los pacientes y apoyo a los programas de promoción y prevención2013 a 2015,. Ruta N (Regalías de la Nación), número del convenio: 512C-2013. Código SUI (Viceinvestigaciones UdeA): 20139080

I would like to thank my advisors professor Jesús Francisco Vargas Bonilla and professor Augusto Enrique Salazar Jiménez. Finally I would like to thank my family for its continuous support during my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 2008 the United States department of health and human services published directions to provide information about the amount of physical activity recommended to keep health and physical aptitude. For substantial benefits, the guidelines recommend that the majority of elderly people take part in at least 150 minutes of moderate aerobic activity, and 75 minutes of vigorous aerobic activity, or an equivalent combination of both each week. The elderly should also engage in strengthening activities involving all of the most important muscular groups at least two times a week. Elderly people might be at risk of falling, and in many cases such falls cause them important injuries.This population at risk of falling can prevent falls by performing exercises to improve or keep balance [1].

Providing these people with tools to exercise at home can greatly improve the quality of their lives, this however presents a challenge from the engineering point of view, since it becomes necessary to develop "a software artifact that observes the user's execution of an activity and compares it to a specification."[2]. Such a system should be capable of observing the execution of an activity by the user, obtaining relevant information to characterize the performance of the user, and then use a technique to perform the decision of either accepting or rejecting the execution based on criteria previously learned either from repetitions of the activities or from a formal specification (model). This verification can be used to inform the user about the mistakes he is making and help him adjust his movement so that it may be closer to the specification developed by experts in physical therapy and conditioning, resulting in less risk of injury for the user when performing the movements and increased benefits for his health such as the increasing of his strength and his stamina.

1.1 Context

Physical activity is necessary for the well being of human beings, and this is particularly true in the case of certain populations, such as the elderly. These people have obstacles to exercise due to the lack of appropriate scenarios and dependency on other people for transportation or guidance. It makes sense then, trying to reach these people with alternatives to exercise that must comply with certain basic criteria: ease of use, non invasiveness, low risk of injury for the user. If a system is to be developed to meet these criteria it becomes necessary to know precisely the specification of the movements to be performed, maximizing the benefits to the users health and minimizing the risk of injury.

1.2 Human Movement in Therapy and Physical Conditioning

1.2.1 Movement Planes

For a complete description of human movement directional terms are used:

Superior and *inferior*, which describe being closer to the head or the feet, respectively.

Anterior means towards the front of the body and *posterior* towards the back of the body. *Medial* or *lateral* indicate positions or movements towards and getting farther from the body.

A cardinal plane is one that divides the body in two equal parts. The three cardinal planes of movement are the sagittal, frontal and transverse plane. These planes are orthogonal, or perpendicular to each other. The sagittal plane, also called median is vertical and divides the body into right and left halves. The frontal plane (also called coronal or lateral) is another vertical plane, but divides the body in anterior and posterior halves. The transverse plane (also called horizontal plane) divides the body into superior and inferior halves.[3],as is shown in Figure 1.1

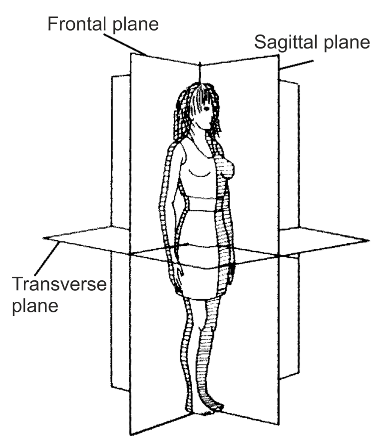


Figure 1.1: Movement planes used in human motion analysis

1.2.2 Types of Movement

Most of the descriptions of movements are based on what is called the anatomical position, this means a person standing fully erect, with the arms relaxed on the sides of the body and the feet and fingers together[3] as shown in figure 1.1

Depending on the plane on which the movement occurs, different terms are used to classify it.

Abduction is a motion in a frontal plane that moves the segment away from the anatomical position. *Adduction* is a frontal plane motion that returns the segment to the anatomical position. *Flexion* is a motion in a sagittal plane, that moves the limb away from the anatomical position. The angle that should be visualized in this convention is the angle formed by the limb of interest and the frontal plane, while in the anatomical position. *Extension* returns the segment to the anatomical position in a sagittal plane and is described as increasing the angle at the joint.*Internal rotation* at the shoulder (also called inward or medial rotation) is motion of the arm segment that rotates it from the palms-forward anatomical position to a posture in which the palms are facing more medially and finally posteriorly.*External rotation* (outward or lateral) is the opposite motion in which the segment is returned to or beyond the anatomical position in a transverse plane. [3]

1.3 Artificial Vision in Human Motion Analysis

1.3.1 State of the Art

Several different technologies exist that allow for the tracking and characterization of movement in human beings. One of these technologies is the use of IMU(Inertial measurement units), these are sensors that integrate a three axis accelerometer, a three axis gyroscope and sometimes a three axis magnetometer. Some of these systems are capable of tracking the full body of a person, such as Noitom industries *Perception Neuron*[4]. The main downside of this technology is that wearing the sensors all over the body can be uncomfortable. Other technology that is very prominent in the field of understanding human motion is Motion Capture (MoCap), these systems make use of a set of markers put on important locations in the human body, usually on a suit that a person needs to wear. The subject wearing the markers is placed in a room that is equipped with several different cameras, and a software outputs the coordinates of the markers, with millimetric precision. The main downside of this technology is its high cost.[5] Markerless motion capture systems also exist that allow for the tracking of body parts of a human being without the need of a special suit, these systems however require a space without any object on it and an arrangement of cameras.[6]. Again the main limitation to the use of these technology is its high cost. In recent years the use of depth cameras for the task of human motion tracking has gained popularity, because these sensors offer reasonable precision at a price that is much less than other motion capture system[7],[8].

In the case of video signals, several techniques use video signals as inputs and try to recognize movements of limbs in human beings. One of them is MHI (Motion History Image) and it is used by Javed et al [9]. The technique consists in keeping a set of segmented images of a human being corresponding to the different phases of the movement that wants to be observed, and calculating gradients on the edges in order to determine a general direction of motion. Other approaches use algorithms that build Markov chains with a given probability distribution as stationary distribution, these algorithms are known as Markov Chain Montecarlo (MCMC),Siddiqui et al [10] use them in data coming from a depth sensor, and train an algorithm to recognize individual body parts, the system however incurs in a high computational cost.

Sucar et al [11] use a system comprised of two RGB cameras placed perpendicularly in order to capture the trajectory of the hand of a human being in three dimensions, such system is proposed as a tool for rehabilitation of patients who have suffered strokes. A disadvantage of such system is that the tracking of the hand is based on color segmentation, meaning that any object similar to a user's hand is erroneously detected and its position processed.

Burke et al [12] develop an exercise system for rehabilitation of patients who suffered a stroke using a single RGB camera, color segmentation and tracking of a persons hand.

The use of depth sensors has greatly expanded in the last years. This technology allows for the calculation of the distance at which an object stands in front of the camera. This information facilitates the segmentation process on images, and allows for estimation of the position of parts of the body, which can be specified by x,y,z coordinates. The most widespread of such systems is the *Microsoft kinect*, a low cost sensor that incorporates a vga camera, a depth camera and an array of multiple microphones, it is also equipped with an embedded algorithm that outputs coordinates of 20 relevant body parts at a rate of 30 frames per second, the algorithm was developed by Shotton et al [13].

Dutta et al [14] evaluated the precision of the *kinect* sensor by comparing with a vicon system, and found out that the average error made by the *kinect* was comparable with the error associated to the movement of the markers commonly used for Mocap systems when they move over the skin of the user, this means that the sensor can be used to estimate posture in human beings.

Kurillo et al. [15] compared the precision of the *kinect* sensor with a movement capture system in the task of detecting the frontiers of the volume in which the upper limbs can move, and arrived to the conclusion that *kinect* data are as precise and trustworthy as MoCap data and are promising for clinical evaluation of upper limb movements applied to neurologic or musculoskeletal illnesses.

Obrdzalek et al [16] studied the precision of the data coming from the sensor, by comparing it to MoCap systems, which use markers attached to the body to determine the 3D position of important body parts of a person's body. In such study they arrive to the conclusion that for controlled positions (standing in front of the sensor) the *kinect* offers a precision similar to that of the systems used for motion capture. They also found that estimations would be improved if an anthropometric model were to be applied to the data, because in each frame the skeletonization algorithm embedded in the device calculates the position of relevant points ,in such a way that an error of 10 cm can be made when calculating the length of the limbs if the person has rotated more than 45 degrees with respect to the sensor. A similar study was conducted by Fernández-Baena [17], it found out that the precision of the *kinect* sensor when estimating the position of the main joints in the body is enough to make it a useful tool for rehabilitation treatments. Zhao et al.[18] study the application of the sensor to the evaluation of a series of physical activities and compare it to a cortex system composed of 8 cameras and arrive to the conclusion that in the absence of severe occlusions the systematic errors introduced by the skeletonization algorithm can be compensated by adjusting the decision rules used for the activities.

Bonnechere et al[19] compared the repeatability of the measurements used by *kinect* and a motion capture systems, finding that the intra-class correlation coefficients were very similar for captures made on the same day and different days, which indicates that the repeatability is similar for both the low cost system and the MoCap system.

The same author [20] compared the precision of the *kinect* sensor and a stereophotogrammetry system for estimation of the morphology of a human being when measuring height, and the length of relevant segments of the human body, like legs and arms. This study found that the differences between both systems were systematically correlated and allow for the use of regression equations to correct the *kinect* results. The conclusion of the study was that systems based on *kinect* sensors can be used for quick estimation of morphology in human beings.

Van Diest et al [21] conducted a comparison study between *kinect* and a Vicon 3D system in the case of movements of the body, and found that using Principal Component Analysis (PCA) both systems captured 90% of the observed variability using the first three principal components. The study arrives to the conclusion that *kinect* can be used as a tool to measure balance in human beings.

The *kinect* was used successfully by [22] to recognize signals made with flags that represent characters in the japanese alphabet. A similar work was carried out by Zahoor et al [23] to recognize characters of the american sign language, in this study they used Hidden Markov Models, which are basically Markov chains for which on every state there is a certain probability of emitting one of several possible symbols.

Other studies have used the technology to recognize movement sequences instead of fixed positions. Chaves et al[24] use this information to generate state vectors, with the angles formed by the joints, these sequences are then compared with the observed sequences in order to recognize specific movements. Liu et al [25] use a similar method, generating a vector model invariant to translations or anatomical variations. Similarity is quantified by calculating a sum of angular differences with appropriate weights to compensate visual differences in the perception of movement, the decision is finally taken by a neural network.

Movement sequences of the hand are of particular interest, Jaemin et al [26] use the tracking provided by the device to isolate the silhouette of the human hand and recognize gestures using an HMM that outputs a mixture of gaussians. An application of these techniques is found in control theory, Ye Gu et al [27] report 89% of success in recognition of hand movements for control of a vehicle.

Coordinate data provided by *kinect* might be affected by considerable noise, due to tracking errors and occlusions, Chaaroui et al [28] have used silhouette data obtained with *kinect* to improve the recognition of actions in these cases, by mixing these data with coordinate data. Bo et al [29] use data from inertial sensors and kalman filters to improve the measurements taken by the *kinect*.

In the field of therapy and physical conditioning, numerous works have been developed that use the *kinect* to create tools for patients who have suffered strokes. Chang et al [30] report the development of an exercise system for physical therapy aimed to be used by children with muscular dystrophy and cerebral palsy.Pastor et al [31] report the development of an application to help with the rehabilitation process of patients who have survived a stroke. A similar tool has been proposed by Acosta et al [32]. A complete survey of the multiple systems developed with the sensor for rehabilitation can be found in [33], where the systems are classified into two categories: the ones who have been clinically reviewed and the ones who have not. What is common to the vast majority of these systems is their use in rehabilitation of patients who suffer from some musculoskeletal or neurological disorder.

There are also commercial applications specifically made for physical conditioning, but they are aimed to be used by healthy people, as is the case of the videogame *Nike Kinect Training*[34], which uses the capabilities of a console to process the *kinect* data and determine whether a person is executing an athletic activity. Sato et al [35] report the development of exergames (exercise oriented games) directed specifically to the elder, using *kinect*, and report the usefulness of the system to improve functions like strength and the gait parameters in this population. An application also aimed at the elder population that offers stepping exercises and simultaneously measures stepping parameters useful for predicting falls in elder people was developed by Garcial et al [36]. Kayama et al [37] prove that the use of an application called dual-task Tai-Chi has positive effects on indicator associated to the risk of falling.

Ravi [38] develops a recognition system focused in physical therapy activities that provides real time feedback using *kinect*. The system tracks all of the joints and uses quaternions to analyze if the movement was correctly executed, by using a technique called Spherical Linear Interpolation(SLERP).

Paiement et al [39] propose a method to determine based on skeleton data provided by the device whether the movements of a person going upstairs deviate from normality. They use a dimensionality reduction method based on non-linear manifolds, and then create two statistical models, one for the fixed poses that are observer while the person executes the movement and another to model the dynamics of the movement. Both models are generated using Parzen windows, which are non parametrical methods to estimate the probability density function of a single random variable.

Agathocleus [40] report the development of a system for physical training, using a kinect sensor, and performs real time validation of the user movements using a model based on finite state machines. A user must pass through all of the states if the movement is to be recognized as completed. Maung et al [41] define clearly an XML grammar associated to the recognition of movements relevant for physical therapy and implement it too, using a model based on a finite state machines, where each movement to be recognized is one such state machine and these are executed in parallel.

Velloso et al[42] explore the use of Programming by Demonstration (PbD) techniques to develop a system that determines a model for movements based on a demonstration performed by an expert and that can determine whether a new user executes the movement closely enough to the demonstration. The authors use a *kinect* sensor and characterize the movements in terms of the angles formed with the coordinate planes commonly used in kinesiology [43].

Da Gama et al [44] report the development of a system for movement recognition for physical therapy using *kinect*, in this work the authors calculate vectors normal to the movement planes defined in physical therapy and kinesiology, to make sure that the movement takes place in a single plane. Kitsunezaki et al [45] evaluate the precision of a similar system to measure angles in the human body by comparing them with manual measurements taken by an expert with a goniometer. The angles calculated by the *kinect* are obtained estimating a vector normal to the body using the data provided by the device.

After reviewing the approaches in the literature we identify two basic approaches: the use of algorithms of artificial intelligence and the use of movement models, usually implemented as state machines. In general the first approach works well for the problem of recognizing gestures coming from a dictionary, while the second approach is favored when the problem is evaluating the performance of a subject when executing a movement. This happens for two reasons, the first one is the impossibility of offering real - time feedback to the person using the system by using machine learning algorithms, since these algorithms only reach a decision once the whole activity has been performed, and the second one is the fact that the use of state machines facilitates increasing the number of movements to be evaluated without having to collect an extensive database.

1.4 The Kinect Sensor ™

The *kinect* sensor is an array of sensors comprising a RGB camera, a depth sensor and an array of microphones. The depth sensor allows acquisition of images in varying light conditions, thanks to an infrared projector and a monochromatic CMOS sensor. The sensor can recognize up to six people but can only skeletonize two of them.

1.5 Objectives of this work

This works will contribute to the field of automated movement evaluation by :

1. Creating a database of relevant physical therapy and conditioning exercises with real subjects executing the movements.
2. Proposing a set of characteristics that can be used to define such movements.
3. Proposing a methodology by which movements may be considered to be correctly or incorrectly executed.
4. Implementing a software prototype to characterize and evaluate the performance of a human being when he tries to execute the movements.

1.6 Summary

This chapter discussed the need for people to exercise in order to improve the quality of their lives, and how computer vision technology can provide them the means to accomplish this. It reviews several different technologies that exist to address the issue of accurate human movement tracking, to focus on a low cost, reliable alternative, the *Microsoft kinect*. It reviews proposals to apply this technology for physical therapy and conditioning, for persons ranging from severely-ill patients to healthy adults. Finally it presents some of the characteristics of the sensor itself, that will be used as the main acquisition device for the discussions presented in the next chapters.

Chapter 2

Databases and characteristics used in this study

The following is the description of the databases that were collected during the present study.

2.1 Database Description

2.1.1 Normal activities database

First a database was built which consisted of repetitions of the activities as close to the specification detailed by the expert in physical therapy as possible. The number of participants was 14. They were asked to perform three repetitions of each of the ten activities of interest.

2.1.2 Dual Kinect database

Experiments performed with an system consisting of two kinects were carried out with 6 participants. They were asked to perform each one of the ten activities of interest three times. This was repeated for each one of the four angles formed between the kinects and each one of the four possible orientations of the persons with respect to the sensor. In order to have samples of the movements according to the specifications of the expert,

2.1.3 Errors database

For each one of the ten activities of interest two deviations from normality were defined. Ten people were asked to perform each one of the deviated movements ten times, meaning that a total of 100 samples of erroneous movements were simulated for each one of the defined deviations. The deviations are described below:

For abduction movements the simulated errors were:
error1: moving the limb away from the frontal plane towards the front of the body.
error2: moving the limb away from the frontal plane towards the back of the body.

For extension and flexion movements the errors were:
error1: moving the limb with respect to the sagittal plane away from the body.
error2: moving the limb with respect to the sagittal plane towards the body.

For rotation movements the errors were:
error1: moving the limb deviating from the transverse plane towards the floor.
error2: moving the limb deviating from the transverse plane towards the ceiling.

2.2 Set of Activities

In the following we show the set of activities to be evaluated, selected in agreement with the physicians who adviced the development of this work. The criteria for selecting the activities were: the order in which a person normally executes a strengthening routine, beginning with the upper limbs and following with the lower limbs, and the use of the most important muscle groups of the limbs.

In chapter Real Time Kinect Evaluation of Therapeutical Gestures two databases were collected:

- A database was taken comprising the following ten physical activities: shoulder abduction, hip abduction, shoulder flexion, shoulder extension,hip flexion, hip extension, shoulder internal rotation, shoulder external rotation, elbow flexion and, elbow extension. The database aims at providing data to compare the performance of a dual kinect array with a single kinect. Two parameters are varied, the first one is the angle formed by the array of two kinects used for data acquisition, the second one is the orientation of the subject with respect to a single sensor. Three repetitions of each activity were recorded.
- A database was taken with 14 people performing each one of the ten physical activities three times. This time a single sensor is used, and only the orientation of the person with respect to the sensor is varied.

In chapter HMM and DTW for Evaluation of therapeutical gestures using kinect a single database is taken:

- A database of common errors in the performance of the activities. Two kinds of errors are considered for each type of movement: for abduction

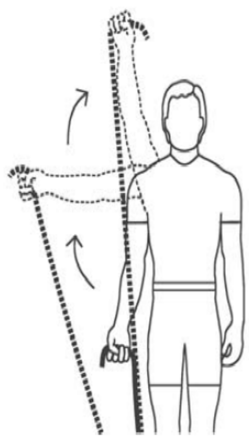
Graphical Description	Detailed Description	Comments
 [46]	Begins in the anatomical position, the elbow must not be flexed. The arm is raised until a 90 degree angle is formed with the body. The arm is slowly returned to the initial position.	The user must stand straight. Unlike the picture, people are asked to perform the movement until an angle of 90 degrees is formed to prevent injuries.

Table 2.1: Shoulder Abduction

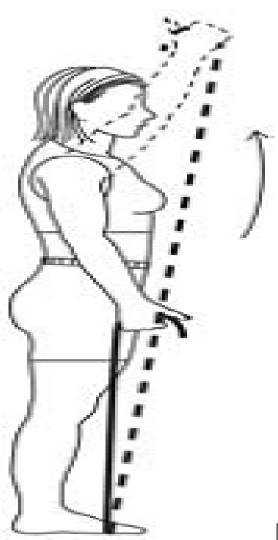
Graphical Description	Detailed Description	Comments
 [46]	Begins in the anatomical position, the elbow must not be flexed and the thumb must me up. The arm is raised until a 90 degree angle is formed with the body. The arm is slowly returned to the initial position.	The graph varies from the actual exercise suggested by the expert in physical therapy in that the expected angular variation in our case is 90 degrees.

Table 2.2: Shoulder Flexion

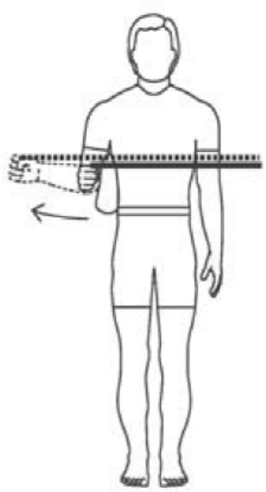
Graphical Description	Detailed Description	Comments
 [46]	Starts in the anatomical position. The elbow is flexed to form a 90 degree angle, this is considered the initial position. The hand is rotated outwards, around the body. The hand is slowly to the initial position.	The elbow must remain in contact with the body during the performance of the activity. This exercise is performed with both hands simultaneously as recommended by the expert in physical therapy.

Table 2.3: Shoulder External Rotation

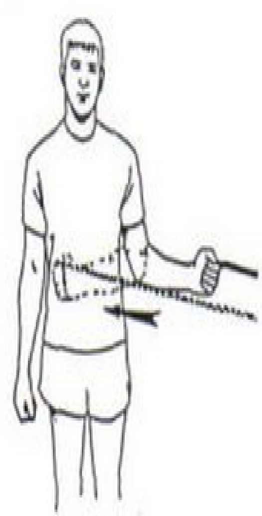
Graphical Description	Detailed Description	Comments
 [46]	Starts in the anatomical position. The elbow is flexed to form a 90 degree angle, this is considered the initial position. The hand is rotated towards the center of the body, around it. The forearm must remain parallel to the floor. The limb is slowly returned to the initial position.	The expected angular displacement is 30 degrees.

Table 2.4: Shoulder internal rotation

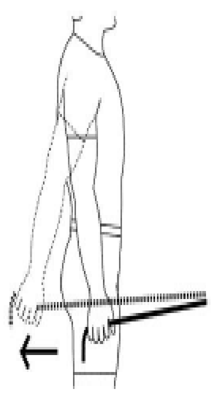
Graphical Description	Detailed Description	Comments
 [46]	Starts in the anatomical position, the elbow must not be flexed. The arm is moved towards the back of the body keeping the elbow extended. The arm slowly returns to the initial position.	The user must stand straight.

Table 2.5: Shoulder Extension

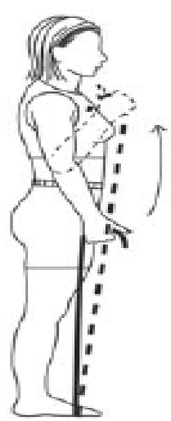
Graphical Description	Detailed Description	Comments
 [46]	Starts in the anatomical position, with the elbow extended. The elbow is flexed until the hand almost reaches the shoulder. The elbow slowly returns to the initial position.	While the activity is being performed, the elbow must remain on one side of the body, still.

Table 2.6: Elbow Flexion

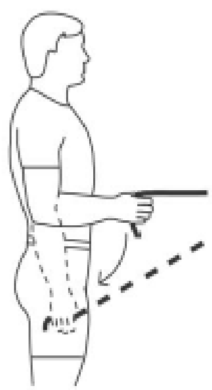
Graphical Description	Detailed Description	Comments
 [46]	Begins in the anatomical position. The elbow forms a 90 degree angle. This is considered the initial position for the activity. The forearm is lowered until it stays on the side of the body. The forearm is slowly returned to the initial position.	While the activity is performed the elbow must remain still on the side of the body.

Table 2.7: Elbow Extension

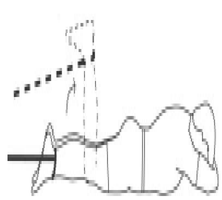
Graphical Description	Detailed Description	Comments
 [46]	Starts with the person standing straight. The knee is flexed until it forms a 90 degree angle. The leg is returned to the initial position.	Unlike the figure, the user performs the activity standing.

Table 2.8: Knee Flexion

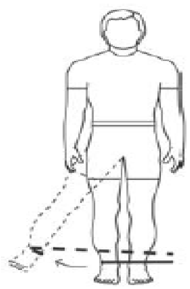
Graphical Description	Detailed Description	Comments
 [46]	Stars with the user standing straight. The leg is moved towards the side of the body until a 45 degree angle is formed. The leg is returned to the initial position.	The leg must be extended during the whole performance of the activity.

Table 2.9: Hip Abduction

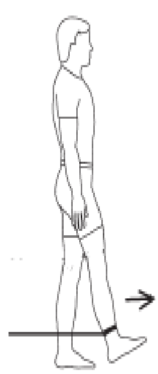
Graphical Description	Detailed Description	Comments
 [46]	Starts with the person standing straight. The leg is moved towards the front, while keeping the knee unbent. The leg is returned to the initial position.	The user must stand straight during performance of the activity.

Table 2.10: Hip Flexion

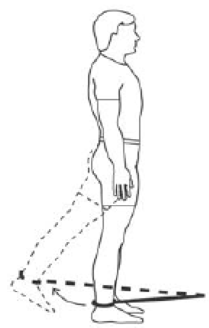
Graphical Description	Detailed Description	Comments
 [46]	Stars with the user standing straight. The leg is moved towards the back of the body, while the knee is unbent. The leg returns slowly to the initial position.	The user must stand straight during the whole performance of the activity.

Table 2.11: Hip Extension

movements, error type 1 is defined as moving the limb of interest towards the front of the body. Error type 2 consists in moving the limb towards the back of the body. For flexion and extension movements error type 1 is defined as moving the limb away from the body, and error type 2 consists in moving the limb towards the body. For rotation movements error type 1 consists in moving the limb keeping it pointing towards the floor, and error type 2 consists in moving the limb while it points upwards. 10 people took part in the study, each one performing 10 repetitions of each kind of mistake, for a total of 200 different erroneous sequences.

In chapter Extension to continuous HMM and comparison between different populations a new database is introduced:

- 8 people between the ages of 65 and 75 were asked to perform the 10 activities of interest in front of the sensor, repeating each one three times. This database was used to determine whether measurements for this population varied significantly from measurements taken from younger subjects.

2.3 Characteristics used in this study

The experiments presented in the present work used two sets of characteristics:

- The first ones are the coordinate measurements taken by a kinect sensor and normalized according to the height of the subject.

- The second set of characteristics are the angles formed by the limb of interest with each of the planes of motion,as defined in chapter Human

Movement in Therapy and Physical Conditioning. Both sets of characteristics are used and compared throughout the following chapters, in order to determine which one offers the best accuracy when analyzing performance in human movement.

Chapter 3

Real Time Kinect Evaluation of Therapeutical Gestures

Results showing the comparison between the use of a dual kinect array and a single sensor to evaluate movement sequences were published in the 20th Symposium on Signal Processing, Images and Artificial Vision, STSIVA-2015, that took place in Bogotá, Colombia in September 2-4, 2015.[47]

3.1 Introduction

In this chapter we study the repeatability of the angles formed with estimates of the frontal, sagittal and transverse planes for limbs of interest across a wide set of movements defined by an expert in physical therapy. We also assess the robustness of these measures relative to rotations of the person with respect to the sensor, already several works have compared the tracking of the device with marker based systems used as ground truth, proving that it is acceptable for rehabilitation systems.[17],[48],[49]. In this chapter we focus on comparing the results yielded by a stereo system with a single *kinect*, to determine whether there is an improvement for the characteristics of interest. We propose a software implementation to evaluate the conditions determined by an expert in physical therapy and test it for the case of shoulder abduction.

3.2 Related Work

The use of several *kinect* sensors as tools for reconstruction of a 3d full human body was studied by Tong et al. [50], the authors use 3 kinect sensors, one to scan the upper part of the body, one for the lower part and one for the middle part, an propose a method of non-rigid alignment to obtain a 3D model of the user. Jo et al [51] develop a system based on an array of four kinect sensors aimed at tracking the motion of a single individual in scenarios where various people are moving, they focus their attention however on tracking the hip center of each person. Azis et al.[52] propose the use of two kinect sensors for the task of human action recognition. One of the sensors is located in front of the user and another one sideways. The strategy used to determine the optimum position of the 3d joints consists in replacing the coordinates of those joints with non-tracked as tracking state with the coordinates of the joints given by the other sensor converted to the common frame of reference. Yeung et al [53], use an optimization which consists in minimizing the sum of the differences of the joint positions with the estimated positions subject to the condition that the bone lengths must remain constant.

3.3 Estimation of the planes of motion in real time

Following the ideas presented in [44] and [45], a method is developed to estimate vectors normal to the movement planes.

The Kinect Software Development Kit (SDK) provides the user with the coefficients of the equation of the floor plane in the form of a vector [A,B,C,D] such that $AX + BY + CZ + D = 0$. This vector is used to obtain a vector perpendicular to the transverse plane.

$$V_{transverse} = (A, B, C, D). \tag{3.1}$$

It is important to note that this vector is directed from the floor towards the ceiling.

A vector normal to the frontal plane is calculated by taking the cross product of two vectors, one connecting the left shoulder to the shoulder center, and one connecting the left and right shoulders. This is done to ensure that the direction of the cross product is always from the person towards the sensor.

$$V_{frontal} = (p_{sc} - p_{ls}) \times (p_{rs} - p_{ls}) \tag{3.2}$$

Where p_{sc}, p_{ls} and p_{rs} are the coordinates of the shoulder center, left shoulder and right shoulder as given by the sensor.

Finally a vector normal to the sagittal plane is calculated using the vectors normal to the transverse and frontal planes. This vector is, according to the definition of the previous vectors, oriented towards the left side of the body.

$$V_{sagittal} = V_{frontal} \times \ V_{transverse} \tag{3.3}$$

The calculation of the angle formed with each plane of motion is finally done by taking a scalar product of these vectors and the vector of the limb that represents the movement better.

$$Frontal = 90 - \arccos \frac{limb \cdot V_{frontal}}{\|limb\| \, \|V_{frontal}\|} \tag{3.4}$$

$$Transverse = 90 - \arccos \frac{limb \cdot V_{transverse}}{\|limb\| \, \|V_{transverse}\|} \tag{3.5}$$

$$Sagittal = 90 - \arccos \frac{limb \cdot V_{sagittal}}{\|limb\| \, \|V_{sagittal}\|} \tag{3.6}$$

Where limb is the body segment in which we are interested.

3.4 Use of two kinect sensors to improve skeletonization

In order to obtain robust measures for the angles formed between the limb of interest and the planes of motion, the use of two *kinect* sensors simultaneously was explored, so that the loss of information due to inferred or not tracked states by one device would be compensated by the information coming from the other one. We start by applying a rotation to the joint positions provided by the Kinect SDK. The rotation matrices relating positions in the coordinate systems of Kinects K_A and K_B are found by using the Moore-Penrose pseudoinverse.

$$R_B = A(B^T B)^{-1} B^T. \tag{3.7}$$

$$R_A = B(A^T A)^{-1} A^T. \tag{3.8}$$

These equations suppose that the Kinect skeletons on both coordinate systems are expressed using homogeneous coordinates; A is the skeleton coordinates on the system of Kinect A and B the set of coordinates on the system of Kinect B. R_B is the matrix that allows us to rotate a skeleton in the coordinate system of Kinect B to the coordinate system of Kinect A.

The idea of having two skeletons is then to determine based on the information of these sources the most probable locations of the points that form a person's skeleton. This is what we will call the real skeleton. This skeleton should be close to both of the skeletons yielded by each one of the devices, and in the cases were there was an incorrect tracking, then we should trust the information provided by one device more than the other. If p_i^* is the vector representing the position of a joint in the new improved skeleton and p_i , q_i are the positions given by the sensors K_A and K_B, and w_i^A and w_i^B are weights assigned depending on the tracking state of the joint, then we have to rotate the coordinates given by the sensor K_B and translate them so that they become coordinates in the system of coordinates of K_A, and then we have to minimize the distance of the improved skeleton to each of the skeletons. This can be formulated as an optimization problem:

$$min \sum_{i \epsilon S_A} w_i^A ||\boldsymbol{p_i^*} - \boldsymbol{p_i}||^2 + w_i^B ||\boldsymbol{p_i^*} - (\boldsymbol{R q_i} + \boldsymbol{t})||^2 \tag{3.9}$$

Taking the partial derivatives and solving the corresponding linear systems the following closed solution is found:

$$p_i^* = (w_i^A p_i + w_i^B (R q_i + t))/(w_i^A + w_i^B) \tag{3.10}$$

The weights are calculated according to the ideas exposed in [53]. Once the optimization is done, the resulting skeleton is used to calculate the angles formed by the limb of interest with respect to the three planes of motion: frontal, transverse and sagittal. In order to deal with the noise in the calculated angles, a median filter is implemented, since threshold based recognition is severely affected by outliers. Empirical tests show that a length of 10 samples is enough for the smoothing of the signal.

3.5 Noise Reduction

It is well known that the coordinate data produced by the sensor are affected by noise, which appears as low intensity noise in all measurements and as noise peaks for the instants in which the joint is no longer being tracked [54]. The first type of noise affects all quantities derived from measurements taken by the device, due to the random error propagation when applying arithmetic operations or trigonometric functions. The second form of noise greatly affects the calculated quantities (like the vectors normal to the coordinate planes) in those specific frames in which the tracking of the joints is not reliable. An example of the noise found in the signals can be found in figure 3.1, this depicts the calculated frontal angle for an instance of the shoulder abduction movement, and the result of using kalman filtering on such angle.

Sources in the literature report the use of Kalman filters to tackle the problem of the noise peaks appearing in the signal [55] [56]. In order to improve the precision of the system, kalman filtering is used for the measured angular characteristics.

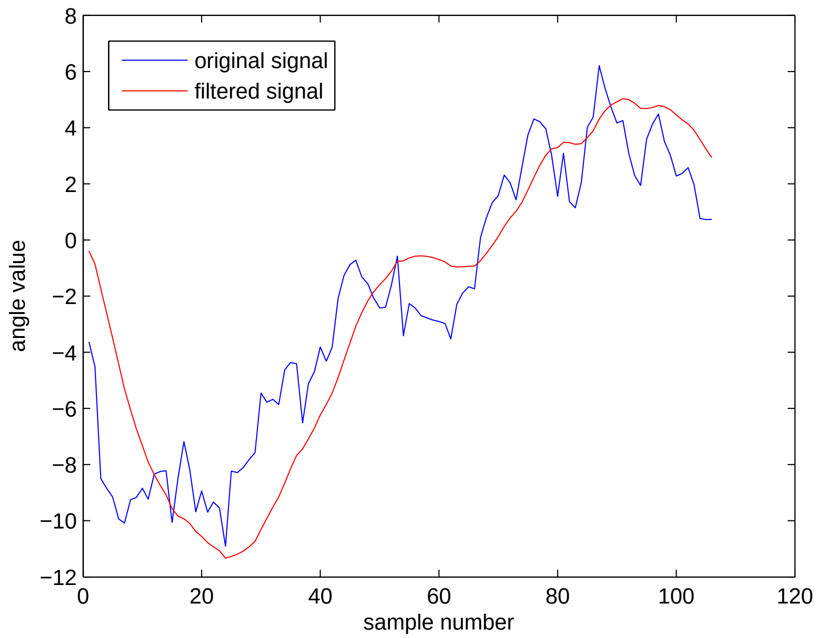


Figure 3.1: Kalman filter applied to data given by the sensor

3.6 Framework for action evaluation

The strategy implemented to perform evaluation is based on a priori knowledge of the correct form of the activities of interest. This knowledge allows for recognition of essential characteristics like the number of phases of movement to consider, and initial and final positions for each phase.

3.6.1 Evaluation of movement

Our approach follows the ideas presented in [57],but instead of defining for each movement a different state machine, considers each relevant movement to be composed of three stages, which are: an initial movement, a stationary stage in which the user is supposed to hold the limb in position for a certain amount of time, and a final movement, to go back to the initial position. This is done according to the specification of an expert in physical therapy, who accompanied us in the development of this project and considers that in order to strengthen the muscles in the limbs the final position of the movement must be sustained.

Each one of the states implements another finite-state machine, which consists of four states, a waiting state that checks for the initial position of that specific phase of the gesture, an idle state in which the initial position has been detected but real-time evaluation has not been triggered, an execution state in which the movement takes place and a final state in which the final position for the movement has been detected. An error occurring in any of the states implies going back to the start of the sequence an sending a message to the user. An error in any of the states triggers a return to the initial state.

3.6.2 Evaluation of a simple action

A simple action can be one of the basic movements defined in biomechanics (see section 1.2.2). In order to evaluate the performance of this action a finite state machine is used. The state machine recognizes the initial position; makes sure that the movement has actually started and validates two conditions that must be true while the action is being performed, that the movement occurs approximately parallel to one of the movement planes and that the movement is taking place in the desired direction; the final condition to be checked is that the person has reached the final position.

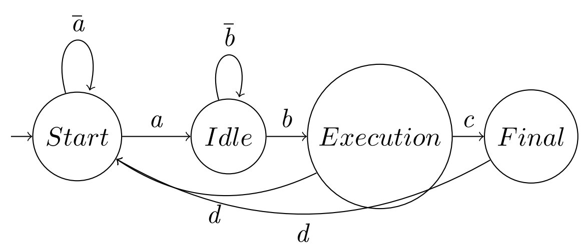


Figure 3.2: State Machine used for recognition of a simple action.

In Figure 3.2 the variable a indicates whether the user has positioned himself in what is considered to be the initial pose for the gesture, determined by ranges of the angles of the limb of interest as well as angles that check the standing position. Variable b is used to determine when the user has started moving, by checking the angle that is supposed to vary the most during the execution of the gesture. Variable c indicates when the user has reached the final position of the movement, and when this happens the general state machine experiences a transition to the next state. d represents the occurrence of an error during recognition, which causes the whole system to reset.

3.6.3 Experimental calculation of thresholds

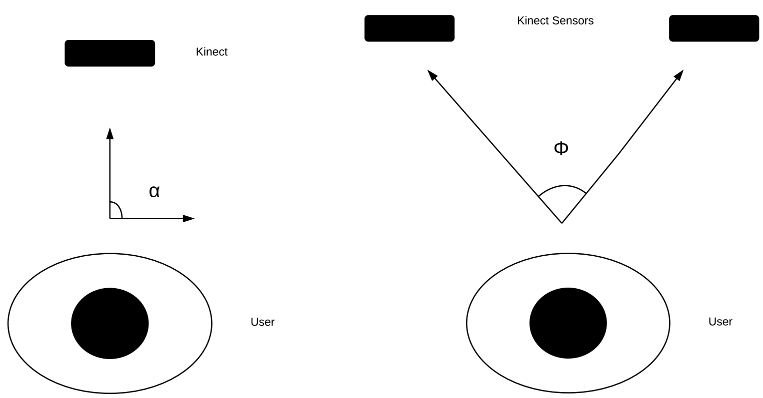
Besides random noise present in measurements taken by the device, an important source of variability to be taken into account when implementing activity models is the morphological and functional variability present in the human population. This variability is hard to model, making it necessary to collect samples of different people performing the activities, in order to calculate intervals for the angles of interest that define the activities that want to be evaluated.

3.6.4 Experimental setup

Experiments were conducted to compare the reliability of angle measurements taken with one and two *kinect* sensors, to determine the robustness of the measurements with respect to different orientations of the user and to test the usability of the proposed state machine based recognition scheme.

For the first experiment five people were asked to perform ten physical activities, these activities were described in chapter **??**. The activities were selected to include both the upper limbs and the lower limbs, as well as movements in each of the planes of motion, the reason for this is that a normal exercise routine includes a variety of movements happening in all planes of motion and involving several different body parts. The complete description of the activities used in this study can be found in the appendix at the end of this document.

Coordinates were recorded using three sensors, two of them forming a dual-kinect system as shown in Figure 3.3. α is the angle formed between the user and a sensor located right in front of him. The values of alpha for which the experiment was performed were α of 90° , 60°, 30° and 0°. ϕ is the angle formed between the two kinect sensors that were used to improve the skeletonization performance, the whole procedure was repeated for values of ϕ of 90°,105°, 120°, and 135°.



(a) (b)

Figure 3.3: Experimental setup for data acquisition (a) acquisition with one sensor (b) acquisition with two sensors, the acquisitions were performed simultaneously

For the second experiment fourteen people were asked to perform the same ten activities, repeating them three times. This experiment was conducted to assess the reliability of the angles measured across subjects, so a greater sample was used. Each set of activities was performed four times, varying the orientation of the person with respect to the sensor. Results exclude the case of the person standing sideways with respect to the sensor, because in such case the tracking was not reliable for the most part. The third experiment was conducted with eight people, who were asked to use the system after receiving basic instruction about the execution of the shoulder left abduction. All the experiments were performed with the sensor placed on a table 70 cm above the floor, and the subjects were asked to stand at a distance of 1.8 m from the sensors, so that their entire skeleton may be tracked by the device.

3.7 Results

This section shows comparisons of the performance of two sensors and a single sensor, we also show the variation of the proposed angles when measured across different subjects and different orientations.

The reliability of the measurements taken with two Kinect sensors and a single Kinect was compared on the basis of the success rate of the tracking. A tracking is successful if the calculated angles with respect to the three planes show distinct repetitions and low noise, as shown in Figure 3.4.

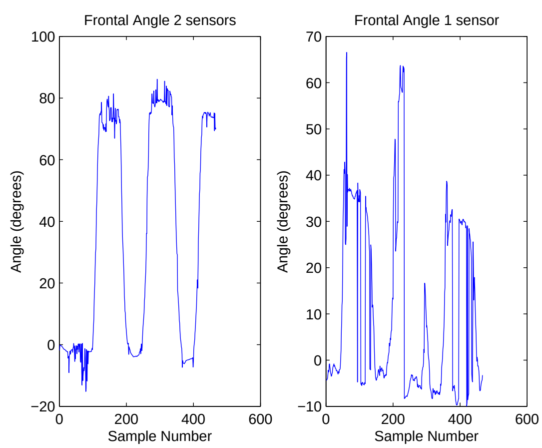


Figure 3.4: Example of both successful and unsuccessful tracking, the activity performed was shoulder flexion, with the person in front of the sensor.

The results of the first experiment are shown in Tables 3.1 and 3.2, which summarize the effect of varying the angle between the two Kinects and varying the angles of the person with respect to the sensor. The angle shown in the Table 3.2 is the angle formed between a vector normal to the frontal plane directed towards the sensor and positive x axis, when the x axis is connecting the shoulders of the user, and its positive direction is considered to be the right. This means that an angle of 90° indicates a person standing in front of the sensor, and an angle of 0° indicates a person that only shows his left side to the sensor. The errors were maximum for the 0° rotation, because one of the shoulders is occluded and the frontal plane can not be estimated accurately.

Table 3.1 shows the effect of varying the angle between the two *kinect* sensors on the success of the tracking. Each row corresponds to an activity, and the percentage of success is displayed for each one of the four angles between the sensors. Additionally the rate of success of a single sensor is presented, in order to help the comparison, the result that is higher when comparing the dual kinect system to a single kinect is highlighted.The success of the tracking with two *kinect* sensors was hardly affected by this angle. In the case of activities difficult to track like internal and external rotations, varying the angle between the sensors did not offer any significant improvement.

Activity	Sensors	Angle (°)			
		90	105	120	135
Hip	2	65	65	**70**	60
Abduction	1	**80**	**90**	60	**95**
Shoulder	2	**90**	80	80	**85**
Abduction	1	80	**85**	80	80
Hip	2	55	60	**75**	60
Extension	1	**60**	60	70	**70**
Elbow	2	**95**	**95**	90	90
Extension	1	85	90	90	**95**
Shoulder	2	**70**	90	65	75
Extension	1	65	90	**75**	75
Hip	2	45	40	40	65
Flexion	1	**50**	**60**	**55**	**75**
Elbow	2	95	85	80	90
Flexion	1	95	**95**	**90**	90
Shoulder	2	**90**	**75**	**80**	**70**
Flexion	1	60	65	55	50
Shoulder	2	**45**	**50**	20	**25**
Ext. Rotation	1	25	35	**30**	20
Shoulder	2	**30**	40	**15**	20
Int. Rotation	1	20	40	10	**60**
Average	2	**68**	68	61,5	64
	1	62	**71**	61,5	**71**

Table 3.1: Successful tracking percentage for different orientations of the two sensors

Table 3.2 shows the effect of the rotation of the person with respect to the *kinect* sensor in the success of the tracking. Each row corresponds to an activity, and the percentage of successful tracking is shown for each rotation, both for two sensors and a single sensor, the best result between the dual kinect system and a single sensor is highlighted. Improvements when using two sensors were found for shoulder flexion. The one sensor approach yields better results when trying to track movements of the lower limbs.

Activity	Sensors	Angle (°)			
		90	60	30	0
Hip	2	90	80	75	25
Abduction	1	**100**	**85**	**90**	**70**
Shoulder	2	90	90	**95**	**60**
Abduction	1	**100**	**100**	75	35
Hip	2	75	70	**70**	35
Extension	1	**85**	70	60	**45**
Elbow	2	85	95	100	**85**
Extension	1	85	**100**	100	80
Shoulder	2	80	**95**	**90**	**55**
Extension	1	**100**	**90**	**85**	*50*
Hip	2	80	80	35	0
Flexion	1	**90**	**100**	**50**	**20**
Elbow	2	100	95	90	65
Flexion	1	100	**100**	**95**	**75**
Shoulder	2	**95**	75	**90**	55
Flexion	1	5	**90**	75	**60**
External	2	35	**35**	**35**	5
Rotation	1	**55**	30	30	**20**
Internal	2	**40**	**40**	20	5
Rotation	1	20	30	**40**	**10**
Average	2	**77**	75,5	70	39
	1	74	**79,5**	70	**46,5**

Table 3.2: Successful tracking percentage for different orientations of the person with respect to the sensors

The means of the angle variation measured for each activity are presented in Table 3.3. These are calculated by taking the difference between the mean of the angle formed in the final stage of the activity and the mean of the angle formed in the initial stage of the activity.

Activity	Angle (°)	Mean angle variation and standard deviation		
		Frontal	Transversal	Sagittal
Hip Abd.	0	2.1 ± 6.8	26.4 ± 4.4	30.8 ± 4.7
	30	8.3 ± 8.7	29.6 ± 6.3	34.0 ± 6.9
	60	11.3 ± 9.2	29.9 ± 7.0	37.3 ± 7.1
Should. Abd.	0	3.9 ± 6.4	80.7 ± 5.9	64.7 ± 4.5
	30	9.8 ± 7.4	74.0 ± 7.0	58.0 ± 4.8
	60	14.8 ± 10.7	72.1 ± 8.0	43.0 ± 29.0
Hip Ext.	0	21.6 ± 8.6	26.9 ± 7.4	6.7 ± 4.3
	30	23.0 ± 5.1	30.1 ± 5.7	12.6 ± 6.5
	60	34.3 ± 2.1	36.1 ± 2.4	12.0 ± 7.2
Elbow Ext.	0	67.7 ± 9.9	79.8 ± 11.1	1.3 ± 8.3
	30	69.2 ± 6.6	71.2 ± 7.2	13.45 ± 8.4
	60	62.0 ± 6.6	74.2 ± 4.2	20.6 ± 12.2
Should. Ext.	0	43.4 ± 6.8	42.2 ± 5.0	10.3 ± 5.8
	30	30.7 ± 7.6	34.6 ± 9.6	15.2 ± 7.2
	60	25.7 ± 5.0	29.0 ± 9.0	13.8 ± 8.1
Hip Flex.	0	39.7 ± 7.2	33.1 ± 4.5	4.7 ± 3.8
	30	35.1 ± 6.1	30.4 ± 6.9	5.4 ± 4.9
	60	31.9 ± 8.1	31.4 ± 16.7	2.2 ± 5.4
Elbow Flex.	0	24.1 ± 5.8	128.2 ± 4.8	2.8 ± 9.0
	30	37.2 ± 8.3	124.4 ± 6.5	12.3 ± 9.0
	60	32.0 ± 5.1	122.7 ± 8.6	12.9 ± 9.5
Should. Flex.	0	54.8 ± 4.6	68.8 ± 5.6	14.3 ± 8.0
	30	72.1 ± 9.9	83.1 ± 7.0	6.5 ± 7.5
	60	82.9 ± 6.7	77.1 ± 4.8	21.2 ± 16.4
Ext. Rot.	0	40.4 ± 10.7	5.4 ± 13.1	41.7 ± 11.9
	30	40.2 ± 16.2	7.7 ± 14.3	34.6 ± 27.6
	60	32.6 ± 5.2	3.8 ± 11.2	31.2 ± 9.7
Int. Rot.	0	38.7 ± 17.1	19.9 ± 19.1	31.6 ± 8.0
	30	41.3 ± 10.9	27.1 ± 14.2	18.2 ± 17.1
	60	39.1 ± 13.2	16.7 ± 12.3	26.1 ± 14.8

Table 3.3: Means of the angle variation for different activities

Together with the standard deviation of the final positions the values of the final angles measured can be used to determine the thresholds for the recognition using Finite State Machines, so long as the standard deviation observed is relatively small. The standard deviations observed for the angles measured in what is considered the final position of the movement are presented in Table 3.3. As expected, even though the amplitudes of the movements do not vary significantly, the measurement error increases when the subject is not standing in front of the sensor, this is manifested in an increase in the standard deviation.

Activity	Angle (°)	Initial	Final
Hip Abduction	Transversal	equal	$\mu_3 > \mu_2$ $\mu_3 > \mu_1$
	Sagittal	$\mu_1 > \mu_3$ $\mu_2 > \mu_3$	$\mu_1 > \mu_3$ $\mu_2 > \mu_3$
Shoulder Abduction	Transversal	equal	equal
	Sagittal	$\mu_1 > \mu_3$ $\mu_2 > \mu_3$	equal
Hip Extension	Frontal	$\mu_2 > \mu_3$	equal
	Transversal	equal	$\mu_3 > \mu_1$ $\mu_3 > \mu_2$
Elbow Extension	Frontal	equal	$\mu_2 > \mu_3$
	Transversal	equal	$\mu_3 > \mu_2$
Shoulder Extension	Frontal	$\mu_2 > \mu_1$ $\mu_3 > \mu_1$	$\mu_2 > \mu_1$ $\mu_3 > \mu_1$
	Transversal	$\mu_2 > \mu_3$	$\mu_2 > \mu_1$ $\mu_2 > \mu_3$
Hip Flexion	Frontal	equal	$\mu_1 > \mu_2$ $\mu_1 > \mu_3$ $\mu_2 > \mu_3$
	Transversal	equal	equal
Elbow Flexion	Frontal	equal	equal
	Transversal	equal	equal
Shoulder Flexion	Frontal	$\mu_2 > \mu_1$	$\mu_2 > \mu_1$ $\mu_3 > \mu_1$
	Transversal	$\mu_3 > \mu_1$ $\mu_2 > \mu_1$	$\mu_3 > \mu_1$ $\mu_2 > \mu_1$
External Rotation	Frontal	$\mu_1 > \mu_2$	equal
	Sagittal	equal	$\mu_1 > \mu_2$
Internal Rotation	Frontal	$\mu_1 > \mu_2$ $\mu_3 > \mu_2$	$\mu_1 > \mu_2$ $\mu_3 > \mu_2$
	Sagittal	equal	$\mu_1 > \mu_2$ $\mu_1 > \mu_3$

Table 3.4: Repeated measures ANOVA result for relevant angles with respect to rotations of the subject

To determine how sensitive the calculated angles are to the rotation of the person in front of the sensor, it is necessary to know whether the means of the angles for the different orientations are statistically different. Since measurements are done over the same subjects for the different orientations, repeated measures ANOVA was performed, the results are shown in Table 3.4. There, μ_1 stands for the mean of the angle when the person is standing in front of the sensor, while μ_2 and μ_3 are the means for rotations of 30° and 60°. An angle is considered robust to small rotations when μ_1 and μ_2 are not statistically different. It was found that for both shoulder and hip abductions the transverse and frontal angles were robust to small rotations. The angles for shoulder extension varied in great extent between the orientations of the user. For flexion movements it was found that the frontal angle is not robust, and recognition would depend on the transverse angle. Shoulder rotations yielded the worst results, since even small rotations cause great variations in the measured angles, making it impossible to determine recognition intervals invariant to rotations.

Results of the third experiment results are summarized in Table 3.5. The table shows for each one of the subjects the number of times that the activity was correctly executed, which in all cases was 10, and the number of times that the software correctly identified them as according to the specification. Only for one subject it was found that the normal movement of the arm was outside the thresholds found by taking the mean of the angles in the database and extending intervals of 2 standard deviations around these means. Errors were usually due to sudden increases in the calculated angles, or incorrect determination of the tendency of the angles either to increase or decrease.

Subject	Score	Subject	Score
1	7/10	5	5/10
2	7/10	6	8/10
3	8/10	7	10/10
4	7/10	8	7/10

Table 3.5: Success rate of the recognition for different subjects

3.8 Conclusion

Experiments were conducted to determine whether the use of two *kinect* sensors can improve the stability of angles formed with the planes of motion for physical activities common in coaching and therapy. The repeatability of such angles was also studied across different subjects with respect to rotations of the person relative to the sensor.

The use of two *kinect* sensors offered a marginal improvement when compared to the use of a single sensor. It proved to be especially useful for shoulder flexion, in which the limb moves towards the camera. It remains interesting to determine whether different optimization approximations could yield better results.

For the case of shoulder flexion, and rotations the proposed angles were not found to be robust enough, this is due to the movement of the arm towards the camera and the occlusions arising from it. For these cases other characteristics must be used.

The evaluation of movement using Finite State Machines proved to be reliable in finding the mistakes made while performing the tested activity.

Chapter 4

HMM and DTW for Evaluation of therapeutical gestures using kinect

Results regarding the ability of the proposed approach to detect abnormal movement sequences were published in the 11th International Symposium on Visual Computing, that took place in Las Vegas, Nevada in December 14-16, 2015.[58]

This chapter presents a method for the detection of deviations from the correct form in movements from physical therapy routines based on Hidden Markov Models, which is compared to Dynamic Time Warping. The activities studied include upper an lower limbs movements, the data used comes from a *kinect* sensor. Correct repetitions of the activities of interest were recorded, as well as deviations from these correct forms. The ability of the proposed approach to detect these deviations was studied. Results show that a system based on HMM is much more likely to determine if a certain movement has deviated from the specification.

4.1 Introduction

Physical therapy is a common step in the rehabilitation process for many injuries and diseases. Movements used in therapy and conditioning routines are well defined in order to strengthen specific body segments and prevent injuries. The usual method by which one of these movements is evaluated is by having a human expert observe it and give feedback to the person executing it. This introduces subjectivity in the evaluation process and implies the need for expert personnel in rehabilitation, increasing the costs for the health system. Developing a low cost, precise evaluation system for therapeutical gestures could help ease the burden on health systems all over the world while increasing objectivity in the measurements.

Most of the research on the computer vision community focused on human activities has been directed towards the recognition of activities on video sequences. An extensive review can be found in Mohamed et al[59]. In the case of rehabilitation and therapy however the main focus lies on the quality of the movement executed by the patient, and not in the identification of the activity. This means that an automatic system must be able to determine whether the patient has executed the movement according to the specification of an expert.

Several different problems arise however when trying to translate movement specifications into software conditions, as clearly evidenced by the work of Velloso et al [42], where several experts in weightlifting miscalculated the ideal angles to be formed during execution of common exercises. Even for a clearly defined activity there remains a set of problems to be tackled, like the noise inherent to the measurements taken by the sensors used to monitor the activity, and self-occlusions when using a computer vision system. Another problem lies in the generalization of the evaluation approach to different subjects, given the variability inherent in the human population.

Two widespread techniques used to measure the adjustment of an observation to a pattern of data in a time series are Dynamic Time Warping (DTW) and Hidden Markov Models (HMM). The former consists in aligning a sequence to another of different length by means of finding the path that minimizes the sum of distances between all corresponding pairs of elements, the technique can be extended to handle vector quantities by using euclidean distance between vectors. HMM parametric statistical models, that can be trained using observations from a process evolving in time, the likelihood of any other sequence of being generated by the model can then be calculated after training.

Most of the works based on the use of DTW to recognize quality of movement do not present any experiment to validate the performance of the method when applied to sequences that deviate slightly from the correct form. In this work we use experimental data collected among 14 subjects who were instructed to perform as set of physical activities according to the specifications of an expert,data from ten different subjects who were instructed to deviate from the correct performance for the activities was also collected. This was used to test two evaluation approaches, one based on Multi-Dimensional Dynamic Time Warping (MDDTW) and another based on Hidden Markov Models.

4.2 Related Work

Most proposals in the field of tele-health based on the tracking offered by the Kinect device make use of Finite State Machines or Dynamic Time Warping (DTW) [60] to determine if the performed movement is close to the specification by the expert. An example of the first approach is Velloso et al.[42], who developed a system to help evaluate movements in a physical therapy context, by recording ten repetitions of a movement and then obtaining a model for the activity. Then a direct comparison is used between such models and the observed values to determine whether the activity has been correctly executed. The second approach has been used by Su et al. [61] who proposed recording a single instance of an activity under supervision and using DTW to compare sequences of coordinates of the hands. The degree of similarity is then determined by using fuzzy membership functions, one for the DTW result and one for the speed of the movement. Cuellar et al. [62] developed a system that aligns a repetition of an activity with a template recorded by a therapist, obtaining a score that is the average of scores for each frame of the sequence after aligning them with DTW, the result becomes the input of an evaluation function which is a Gaussian Function or a Gaussian Bell Function, used to assign a score to the movement.

The use of classification methodologies has also been studied as a means to detect mistakes, this is the case of Staab[63], who uses Support Vector Machines with different kernels to recognize common mistakes for the case of three physical activities. The main downside of this approach is the fact that it becomes unfeasible to train classes for all the different kinds of errors that can appear while performing a certain activity, as stated in [2].

Alternatives based on the use of statistical models to determine deviations from normality have also been proposed, Paiement et al. [39] use such a model to detect abnormalities in a sequence of movements of a person moving on stairs, and determine empirically the threshold to consider an instance of observations as normal.

Outlier recognition based on HMM is applied to fault detection in antennae by Smyth et al. [64]. Yan et al. [65] propose a methodology based on the Wavelet transform and HMM to detect outliers and test it on data coming from depth measurements. Zhu et al [66] use a modified HMM model to detect faults in industrial processes. The advantages of a system based on HMM to detect abnormal movement sequences are that it can be trained only with repetitions of movements considered to be normal, that it can model sequential data and that depending on the training sequences it can be adjusted to be more or less tolerant to deviations from normality.

4.3 Methods

This section shows the calculations performed in order to characterize the movements taken into account in the database collected, it also presents the algorithms used for evaluation of performance as well as the mechanism used to determine whether a sequence is considered to be correct or not.

4.3.1 MDDTW

Let X and Y be two sequences, $X = [x_1, x_2 ... x_N]$ $Y = [y_1, y_2, ..., y_M]$. The DTW technique finds a matching of the two sequences that minimizes the distance between the two, for this two warping functions w_x and w_y are defined such that $w_x(1) = 1$, $w_x(K) = N$, $w_y(1) = 1$, $w_y(K) = M$, so that the elements of the alignment would be pairs of corresponding warped entries (w_x,w_y).

Let $DTW(X,Y)$ be the accumulated distance between two sequences. The purpose is to find the alignment between the sequences that minimizes the distance between the two:

$$DTW(X,Y) = \min_{w_x,w_y} \sum_{k=1}^{K} d(x_{w_x}(k), y_{w_y}(k)) \tag{4.1}$$

The dynamic programming algorithm proposed by [60] computes the elements of the distance matrix between two sequences iteratively as

$$DTW(i,j) = min \begin{cases} DTW(i-1,j) + d(i,j) \\ DTW(i,j-1) + d(i,j) \\ DTW(i-1,j-1) + d(i,j) \end{cases} \tag{4.2}$$

This means that in each step the algorithm attempts to minimize the accumulated distance between the sequences. In order to avoid computing all of the possible entries of the accumulated distance matrix, a restriction is usually enforced between the pairs of values of the sequences being considered

$$j - r \leq i \leq j + r \tag{4.3}$$

Multi-Dimensional Dynamic Time Warping is an extension of the DTW algorithm in which the warping path and the distance between two sequences are computed by taking the distance between two vectors on each step of the algorithm. This allows the use of a pattern defined as a trajectory in an hyperspace instead of a single valued function.

Let two example sequences be

$$S_1 = \begin{bmatrix} -0.60 & -0.65 & -0.71 & -0.58 & -0.17 & 0.77 & 1.94 \\ -0.46 & -0.62 & -0.68 & -0.63 & -0.32 & 0.74 & 1.97 \end{bmatrix} \tag{4.4}$$

$$S_2 = \begin{bmatrix} -0.87 & -0.84 & -0.85 & -0.82 & -0.23 & 1.95 & 1.36 & 0.60 & 0.0 & -0.29 \\ -0.88 & -0.91 & -0.84 & -0.82 & -0.24 & 1.92 & 1.41 & 0.51 & 0.03 & -0.18 \end{bmatrix} \tag{4.5}$$

0	Inf	Inf	Inf	Inf	Inf	Inf	Inf
Inf	0.0294	0.0494	0.0645	0.0870	Inf	Inf	Inf
Inf	0.0594	0.0498	0.0649	0.0869	0.1395	Inf	Inf
Inf	0.0861	0.0673	0.0623	0.0824	0.1327	0.2659	Inf
Inf	0.1109	0.0827	0.0727	0.0803	0.1285	0.2595	0.4904
Inf	Inf	0.1160	0.1110	0.1036	0.0862	0.1685	0.3507
Inf	Inf	Inf	0.3298	0.3149	0.2676	0.1843	0.1715
Inf	Inf	Inf	Inf	0.4805	0.4034	0.2368	0.2189
Inf	Inf	Inf	Inf	Inf	0.4700	0.2537	0.3355
Inf	Inf	Inf	Inf	Inf	Inf	0.3153	0.4150
Inf	Inf	Inf	Inf	Inf	Inf	Inf	0.4975

Figure 4.1: Example application of the implemented MDDTW algorithm applied to two toy sequences

Figure 4.1 shows the result of applying the technique to two example sequences. The two example sequences were taken from [67]

Given a set of trajectories considered as correct movements $X_1, X_2, ..., X_N$, the idea is to extract from these set a trajectory that best represents the movement of interest. This however poses a problem, due to the variability that can be found in any human population. With this in mind, the trajectories were normalized. The first problem with the normalization is that people might be located differently with respect to the sensor. To account for this the origin of coordinates is shifted, this is done by subtracting the coordinates of the shoulder center from the coordinates of the limb of interest when the activity is performed with the upper limb, and the coordinates of the hip center when the activity is performed with the lower limbs.

Let a sequence of interest be X_i as defined by Equation 4.6.

$$\mathbf{X}_i = \left\{ \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} ... \begin{pmatrix} x_m \\ y_m \\ z_m \end{pmatrix} \right\} \tag{4.6}$$

Then the coordinates of the shoulder center evolve in time to create an analogous sequence, as is shown in Equation 4.7.

$$\mathbf{S}_c = \left\{ \begin{pmatrix} Sc_1x \\ Sc_1y \\ Sc_1z \end{pmatrix} ... \begin{pmatrix} Sc_mx \\ Sc_my \\ Sc_mz \end{pmatrix} \right\} \tag{4.7}$$

The sequences are modified according to Equation 4.8, where the subtraction is performed for each element of the sequences.

$$\mathbf{X}_i = \mathbf{X}_i - \mathbf{S}_c. \tag{4.8}$$

Other problem that arises is the fact that the limbs from different people have different lengths. In order to deal with this the coordinates are divided by the distance between the shoulders, when analyzing upper limb movements, and by the distance between the hips for sequences of the lower limbs. Equations

4.9 and 4.10 show the sequences of coordinates of both shoulders, from which a new sequence **D** is obtained by subtracting each of the corresponding elements of the sequence.

$$\mathbf{R}_s = \left\{ \begin{pmatrix} Rs_1x \\ Rs_1y \\ Rs_1z \end{pmatrix} ... \begin{pmatrix} Rs_mx \\ Rs_my \\ Rs_mz \end{pmatrix} \right\}. \tag{4.9}$$

$$\mathbf{L}_s = \left\{ \begin{pmatrix} Ls_1x \\ Ls_1y \\ Ls_1z \end{pmatrix} ... \begin{pmatrix} Ls_mx \\ Ls_my \\ Ls_mz \end{pmatrix} \right\}. \tag{4.10}$$

$$\mathbf{D} = \mathbf{R}_s - \mathbf{L}_s. \tag{4.11}$$

The final sequence X_i is obtained by dividing each element of X_i by each element of the sequence D, as shown in Equation 4.12.

$$\mathbf{X}_i = \mathbf{X}_i/\mathbf{D}. \tag{4.12}$$

Once this is done MDDTW is calculated for all the possible pairs of sequences.

Then a trajectory is selected for which the sum of these distances is minimal, this is considered the template for the activity under consideration.

Once this is done, the distance of all the sequences to the template is calculated, and an interval constructed around the mean of this distance. For evaluation the distance of the test sequences to the template is calculated, and those whose distance lies within the interval are considered to be close enough to the correct repetitions to be also considered correct.

4.3.2 HMM

Hidden Markov Models are stochastic models that consider an observed signal as the result of the transition of a system between several states, in each of these states there is a certain probability that one symbol might be observed. They have been used in speech processing [68], and outlier detection[65] ,[66].

An HMM is characterized by:

- N,the number of states of the model, these states are hidden, meaning that it is impossible to know precisely in which of the states the model is.
- M the number of observations, these are the symbols that are supposed to be generated probabilistically in each of the states.
- A, the state transition probability distribution.

  If the system has N different possible states $S_1, S_2, ...S_N$ and the state at time t is q_t, the elements of the transition matrix A describe the probability of passing from one state to another.

$$a_{ij} = P[q_t = S_j|q_{t-1} = S_i] \quad i \leq N \quad j \leq N \tag{4.13}$$

- B, the observation symbol probability distribution in each state. The observation matrix B describes the probability of observing a certain symbol while the system is in each of its states. If the system has N different possible states $S_1, S_2, ..., S_N$ and in each state M different symbols may be observed, the elements of the observation matrix B are

$$b_{ij} = P[v_j \quad at \quad t|q_t = S_i] \quad i \leq N \quad j \leq M \tag{4.14}$$

  where $v_1, v_2, ..., v_M$ is the set of all the possible symbols that can be emitted by the system.

- the initial state distribution π, is a vector that gives the probabilities of the system being in one of the states for the initial observation.

$$\pi = P[q_1 = S_i] \quad i \leq N \tag{4.15}$$

An HMM model is totally specified when the transition matrix A, the observation matrix B and the vector of initial states π are known.

Two broad categories of HMM models exist, the left-right HMM in which the transitions between state have the property that it is impossible to pass from one state to the previous one as shown in Figure 4.2, and the ergodic HMM in which a transition can occur between any pair of states, as shown in Figure 4.3.

$$a_{ij} = 0 \quad i \leq j \tag{4.16}$$

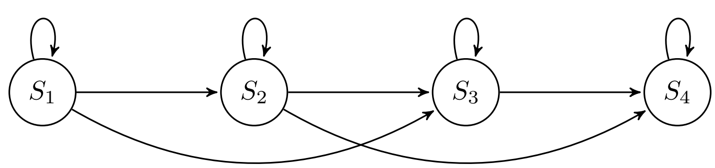


Figure 4.2: Example of a left-right HMM

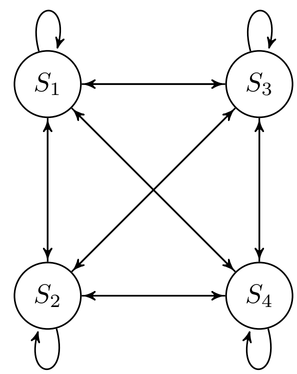


Figure 4.3: Example of an ergodic HMM

There are three fundamental problems associated to the use of HMMs.

- How to compute the probability that a given sequence was generated by the model. This problem is solved by using the forward algorithm.
- How to know what was the sequence of states that the model went through for a given observation sequence. This problem is solved by using the Viterbi algorithm.
- How to adjust the parameters of the model so that the probability of observing a given set of observation sequences. This problem has no exact solution, but the Baum-Welch algorithm is used to arrive to local solutions.

In order to specify a certain movement by using HMMs we first determine three characteristics of the movement which are the angles formed with each of the three movement planes shown in Figure 1.1.To accomplish this, vectors normal to the planes must be estimated first by using the data coming from the sensor. These calculations are performed according to equations 3.1, 3.2, 3.3. Once the vectors are estimated, the angles formed with each one of the planes of motion can be estimated using the equations 3.4, 3.5, 3.6.

Once the sequences of angles are obtained, they need to be quantized in order to use the discrete version of HMM, this quantization is done according to Figure 4.6. This figure shows the coordinate system used by the sensor and the planes of motion. Frontal angle values are considered positive when the limb stands on the side of the frontal plane determined by the direction of the positive z axis, and negative otherwise. Transverse angle values are considered positive if the limb stands in the side of the transverse plane marked by the direction of the positive y axis, and negative otherwise. Sagittal angle values are considered positive if the limb stands in the side of the sagittal plane marked by the direction of the positive x axis, and negative in the opposite sense. Since the angle between a vector and a plane can be any value between 0 and 90 degrees, the quantizer considers a range of -90 to +90 degrees. For any angular value there exists a symbol assigned to it, so that the sequence of angles becomes a discrete sequence, all of the possible values of the calculated angles are considered, so that any observed sequence can be assigned a certain probability. In order to account for the noise in the measurement a variation of ten degrees was considered as sufficient to change the emitted symbol.

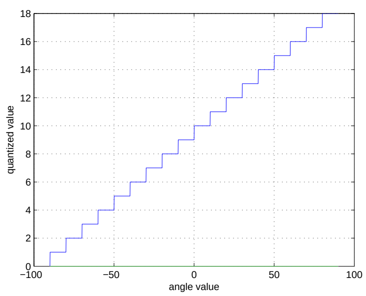


Figure 4.4: quantizer

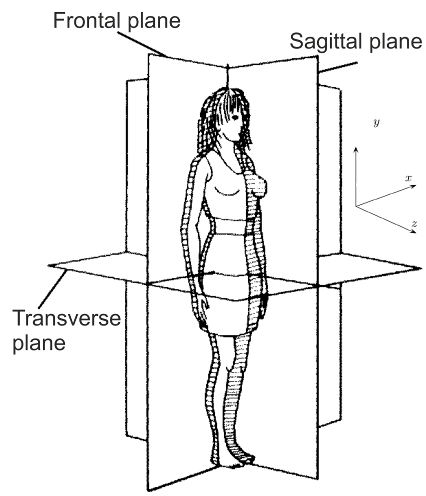


Figure 4.5: coordinate system

Figure 4.6: Coordinate system and values for the quantizer

Once the sequences corresponding to the correct repetitions of the activities have been quantized, they can be used to train the model, this means estimating the structures of matrices A and B that maximize the probability of observing such sequences. Baum -Welch algorithm is used for this purpose.

As an example we show these matrices for the case of Hip Flexion:

$$A = \begin{bmatrix} 0.9050 & 0.0950 & 0.0000 \\ 0.0031 & 0.9919 & 0.0050 \\ 0.0000 & 0.0706 & 0.9294 \end{bmatrix}. \tag{4.17}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9714 & 0.0286 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0012 & 0.9988 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0224 & 0.9776 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.18}$$

It is clear from the structure of the matrix A, that the HMM model generated is a left-right model. This is due to the sequential nature of the movement data used, in which the the signal values have a tendency to increase or decrease for the whole duration of the training sequence.

Matrix B shows an important characteristic of the trained models, the fact that they are being trained only with a single phase of the activity implies that

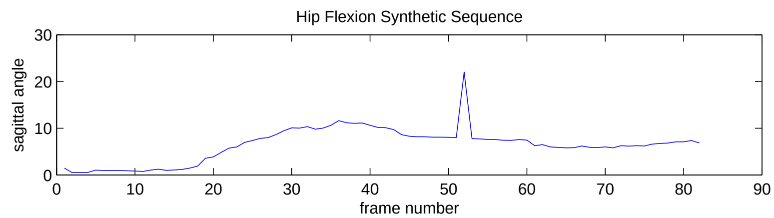


Figure 4.7: Modified sequence containing an outlier

the symbols that the model can emit are very limited, and the probability of emitting a different symbol becomes an extremely small number.

The model can be modified manually so that the presence of a single outlier does not affect detecting the movement as occurring according to the specification. This is achieved by modifying the structure of matrix B and recalculating the probabilities for the sequences. Figure 4.7 shows a synthetic error sequence containing a single outlier for the sagittal angle during a Hip Flexion movement.

Calculation of the probability of observing this sequence yields a numerical error result on Matlab, owing to the fact that it is an extremely small number, but upon modyfing matrix B to make it probable to emit 12 as a symbol as shown in equation 4.19, the probability of observing the synthetic sequence becomes finite, and is equal to -27.8073.

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9714 & 0.0286 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0012 & 0.9988 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0224 & 0.8 & 0.1776 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.19}$$

By taking all the sequences considered to be correct and calculating their probabilities according to the trained HMM, an interval can be constructed of two standard deviations around the mean, for this particular case the interval found is $[-28.9661\ 8.7130]$.The synthetic sequence then is classified as a sequence that might be generated by the model. This means that the manually modified model is robust to the presence of a few outliers, in a scenario where a threshold detection would have failed. Introducing two outliers, however, yields a numerical error, meaning that the number is extremely small.

The identification of mistakes can be done by analyzing the symbols emitted in the erroneous repetitions and comparing them with the symbols emitted in the training repetitions.Symbols that are greater or smaller than the ones present in the training repetitions indicate angles that are either too big or too small. This allows for feedback expressed in terms of the angular quantities, at the end of the recognition of each phase of the activities.

With the trained model all of the trained sequences are assigned a probability by using the forward algorithm, and an interval is built around the mean of such probabilities, by taking two standard deviations.

Evaluation of new sequences proceeds as follows: error sequences are quantized and then assigned a probability by using the forward algorithm. If this probability lies on the interval built with the correct repetitions, the sequence is considered to be correct.

4.4 Experimental Setup

Experiments were conducted to determine the effectiveness of the proposed algorithms to reject sequences that constitute deviations from what is considered a correct execution of the movements of interest.

Correct repetitions of the activities were recorded by asking 14 subjects to perform each one of the ten physical activities according to the specification three times.The activities considered are exercises commonly found in physical therapy and rehabilitation routines. The list of activities is shown in table 4.1.

Table 4.1: Activities of interest

1. Shoulder Extension	6. Hip Abduction
2. Shoulder Flexion	7. Shoulder Internal Rotation
3. Shoulder Abduction	8. Shoulder External Rotation
4. Hip Extension	9. Elbow Flexion
5. Hip Flexion	10. Elbow Extension

For a more complete description of the activities the reader can see appendix 2.2

This means that 42 repetitions were recorded for each one of the activities of interest. Some of these were excluded from the analysis based on the noise present in the signals, in order to obtain a pattern as well defined as possible.

A database of mistakes during the execution of activities was collected by asking ten persons to perform the incorrect form of the activity ten times. This gives us a total of 100 incorrect repetitions for each error considered. The number of ways in which an activity can be incorrectly executed is potentially infinite, so in this study we chose to select errors that corresponded to common deviations in the normal performance of the activities. Since the evaluation is based on differences between the observed sequences, it is logical to think that much greater deviations than the ones considered here would be detected even more easily by an algorithm able to detect the ones presented here.

The mistakes are explained as follows:

For the abduction exercises, *Error 1* consists in deviating the limb of interest towards the front of the body while performing the activity, increasing the angle formed with the frontal plane, this is shown in figure 4.8.*Error 2* consists in deviating the limb towards the back of the body, decreasing the angle formed with the frontal plane, this is shown in figure 4.9.

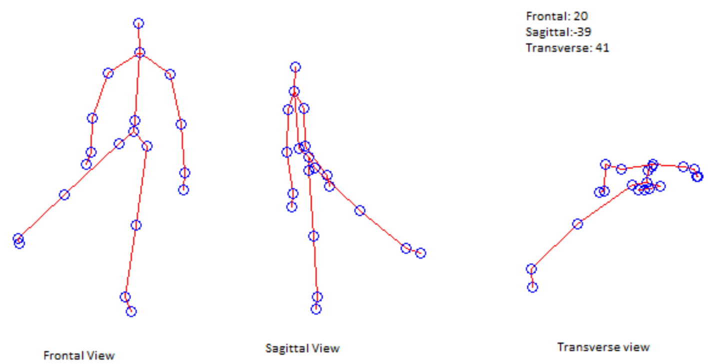


Figure 4.8: Erroneous performance consisting in moving the limb towards the front of the body

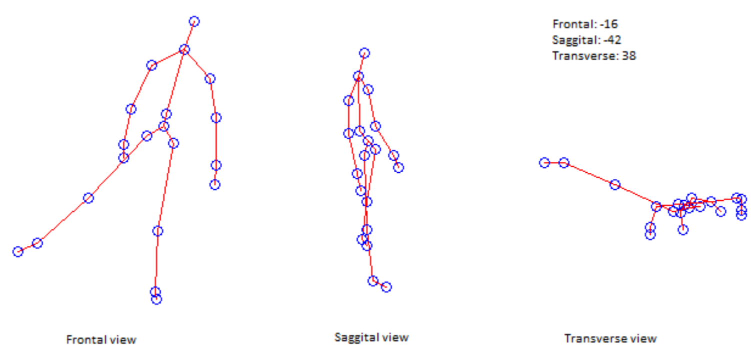


Figure 4.9: Erroneous performance consisting in moving the limb towards the back of the body

For extension and flexion exercises *Error 1* consists in deviating the limb away from the body , increasing the value of the angle formed with the sagittal plane, this is shown in Figure 4.10. *Error 2* consists in moving the limb towards the body, decreasing the angle formed with the sagittal plane, this is shown in Figure 4.11.

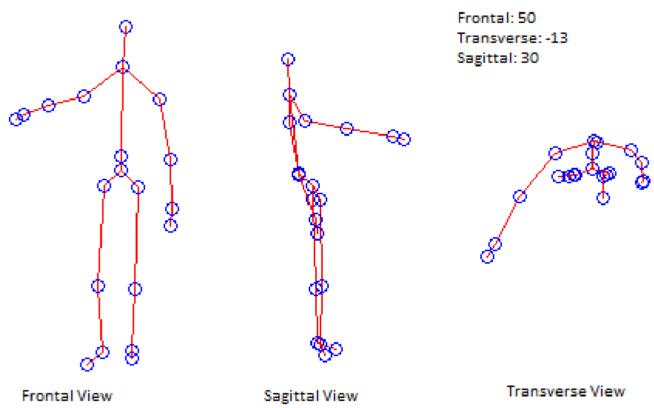


Figure 4.10: Erroneous performance consisting in moving the limb away from the body plane

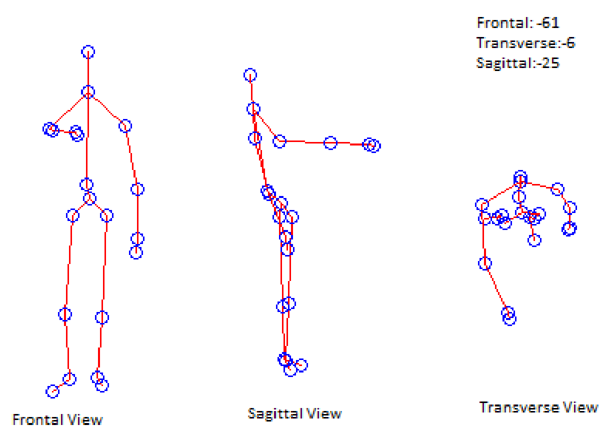


Figure 4.11: Erroneous performance consisting in moving the limb closer to the body

For shoulder rotation exercises *error1transversal* consists in deviating the limb towards the floor, decreasing the value of the angle formed with the transverse plane,this is shown in Figure 4.12.*error2transversal* consists in deviating the limb upwards, increasing the value of the angle formed with the transverse plane, this is shown in Figure 4.13.

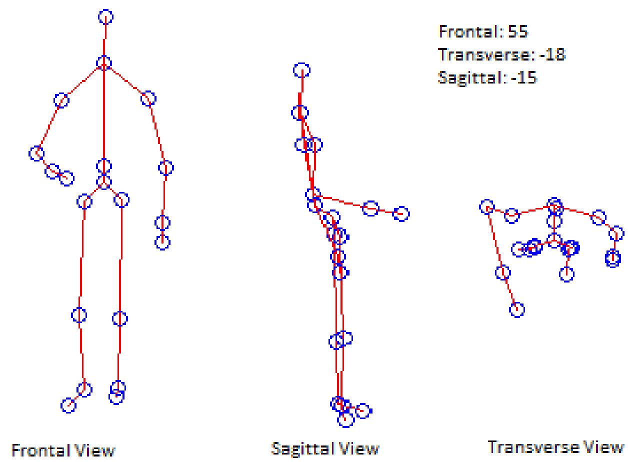


Figure 4.12: Erroneous performance consisting in moving the limb downwards

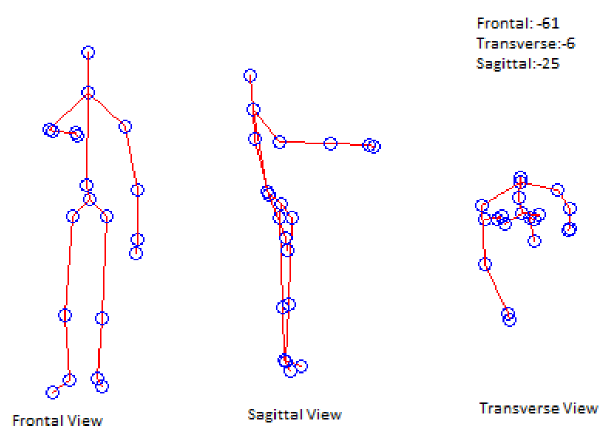


Figure 4.13: Erroneous performance consisting in moving the limb upwards

In our analysis of performance we have chosen to divide the activities performed in two basic movement phases, one that consists in the limb of interest moving away from the body and another that moves it back to the initial position. This allows an evaluation that is closer to the ideal of real-time performance for such a system, since it would be frustrating for the user to receive a negative evaluation of his or her performance once the complete movement has been performed.

The first experiment was done in order to test the ability of a DTW based system to reject movements that deviate from the specification. It consisted in determining using the correct repetitions for the activities both a template and an interval of distances to consider a repetition as correct. The data used in these experiments consists in coordinates of the limb of interest normalized to take into account relative displacement to the sensor and person size. The distance to the template is then calculated for all the incorrect repetitions on both phases of the movement and if it lies within the interval, the repetition can be considered correct.

The second experiment varied the data fed to the DTW recognition scheme, by using the estimated angles with respect to the planes of motion. A sequence of vectors containing the angles formed with these planes is used and a template calculated that minimizes the distance to the other sequences. Similar angular sequences are also calculated for the incorrect repetitions.

The third experiment consisted in training HMM models using the correct repetitions of the activities, the data used to train the model consists in the same coordinate sequences used for the first experiment. Since a model can be trained using discrete sequences, an HMM model is trained for each one of the characteristics of interest, for each one of the movement phases.This means that for each phase of an activity, three HMM models are generated. The models were then tested with the incorrect repetitions, by using the forward algorithm, to determine the model's ability to reject them.

The fourth experiment consisted in training HMM models varying the sequences used for training, this time the angular sequences from experiment two are used. Once again an interval of probability values is obtained and the incorrect repetitions are tested by determining whether the probability assigned to them by the model lies within the interval.

4.5 Results

This section shows the recognition percentages for the errors taken into account for each one of the ten activities and for the two methods studied.

Table 4.2 shows the results of the first experiment, the percentages correspond to erroneous sequences within the error database that were detected as such. Percentages of error recognition are shown for both the phases in which the movement has been divided. Since the technique takes into account vectors used to describe the trajectories, only one metric distance is obtained for each error sequence and based on this distance the percentage of correctly refused erroneous sequences is calculated.

Table 4.2: Percentage of error detection using MDDTW with coordinate sequences (Experiment 1)

	Phase 1		Phase2	
Activity	Error 1	Error 2	Error 1	Error 2
Shoulder Abduction	47	38,88	100	61,11
Hip Abduction	18,75	48,10	15	29,11
Shoulder Extension	22,73	6	0	1
Hip Extension	29,21	11	16,85	11
Elbow Extension	47,37	50	46,32	50
Shoulder Flexion	51	5,15	46	4,12
Hip Flexion	20	12,12	24	13,13
Elbow Flexion	47	71	44	60
Internal Rotation	71,59	7,29	9,09	16,67
External Rotation	52	40	49	14

Table 4.3 shows the results of the second experiment, percentages show the success of the technique in rejecting erroneous sequences executed by the participants.

Table 4.3: Percentage of error detection using MDDTW with angle sequences (Experiment 2)

	Phase 1		Phase 2	
Activity	Error 1	Error 2	Error 1	Error 2
Shoulder Abduction	34,29	7,78	62,9	18,89
Hip Abduction	22,5	11,4	13,75	2,53
Shoulder Extension	1,13	1	0	1
Hip Extension	3,37	0	4,5	2
Elbow Extension	22,11	30,21	8,42	11,46
Shoulder Flexion	12	1,03	3	1,03
Hip Flexion	0	1,01	0	0
Elbow Flexion	35	51	36	10
Internal Rotation	0	0	1,14	0
External Rotation	1	1	1	1

Table 4.4 show the results of the third experiment. A different HMM model is trained for each one of the three angles of motion (sagittal, frontal and transversal). Three results are then shown for each one of the phases of movement and for each one of the types of errors considered. The highest result of the three can be considered the actual recognition of errors accuracy for HMM based models.

Table 4.4: Error Recognition Results for HMM using angle sequences

	Phase 1						Phase 2					
Activity	Error1 %			Error2 %			Error1 %			Error2 %		
Shoulder Abduction	90.0	**92.9**	21.4	34.4	**96.7**	26.7	77.1	75.7	**82.9**	52.2	78.9	**92.2**
Hip Abduction	32.5	**62.5**	41.3	**36.7**	30.4	8.9	22.5	**60.0**	41.2	10.1	**21.5**	7.6
Shoulder Extension	14.8	38.63	**64.8**	**45.0**	3.0	7.0	25.0	36.4	**76.1**	**51.0**	1.0	9.0
Hip Extension	**78.7**	40.4	71.9	**72.0**	30.0	25.0	21.3	28.9	**95.5**	37.0	24.0	**57.0**
Elbow Extension	17.9	35.8	**54.7**	26.0	**57.3**	21.9	63.1	28.4	**87.4**	79.2	32.3	**98.9**
Shoulder Flexion	17.0	54.0	**68.0**	10.3	63.9	**77.3**	33.0	**72.0**	25.0	30.9	**78.3**	47.4
Hip Flexion	12.0	7.0	**99.0**	27.3	16.2	**29.3**	0.0	7.0	**97.0**	22.2	11.1	**28.3**
Elbow Flexion	89.0	89.0	**99.0**	**90.0**	81.0	25.0	19.0	46.0	**78.0**	18.0	**78.0**	55.0
Internal Rotation	68.2	**77.3**	32.9	43.3	**75.3**	39.2	48,9	**73.9**	27.3	39.2	**77.3**	24.7
External Rotation	16.0	**88.0**	25.0	17.0	**68.0**	38.0	30.0	**91.0**	43.0	38.0	38.0	**59.0**

Table 4.5 show the results of the fourth experiment. Three models are trained, one for each one of the coordinates X,Y,Z of the sequence describing the activity. The highest percentage can be considered the recognition of deviation from normality for HMM models under these circumstances.

Table 4.5: Error Recognition Results for HMM using coordinate sequences

	Phase 1						Phase 2					
Activity	Error1 %			Error 2 %			Error1%			Error2%		
Shoulder Abduction	**95.7**	17.1	84.3	96.7	12.2	**98.9**	**94.3**	61.4	75.7	**98.9**	30.0	97.8
Hip Abduction	70	**83.8**	71.3	72.2	**81.0**	54.4	50.0	40.0	**70.0**	36.7	27.8	**48.1**
Shoulder Extension	**68.2**	35.2	4.5	**56.0**	37.0	16.0	**67.0**	50.0	1.1	**56.0**	44.0	4.0
Hip Extension	73.03	**84.3**	80.9	67.0	**83.0**	80.0	69.7	74.2	**74.2**	63.0	**77.0**	70.0
Elbow Extension	**36.8**	7.3	14.7	**62.5**	13.5	14.6	**14.7**	7.4	10.5	15.6	12.5	**15.6**
Shoulder Flexion	**89.0**	23.0	75.0	**81.4**	18.6	69.1	74.0	5.0	**88.0**	73.2	4.1	**82.5**
Hip Flexion	**81.0**	0.0	1.0	**52.5**	4.1	0.0	**52.0**	0.0	2.0	**34.3**	3.0	5.0
Elbow Flexion	**94.0**	66.0	49.0	**94.0**	79.0	71.0	**84.0**	6.0	73.0	**91.0**	7.0	86.0
Internal Rotation	**50.0**	46.0	1.1	37.5	**51.0**	2.1	28.41	**36.4**	1.1	26.0	**35.4**	2.1
External Rotation	0.0	**31.0**	2.0	1.0	**41.0**	2.0	0.0	**75.0**	3.0	2.0	**81.0**	9.0

It is evident from tables 4.2 and 4.3 that the performance of MDDTW for recognizing deviations is reduced when using angular characteristics instead of normalized coordinates.

From Table 4.6 it can be seen that the performance of HMM is reduced when using coordinate sequences, in comparison to the use of estimated angles formed with the planes of motion.

Table 4.6: Average recognition percentages for each technique for each combination of factors

<table>
<tr><td rowspan="8">Phase 1</td><td rowspan="4">Coordinates</td><td rowspan="2">Error 1</td><td>HMM</td><td>71,3</td></tr>
<tr><td>DTW</td><td>40,6</td></tr>
<tr><td rowspan="2">Error 2</td><td>HMM</td><td>70,1</td></tr>
<tr><td>DTW</td><td>28,9</td></tr>
<tr><td rowspan="4">Angles</td><td rowspan="2">Error 1</td><td>HMM</td><td>78,1</td></tr>
<tr><td>DTW</td><td>13,1</td></tr>
<tr><td rowspan="2">Error 2</td><td>HMM</td><td>64,7</td></tr>
<tr><td>DTW</td><td>10,4</td></tr>
<tr><td rowspan="8">Phase 2</td><td rowspan="4">Coordinates</td><td rowspan="2">Error 1</td><td>HMM</td><td>65,5</td></tr>
<tr><td>DTW</td><td>35,0</td></tr>
<tr><td rowspan="2">Error 2</td><td>HMM</td><td>62,0</td></tr>
<tr><td>DTW</td><td>26,0</td></tr>
<tr><td rowspan="4">Angles</td><td rowspan="2">Error 1</td><td>HMM</td><td>81,3</td></tr>
<tr><td>DTW</td><td>13,0</td></tr>
<tr><td rowspan="2">Error 2</td><td>HMM</td><td>64,7</td></tr>
<tr><td>DTW</td><td>47,9</td></tr>
</table>

It was found that the average error recognition rate for HMM is consistently higher than for MDDTW, in both phases of the activities and for any characteristic, as can be seen in table 4.6. This table shows the average success rate for each technique under each condition.

Generally speaking the best combination of characteristics and technique was using HMM models together with estimated angles formed with the planes of motion.

There were certain errors very hard to recognize with any combination of technique and characteristics, this is the case for an error 2 in the second phase of hip abduction (leg returning to its initial position), for which the maximum recognition percentage was merely 48,1%.

4.6 Conclusion

Experiments were conducted to test the ability of two algorithms to detect deviations from normality during the execution of activities common in physical therapy routines. The movements tested comprise activities for both the upper and lower limbs. For each one of the movements sequences corresponding to correct executions and deviations from these were recorded. The deviations considered afected the position of the limb only with respect to one of the planes of motion. For abduction movements the deviation was introduced with respect to the angle formed with the frontal plane, for flexion and extension movements the deviation was introduced with respect to the angle formed with the sagittal plane, and for rotation movement the deviation was introduced with respect to the transverse plane.

While MDDTW is a useful tool to determine the similarity between time series coming from observations of human beings executing movements, has proven unable to reject sequences of movement which are similar to the standard but deviate in one of the directions relative to the planes of motion.

HMM outperforms MDDTW for the task of detecting deviations in movement sequences defined for physical therapy and rehabilitation. This is due to the ability of these models to penalize heavily those sequences of symbols that contain symbols not present in the sequences that were used to train the model, and those sequences containing symbols that imply a very unlikely transition between states.

For shoulder abduction, rotation and flexion results of the combination of HMM and angles with the planes of motion were satisfactory. For shoulder and elbow extension and hip abduction the proposed technique does not perform well, probably due to the noise added with loss of precision of the sensor with distance.

It remains interesting as future work to determine if the use of quaternion data coming from the sensor might yield even better results for the task of detecting deviations from normality. Another future work might be the use of a discriminative model such as Conditional Random Fields to be used as outlier detectors.

Chapter 5

Extension to continuous HMM and comparison between different populations

Results of the comparison in performance between continuous and discrete HMMs were published in the 12th International Symposium on Visual Computing, that took place in Las Vegas, Nevada in December 14-16, 2016.[69]

This chapter presents an extension of the methodology presented in the previous one for the detection of deviations from the correct form in movements from physical therapy routines based on Hidden Markov Models. Continuous HMM are proposed as models to evaluate performance of subjects performing the activities of interest. The same set of activities from the previous chapter is used. A comparison is presented between continuous and discrete models. Results show that the use of continuous HMM offers improvement in the ability to recognize deviations in movements. In order to further test the ability of the proposed method to detect deviations, it was tested on a dataset made public by Velloso et al. and referenced in [2]. Finally a comparison of the measurements taken with the *kinect* sensor is presented for the case of two samples taken from different populations. One of the samples was a group of 14 young adults between the ages of 20 to 30, and the other was a group of 8 elderly people with ages between 65 and 75. Results show that for the majority of movements of interest the measured characteristics are statistically equal in both populations.

5.1 Introduction

The recognition of the quality of performance in human movements has many applications, especially in the fields of physical therapy and sports. This problem is essentially different from the task of human action recognition in the sense that the idea is not to determine, out of a certain set of actions, which one is being executed, but to measure how close to a certain specification is the performance of a certain subject. When considered as a classification problem, it turns out that the number of possible classes becomes infinite, due to the infinite possible ways that a person can deviate a movement from the specification[2]. Most works make use of distance measures to try to calculate a score that indicates the success of a person while performing a physical activity[61],[62]. These measured distance has to be calculated for each person attempting to use the system, so that it has to be tailored to meet individual needs. Some other works use finite state machines to model the desired movement and then determine using thresholds if the execution is satisfactory[42]. These thresholds will be highly dependent on the noise present in the data and also in the population that was used to gather the data, so that some individuals may not comply with the specifications. This can be solved by adjusting the thresholds depending on each one of the users. The task of qualitative action recognition can be accomplished by treating it like an anomaly detection problem. This means that a model could be trained to capture the nature of the data that characterizes the performance, and then the same model would have to be used to reject the performances that do not comply with what is expected. A very powerful approach to characterize sequential data coming from a noisy source are Hidden Markov Models. Their power lies in their ability to adapt the model parameters to maximize the likelihood of the observed sequences. This in fact means that the likelihood of sequences different from the training ones is reduced, and allows the model to be used to detect abnormal behaviors in data.

5.2 Related Work

The use of continuous HMM to detect outlier sequences of data has been explored by Wang et. al[70] in the context of wireless sensor networks. The authors chose the model for its ability to detect what they call high semantic outliers, which are long term deviations from normal behavior patterns. Allahdadi et. al.[71] also use continuous HMM models to detect abnormal behaviors in 802.11 wireless networks, and found that the models are successful in identifying several different kinds of deviations. Yang et al.[72] used continuous HMM to model the behavior of people in video data, and then applied the model to the task of recognizing unusual events by modeling the distribution of people and its variation with time. This means that abnormal movements of a crowd are assigned a small probability by the model. Cai et al.[73] use HMM for the task of detecting abnormal movement patterns of objects in a video sequence; the models are applied to filtered trajectories of vehicles, and a threshold is calculated for the probabilites given by the model of new trajectories, whenever the probability is below the threshold, an abnormal movement is considered to have taken place. Yuan et al.[74] use a modified version of HMM models, called Hidden Semi-Markov Model, to detect the changes in the land use in Beijing, this task is done by obtaining trajectories for the pixels in the image and then calculate the likelihoods produced by the models previously trained, an important drop in the likelihood signals a change in the characteristics of pixels on the image. [75] used kinematic data, position, heading, speed and a timestamp, for vessels entering and leaving Cape Town's harbour, and trained HMMs to determine using the sequences of data what the behavior of new vessels is, however the dataset used by the authors is limited.

5.3 Methods

This section presents the mathematical models and tools that were used for the task of deviation detection in human movements.

HMM models can be used for observations that are continuous in nature (e.g. kinect coordinates,accelerometer data). This means that there is no longer a need to quantize the observations, and the degradation associated with the quantization can be avoided. Such models consider the observations to come from a multivariate normal distribution or a sum of such distributions. This means that any value in a continuous interval has a certain probability of being emitted by the model. Another advantage of this type of models is its ability to model multivariate sequences, this in contrast to the approach presented in chapter 3, which needed three models, one for each characteristic of interest, to be trained. The general form of the probability density function used for each one of the states of the model is

$$b_j(\boldsymbol{O}) = \sum_{m=1}^{M} c_{jm}\mathcal{N}(\boldsymbol{O}, \boldsymbol{\mu_{jm}}, \boldsymbol{\Sigma_{jm}}) \quad 1 \leq j \leq N \tag{5.1}$$

where the following restrictions apply:

$$\sum_{m=1}^{M} c_{jm} = 1 \quad 1 \leq j \leq N \tag{5.2}$$

$$c_{jm} \geq 0 \quad 1 \leq j \leq N, 1 \leq m \leq M \tag{5.3}$$

$b_j(\boldsymbol{O})$ is the probability of observing the vector $\boldsymbol{O}$ in the state j. c_{jm} is the weight associated with the $m-th$ element of the mixture of gaussians in the state j.

The library *pmtk3* [76] is used to train the models of interest. This library provides an implementation of the Expectation Maximization Algorithm. This essentially implements the equations presented in [68] for the reestimation of the means and covariance matrices of each one of the gaussian pdfs as well as the weights.

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T}\sum_{k=1}^{M} \gamma_t(j,k)} \tag{5.4}$$

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot \boldsymbol{O_t}}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{5.5}$$

$$\bar{\Sigma}_{jk} = \sum_{t=1}^{T} \gamma_t(j,k) \cdot (\boldsymbol{O_t} - \boldsymbol{\mu_{jk}})(\boldsymbol{O_t} - \boldsymbol{\mu_{jk}})' \tag{5.6}$$

$\gamma_t(j,k)$ is the probability of being in state j at time t, with the $k-th$ component of the mixture of gaussians accounting for the observed vector O_t.

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j))}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \right] \left[\frac{c_{jk}\mathcal{N}(\boldsymbol{O_t}, \boldsymbol{\mu_{jk}}, \boldsymbol{\Sigma_{jk}})}{\sum_{m=1}^{M} c_{jm}\mathcal{N}(\boldsymbol{O_t}, \boldsymbol{\mu_{jm}}, \boldsymbol{\Sigma_{jm}})} \right] \tag{5.7}$$

In this work we use a single gaussian probability density function for each one of the states of the models. Experimentally it was found that using more than three states for each model yielded no significant increase in their ability to label sequences as either according to the specification or deviated from it. For the case of other sensor data it was found that increasing the complexity of the model in terms of the number of its states improved its ability to classify the sequences correctly, this however might lead to an arbitrarily complex model. Experiments showed that 12 states were enough to achieve a high discrimination ability.

The probability of a sequence being generated by the model is obtained by using the forward algorithm, after estimating a matrix of emissions using the logarithmic probability of each observation of the sequence of interest. B_{ij} is the probability that being in the state i, the $j-th$ element of the sequence of characteristics is emitted. This probability is obtained by first evaluating the $j-th$ vector belonging to the sequence in each one of the multivariate gaussians found by the model, according to the following equation:

$$logf(x) = -\frac{k}{2}log(2\pi) - \frac{1}{2}log\,|\Sigma| - \frac{1}{2}(x-\mu)^T(x-\mu) \tag{5.8}$$

The values obtained are then normalized and a matrix obtained by evaluating the exponential function on each of them. This means that for each one of the sequences of vectors a matrix B is obtained, which is then used in the forward algorithm.

The models were tested both for the sequence of angles of interest and sequences of coordinates. Since it was found that the probabilities assigned to the sequences did not always follow a normal distribution, the methodology by which the success of the approach was tested was building an interval around the mean of the probabilities for the training sequences, and then determining for the testing sequences if the probability assigned by the model lied in this interval or not. The interval is then expanded around the mean, until finally it becomes the whole interval of probabilities possible for the training sequences.

Each one of the values of interval width allows for the calculation of a false positive rate (FPR) and a true positive rate (TPR), and finally a ROC curve can be constructed showing the performance of the model seen as a two-class classifier, with one class being the correct performance and the other one being the specific deviation that we are interested in. The process is repeated 10 times, this because the training sequences are randomly divided in 10 groups, and each time 9 of these groups would be used for training while the other one would be used for testing. This guarantees that only some of the correct sequences are used to train the model, while some others are used for testing. The final ROC curve is simply the average of the 10 curves obtained for each execution of the algorithm.

Finally, the area under the curve is taken as a metric to give an idea of the classification accuracy. The value of these areas is presented in table 5.3.

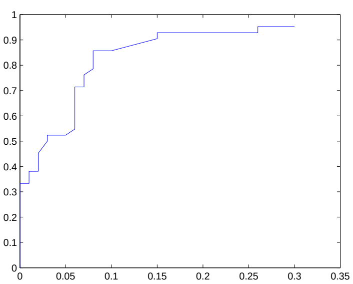


Figure 5.1: Angles

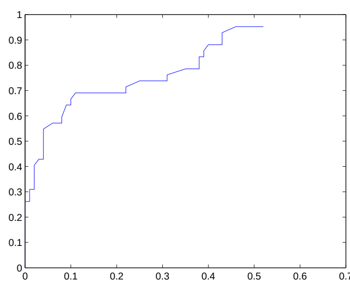


Figure 5.2: Coordinates

Figure 5.3: ROC curve using both angles (a) and coordinates (b), the activity considered is shoulder abduction and the error is a type 2 error on the phase 2

For the comparison of the two populations of interest, four characteristics were calculated based on the three movements each subject was asked to perform. Figure 5.4 illustrates one of the signals calculated using *kinect* data. Each one of these signals is segmented after taking its Discrete Fourier Transform and locating the first peak on the spectrum, this frequency is used to calculate the value of the signal's period, and for each one of the repetitions thus obtained four quantities are calculated, these are, the mean of the angle value in the upper portions of the signal, the mean of the angle value in the lower portions, the average and the amplitude of the repetition (which corresponds to the difference between the mean of the upper portions and the mean of the lower portions).

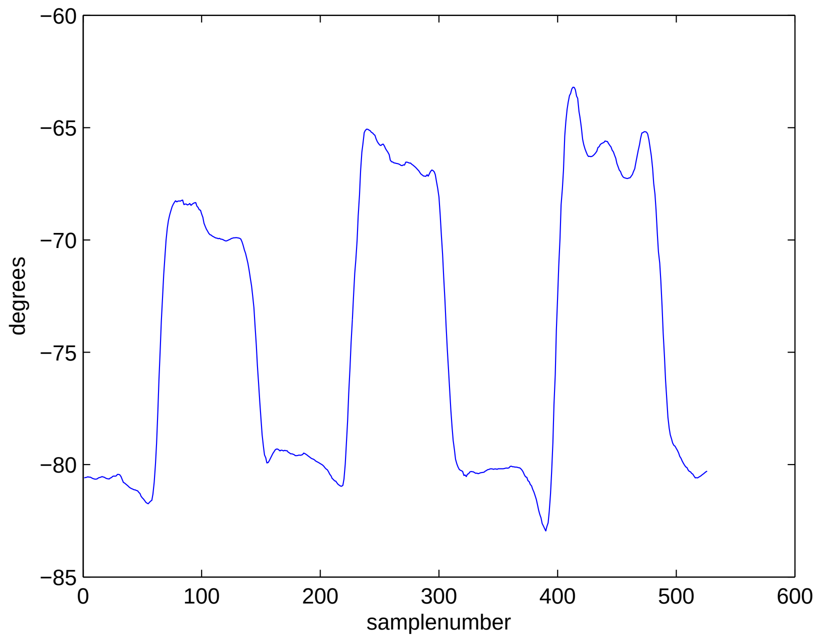


Figure 5.4: Transversal angle calculated for the Hip Flexion repetitions of a subject

5.4 Experimental Setup

Initially, 14 people were asked to perform three repetitions of the activities of interest. These are shoulder abduction, hip abduction, shoulder flexion, shoulder extension,hip flexion, hip extension, shoulder internal rotation, shoulder external rotation, elbow flexion and, elbow extension. They were asked to keep the limb of interest for about five seconds in the final position of the activity before returning it to the initial position. These database is considered as the basis for training the models.

In order to obtain examples of deviation sequences, 10 people were asked to perform movements close to the specification, but deviating from them in one of the planes of motion1.1. In the case of abduction movements deviations were introduced with respect to the frontal plane, which means that the subjects were asked to extend their limbs towards the front of their bodies (error type 1) or towards the back (error type 2). In the case of extension and flexion movements the deviations introduced affected the movement with respect to the sagittal plane, and people were asked to move the limb of interest either towards the left or the right. This database is used to test the ability of the models to detect deviations.

In order to obtain data from an elderly population, 8 people between the ages of 65 and 75 were asked to perform the following physical conditioning activities: shoulder abduction, hip abduction, shoulder flexion, shoulder extension,hip flexion, hip extension, shoulder internal rotation, shoulder external rotation, elbow flexion and, elbow extension. They were asked to perform three repetitions of each one of the 10 activities, keeping the limb of interest still for about five seconds before returning to the initial position. This database is used to compare the measurements taken by the sensor for different age groups.

All measurements were performed with the sensor standing in a table 70 cm above the ground. All of the subjects were standing at 1.8 m in front of the sensor.

In order to test the validity of the proposed approach to detect deviations from the correct movement, an additional dataset was analyzed. These data comes from the work of Velloso et al.[2], in which a weightlifting exercise, unilateral dumbbell biceps curl, is studied. Five different executions of the activity are studied, class A is the execution of the activity according to the specification, class B is the execution throwing the elbows to the front, class C is lifting the dumbbell only halfway, class D is lowering the dumbbell only halfway and class E is throwing the hips to the front. A total of six people were asked to perform 10 repetitions of each one of the activities.

5.5 Results

This section presents the results of analyzing the performance of the models trained to recognize sequences of movement, when the model is trained with correct repetitions and tested with sequences of deviations.In the case of the *kinect* data for each one of the two kinds of errors considered in each one of the 10 different activities of interest. Each one of the movements considered in the database is segmented in two phases, one in which the limb is moving from the initial position towards the final position, and another one that returns the limb to the initial position, hence the existence of four kinds of errors, depending on which one of the phases of the movement is being considered.

Results are presented both for discrete HMMs and continuous HMM.Also a comparison between the performance of the two kinds of models is presented.

Results are also presented for the use of the methodology on the dataset by Velloso et al. The ROC curves are presented, which prove that varying the width of the probability interval obtained with correct performances does not increase much the false positive rate (FPR), or in other words, that the model trained for the correct performance is capable of detecting all the different kinds of deviations introduced and label them as abnormalities.

Table 5.1: Error detection accuracy for continuous HMM models

	Type 1 phase 1		Type 2 phase 1		Type 1 phase 2		Type 2 phase 2	
	angles	coord.	angles	coord.	angles	coord.	angles	coord.
Shoulder Abduction	0.774	**0.915**	**0.912**	0.830	0.800	**0.952**	**0.882**	0.835
Hip Abduction	**0.812**	0.762	**0.844**	0.759	**0.784**	0.756	**0.781**	0.732
Hip Extension	**0.835**	0.603	**0.890**	0.637	**0.859**	0.517	**0.911**	0.592
Elbow Extension	0.851	**0.943**	0.776	**0.907**	0.835	**0.896**	0.744	**0.895**
Shoulder Extension	**0.878**	0.612	**0.858**	0.572	**0.886**	0.684	**0.905**	0.609
Hip Flexion	0.684	**0.887**	0.750	**0.793**	0.652	**0.843**	**0.758**	0.632
Elbow Flexion	**0.946**	0.907	0.889	**0.944**	**0.923**	0.871	0.849	**0.938**
Shoulder Flexion	0.885	**0.908**	0.722	**0.903**	0.912	0.912	0.771	**0.941**
External Rotation	**0.834**	0.804	**0.901**	0.818	0.807	**0.808**	**0.878**	0.794
Internal Rotation	**0.876**	0.799	**0.872**	0.719	**0.869**	0.800	**0.837**	0.752

It is clear from Table 5.1 that for most of the deviations in the database the approach based on continuous HMM is capable of labeling the incorrect sequences as such in about 80% of the cases by using either one of the sequences of characteristics, be it angles or coordinates. In total four kinds of mistakes were considered for each of the activities, and that means 40 different errors are being considered. It was found that the approach based on angles was superior to the one based on coordinates, in 24 of the 40 cases.

The continuous HMM approach can be compared to the discrete approach by using the same metric (area under the ROC curve). Table 5.2 shows the result of performing the same experiment but with discrete HMM models. Since it was necessary to use three different models for each one of the activities under study, the results presented in the table are the ones for the most informative model, according to the error detection percentages that were presented in chapter 3. For example, in the case of Hip Flexion, the analysis were carried out using the sagittal model, which is more capable of labeling sequences as deviations from normality.

Table 5.3 compares the best performance for any characteristic of the two proposed approaches. It is found that the use of continuous HMM models allows for a better detection of abnormal sequences in 27 of the 40 different errors for the 10 activities that were considered.

Table 5.2: Error detection accuracy for discrete HMM models

	Type 1 phase 1		Type 2 phase 1		Type 1 phase 2		Type 2 phase 2	
	angles	coord.	angles	coord.	angles	coord.	angles	coord.
Shoulder Abduction	**0.888**	0.746	**0.940**	0.674	**0.901**	0.711	**0.896**	0.638
Hip Abduction	**0.768**	0.385	**0.584**	0.180	**0.708**	0.451	**0.544**	0.231
Hip Extension	**0.810**	0.214	**0.717**	0.293	0.859	**0.925**	0.781	**0.839**
Elbow Extension	**0.830**	0.019	**0.839**	0.300	0.879	**0.906**	0.844	**0.893**
Shoulder Extension	**0.623**	0.488	**0.715**	0.535	0.661	**0.781**	0.577	**0.787**
Hip Flexion	0.664	**0.907**	0.564	**0.813**	0.673	**0.833**	0.508	**0.812**
Elbow Flexion	0.721	**0.844**	0.746	**0.825**	0.604	**0.937**	0.756	**0.929**
Shoulder Flexion	**0.561**	0.094	**0.659**	0.277	**0.649**	0.092	**0.747**	0.184
External Rotation	0.624	**0.909**	0.636	**0.900**	**0.597**	0.223	**0.691**	0.283
Internal Rotation	0.529	**0.877**	0.559	**0.889**	0.556	**0.902**	0.502	**0.902**

Table 5.3: Comparison of continuous and discrete HMM performances

	Type 1 phase 1		Type 2 phase 1		Type 1 phase 2		Type 2 phase 2	
	cont.	discrete	cont.	discrete	cont.	discrete	cont.	discrete
Shoulder Abduction	**0.915**	0.889	0.912	**0.940**	**0.952**	0.901	0.882	**0.895**
Hip Abduction	**0.812**	0.786	**0.845**	.0584	**0.784**	0.708	**0.781**	0.544
Hip Extension	**0.835**	0.810	**0.999**	0.717	0.859	**0.925**	**0.911**	0.839
Elbow Extension	**0.943**	0.830	**0.907**	0.839	0.896	**0.906**	**0.895**	0.893
Shoulder Extension	**0.878**	0.623	**0.857**	0.715	**0.886**	0.781	**0.905**	0.787
Hip Flexion	0.887	**0.907**	0.793	**0.813**	**0.843**	0.833	0.758	**0.812**
Elbow Flexion	**0.945**	0.844	**0.944**	0.825	0.923	**0.937**	**0.938**	0.929
Shoulder Flexion	**0.908**	0.561	**0.903**	0.659	**0.912**	0.649	**0.941**	0.747
External Rotation	0.834	**0.909**	**0.901**	0.900	**0.808**	0.597	**0.878**	0.691
Internal Rotation	0.876	**0.877**	0.872	**0.889**	0.869	**0.902**	0.837	**0.902**

Experiments were performed to determine if the measured characteristics (the angles formed with respect to the planes of motion) were consistent both in a group of young people (aged 20 to 30 years old) and in a group of elder people (aged 65 to 75 years old). In order to do this the average value of 4 characteristics was calculated for each one of the subjects on each one of the two databases. The characteristics of interest were the mean value of each one of the angles on both the final and initial position of the activities, the amplitude of the movement (that is to say, the difference between the final and initial positions) and the mean of the angle sequence. For each one of the characteristics a two tailed t-test supposing unequal variances was used to determine if there was enough statistical evidence to reject the null hypothesis that the means of the two populations were different. The importance of the analysis lies in the ability to use data taken from a population and use it to create a model than can be applied to both populations.

The results of the analysis are shown in table 5.4. The table shows for each one of the activities of interest and each one of the angles calculated, and for each one of the characteristics calculated if the means of the two populations are equal or not. In the cases were there is no statistical evidence to reject the null hypothesis that the means are different, the table shows equal. In the opposite case the table shows different.

Table 5.4: Result of the t-test for comparison of the features of interest between the two populations

Activity	Angle	High	Low	Amplitude	Mean
Hip Abduction	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	different	equal
	Sagittal	different	equal	different	equal
Shoulder Abduction	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	equal	different
	Sagittal	equal	equal	equal	equal
Hip Extension	Frontal	equal	equal	different	different
	Transverse	equal	equal	equal	equal
	Sagittal	equal	equal	equal	equal
Elbow Extension	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	equal	equal
	Sagittal	different	equal	different	different
Shoulder Extension	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	equal	equal
	Sagittal	equal	different	equal	equal
Hip Flexion	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	equal	equal
	Sagittal	different	equal	equal	equal
Elbow Flexion	Frontal	different	different	equal	different
	Transverse	equal	different	equal	different
	Sagittal	equal	equal	equal	different
Shoulder Flexion	Frontal	different	equal	different	different
	Transverse	different	equal	different	different
	Sagittal	equal	equal	equal	equal
External Rotation	Frontal	different	different	equal	different
	Transverse	equal	equal	equal	different
	Sagittal	equal	equal	equal	equal
Internal Rotation	Frontal	equal	equal	equal	equal
	Transverse	equal	equal	equal	different
	Sagittal	equal	different	equal	equal

Only for two activities, shoulder flexion and elbow flexion it was found that there were statistically significant differences in all three characteristics, which means that thresholds calculated using data taken from young people will not work when it comes to using the system with elderly people.

Figure 5.7 shows the ROC curve for the model trained with repetitions of the correct form. It is clear from the curves that the model is capable of detecting those two kinds of mistakes, since the FPR does not increase consistently while the width of the probability interval was varied, instead it remained at less than 0.3. What this means is that The horizontal line was added to the graphs because after the anaylisis ended it was found that the point (1,1) was never reached.

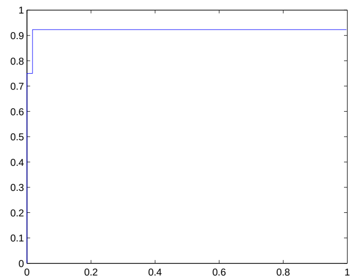


Figure 5.5: Error type B

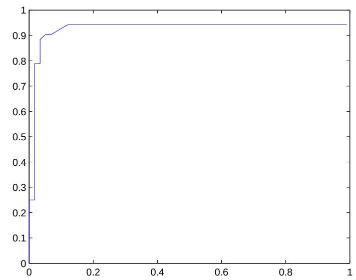


Figure 5.6: Error type C

Figure 5.7: ROC curves for errors type B and C

Figure 5.10 shows the result of the model considered as a binary classifier when analyzing error types D and E. Again it was found that the FPR did not increase significantly when the width of the interval of probabilities is increased. The TPR however reaches an acceptable value of about 0.9, meaning that the vast majority of the sequences considered as correct are classified as such.

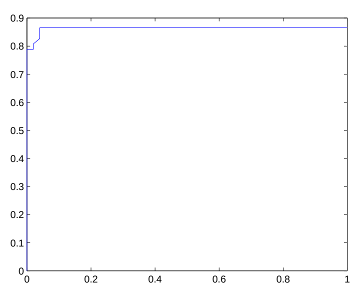


Figure 5.8: Error type D

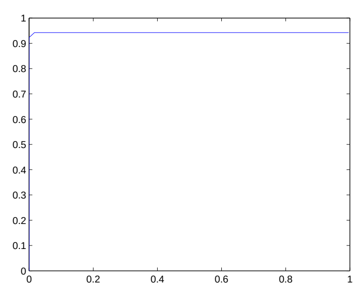


Figure 5.9: Error type E

Figure 5.10: ROC curves for errors type D and E

5.6 Conclusion

It was found that it is possible to use continuous Hidden Markov Models to detect deviations from the specification for movements in the field of physical therapy and conditioning.

Experiments show that continuous Hidden Markov Models outperform discrete Hidden Markov Models for the task of detecting deviations in movement sequences. They also have the additional advantage of being able to deal with multivariate data, which means that they can be readily applied to abnormality detection of any sequential process represented with multivariate series of data.

Experiments show that for a wide range of movements common in the fields of physical therapy and conditioning the ranges of movement measured with the kinect are statistically the same for elder people and young people. This validates the use of data gathered with young people for the purpose of developing coaching applications for the elder, when it comes to training the models used for recognition of correct performances.

The methodology proposed can also be applied to other sensor data and renders good results, by being able to detect correctly sequences that correspond to different kinds of deviations while at the same time successfully classifying the correct sequences as such.

Chapter 6

Software prototype

?? and **??** are diagrams showing the matlab script integrating the characterization and evaluation tasks. The best technique according to the experimental results was selected to perform the evaluation step, this is continuous HMM. The idea is for these routine to be able to identify as many erroneous sequences as possible.

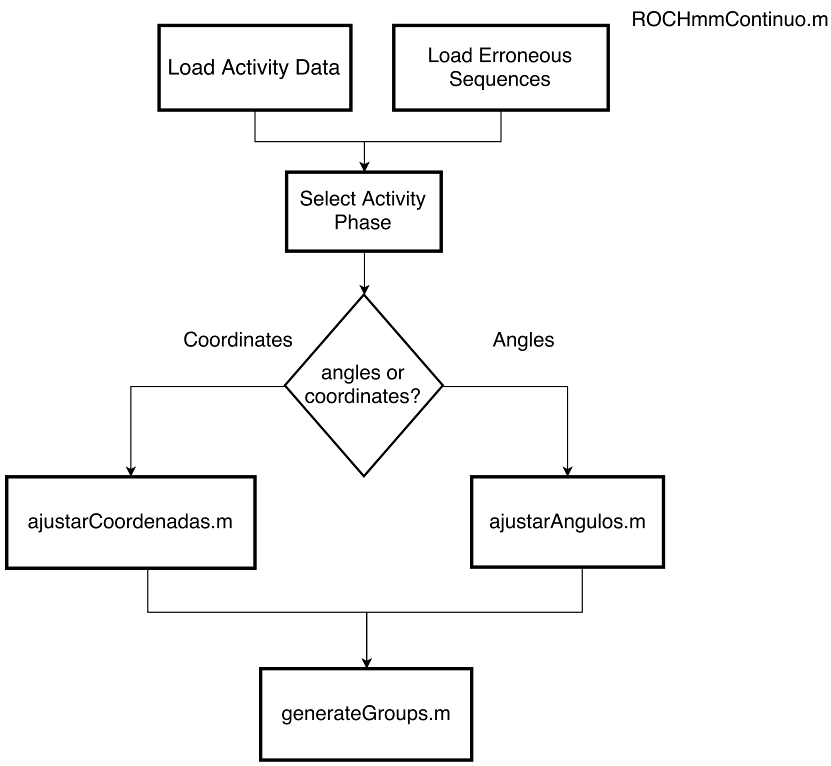


Figure 6.1: Workflow of the software prototype

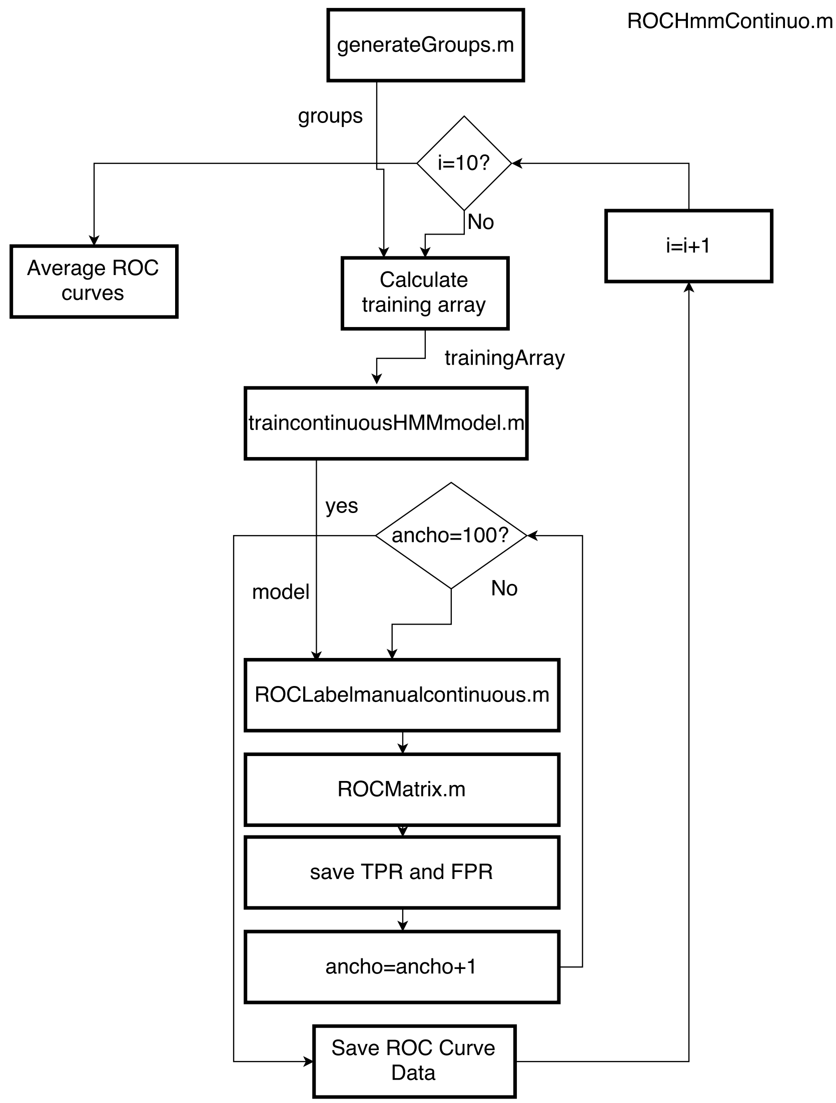


Figure 6.2: Workflow of the software prototype (continued)

It is important to note that part of the libraries developed during this study were used as the core of a software aimed at providing the elderly with a tool to exercise at home. The name of the software is SAMSOM (Sistema de Asistencia al Mejoramiento Supervisado y Objetivo de la Movilidad). Figures 6.3 and 6.4 illustrate the developed software.

Figure 6.3: SAMSOM

Figure 6.4: SAMSOM

Chapter 7

Conclusions and future work

In this work several databases of human beings performing activities common in the fields of physical therapy and conditioning were created. The first database consists of 14 people performing the movements of interest three times. It was created using both a single kinect system and a dual kinect array. The second database consists of 10 people performing the same activities but introducing deviations from the correct form, two error classes were defined for each one of the activities. Finally a database of 8 elderly people performing the activities was created, in order to compare their performances to those of the original subjects.

In this work a set of characteristics was used, that consists in the angles formed by the limb of interest with respect to the planes of motion commonly used to define activities in the fields of physical therapy and conditioning. These characteristics were tested and proved to be superior to the use of the coordinates provided by the sensor for the task of defining a wide variety of movements, even when these coordinates were normalized to account for the variability inherent in the human population.

The use of information coming from two *kinect* sensors was studied. This was done to deal with self occlusion problems that are present in some of the activities studied. It was found that an optimization scheme that minimizes the distance between the coordinates of the sensors offers improvement when the movements to track imply self-occlusions.

The use of distance measures was also studied, in order to try to define a sequence that models the correct performance for a variety of people.

The use of statistical models for the task of abnormal performance recognition was studied. These models have to be trained with samples of what is considered normal instances of the activities. They are then able to assign a probability to every new sequence arriving from the sensor and determine whether this sequence corresponds to a normal or abnormal situation.

Comparison between the distance approaches and statistical modeling showed that the first approach is incapable of obtaining a model that might be used for a variety of subjects. On the other side, statistical approaches can generate models applicable to different subjects that yield an acceptable recognition rate of deviations.

Two types of statistical models were studied in this work, discrete Hidden Markov Models, and continuous Hidden Markov Models. Experiments show that continuous Hidden Markov Models yield better results for the task of detecting movements that deviate from the specification in human beings.

A methodology that consists in the acquisition of skeleton data from a kinect sensor, calculation of the angles formed with the planes of motion, training of a statistical model with repetitions that are considered to be correct and evaluation of sequences by using these model was proposed and implemented on Matlab™.

During this study it was found that commercial, affordable technology for real-time skeletonization of a human being still lacks the ability to extract a human being's skeleton with accuracy when the person rotates in front of the sensor. In the case of movements that take place while the persons remains seated or is laying on his back, the technology used is incapable of extracting any kind of skeleton from him. This limitation severely affects computer vision based applications that aim to create tools for rehabilitation or physical therapy, due to the existence of many activities of interest in such areas that are performed laying on a hospital bed. Further research is needed in order to address this limitation.

During this study it was found that statistical models used to model sequences of data can be used to detect abnormalities in sequences coming from the performance of a physical activity by a human being. It still remains of interest to determine if further processing might allow for the recognition of the specific deviation that took place. This would mean that the statistical model should be capable of labeling the sequences that are considered abnormal.

Bibliography

[1] Bassem Elsawy and Kim E Higgins. “Physical activity guidelines for older adults”. In: *American family physician* 81.1 (2010), pp. 55–9.

[2] Eduardo Velloso et al. “Qualitative activity recognition of weight lifting exercises”. In: *Proceedings of the 4th Augmented Human International Conference*. ACM. 2013, pp. 116–123.

[3] J. McLester and P.S. Pierre. *Applied Biomechanics: Concepts and Connections*. Cengage Learning, 2007. ISBN: 9780495105862. URL: `http://books.google.com.co/books?id=JsFvOy1w-1QC`.

[4] Noitom Ltd. *’Perception Neuron’*. `https://neuronmocap.com/`. Apr. 2015.

[5] Vicon Motion Systems Ltd. UK. *’Vicon’*. `http://www.vicon.com/`. Feb. 2016.

[6] Organic Motion. *’Organic Motion’*. `http://www.organicmotion.com/`. Feb. 2006.

[7] ASUSTeK Computer Inc. *’Asus Xtion’*. `https://www.asus.com/3D-Sensor/Xtion_PRO/`. Feb. 2016.

[8] Microsoft. *’Microsoft Kinect’*. `https://dev.windows.com/en-us/kinect`. Feb. 2016.

[9] Joveria Javed, Hashim Yasin, and Syed Faisal Ali. “Human movement recognition using euclidean distance: a tricky approach”. In: *Image and Signal Processing (CISP), 2010 3rd International Congress on*. Vol. 1. IEEE. 2010, pp. 317–321.

[10] Matheen Siddiqui and Gérard Medioni. “Human pose estimation from a single view point, real-time range sensor”. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE. 2010, pp. 1–8.

[11] L Enrique Sucar et al. “Gesture therapy: A vision-based system for arm rehabilitation after stroke”. In: *Biomedical Engineering Systems and Technologies*. Springer, 2009, pp. 531–540.

[12] JW Burke et al. “Vision based games for upper-limb stroke rehabilitation”. In: *Machine Vision and Image Processing Conference, 2008. IMVIP’08. International*. IEEE. 2008, pp. 159–164.

[13] Jamie Shotton et al. "Real-time human pose recognition in parts from single depth images". In: *Communications of the ACM* 56.1 (2013), pp. 116–124.

[14] Tilak Dutta. "Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace". In: *Applied ergonomics* 43.4 (2012), pp. 645–649.

[15] Gregorij Kurillo et al. "Evaluation of upper extremity reachable workspace using Kinect camera". In: *Technology and Health Care* 21.6 (2013), pp. 641–656.

[16] S Obdrzalek et al. "Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population". In: *Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE*. IEEE. 2012, pp. 1188–1193.

[17] Adso Fernandez-Baena, Antonio Susín, and Xavier Lligadas. "Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments". In: *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on*. IEEE. 2012, pp. 656–661.

[18] Wenbing Zhao et al. "A feasibility study of using a single Kinect sensor for rehabilitation exercises monitoring: A rule based approach". In: *Computational Intelligence in Healthcare and e-health (CICARE), 2014 IEEE Symposium on*. IEEE. 2014, pp. 1–8.

[19] Bruno Bonnechere et al. "Determination of Repeatability of Kinect Sensor". In: *Telemedicine and e-Health* 20.5 (2014), pp. 451–453.

[20] Bruno Bonnechere et al. "Determination of the precision and accuracy of morphological measurements using the Kinect™ sensor: comparison with standard stereophotogrammetry". In: *Ergonomics* 57.4 (2014), pp. 622–631.

[21] Mike van Diest et al. "Suitability of Kinect for measuring whole body movement patterns during exergaming". In: *Journal of biomechanics* 47.12 (2014), pp. 2925–2932.

[22] Noriyuki Iwane. "Arm movement recognition for flag signaling with Kinect sensor". In: *Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2012 IEEE International Conference on*. IEEE. 2012, pp. 86–90.

[23] Zahoor Zafrulla et al. "American sign language recognition with the kinect". In: *Proceedings of the 13th international conference on multimodal interfaces*. ACM. 2011, pp. 279–286.

[24] Thiago Chaves et al. "Human Body Motion and Gestures Recognition Based on Checkpoints". In: *Virtual and Augmented Reality (SVR), 2012 14th Symposium on*. IEEE. 2012, pp. 271–278.

[25] Linwan Liu et al. "Static Human Gesture grading based on Kinect". In: *Image and Signal Processing (CISP), 2012 5th International Congress on.* IEEE. 2012, pp. 1390–1393.

[26] Lee Jaemin et al. "A robust gesture recognition based on depth data". In: *Frontiers of Computer Vision,(FCV), 2013 19th Korea-Japan Joint Workshop on.* IEEE. 2013, pp. 127–132.

[27] Ye Gu et al. "Human gesture recognition through a kinect sensor". In: *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on.* IEEE. 2012, pp. 1379–1384.

[28] Alexandros Andre Chaaraoui, José Ramón Padilla-López, and Francisco Flórez-Revuelta. "Fusion of Skeletal and Silhouette-based Features for Human Action Recognition with RGB-D Devices". In: *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on.* IEEE. 2013, pp. 91–97.

[29] Antonio Bo, Mitsuhiro Hayashibe, Philippe Poignet, et al. "Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect". In: *EMBC'11: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 2011, pp. 3479–3483.

[30] Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang. "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities". In: *Research in developmental disabilities* 32.6 (2011), pp. 2566–2570.

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

[32] Isaac Pastor Acosta. "Upper Limb Rehabilitation of Stroke Patients Using Kinect and Computer Games". PhD thesis. The University of Utah, 2012.

[33] Hossein Mousavi Hondori and Maryam Khademi. "A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation". In: *Journal of Medical Engineering* 2014 (2014).

[34] *Nike Kinect Training Home Page.* `http://http://www.nike.com/us/en_us/c/training/nike-plus-kinect-training`. Accessed: 2015-01-22.

[35] Keizo Sato et al. "The effects of exercise intervention using Kinect TM on healthy elderly individuals: A quasi-experimental study". In: *Open Journal of Therapy and Rehabilitation* 2014 (2014).

[36] Jaime A Garcia et al. "Exergames for the elderly: Towards an embedded Kinect-based clinical test of falls risk". In: *Stud Health Technol Inform* 178 (2012), pp. 51–7.

[37] Hiroki Kayama et al. "Efficacy of an exercise game based on Kinect in improving physical performances of fall risk factors in community-dwelling older adults". In: *GAMES FOR HEALTH: Research, Development, and Clinical Applications* 2.4 (2013), pp. 247–252.

[38] Aarthi Ravi. "Automatic Gesture Recognition and Tracking System for Physiotherapy". In: *Electrical Engineering and Computer Sciences University of California at Berkeley* (2013).

[39] Adeline Paiement et al. "Online quality assessment of human movement from skeleton data". In: *Computing* 27.1 (2009), pp. 153–166.

[40] Dionysia Agathocleous. "Virtual Personal Training". MA thesis. the Netherlands: University of Amsterdam, 2013.

[41] David Maung et al. "Games for therapy: Defining a grammar and implementation for the recognition of therapeutic gestures." In: *FDG*. 2013, pp. 314–321.

[42] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. "MotionMA: Motion modelling and analysis by demonstration". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 1309–1318.

[43] R.T. Floyd and C.W. Thompson. *Manual of Structural Kinesiology*. McGraw-Hill, 2004. ISBN: 9780072558913. URL: `http://books.google.com.co/books?id=Fc9qAAAAMAAJ`.

[44] Alana Da Gama et al. "Guidance and movement correction based on therapeutics movements for motor rehabilitation support systems". In: *Virtual and Augmented Reality (SVR), 2012 14th Symposium on*. IEEE. 2012, pp. 191–200.

[45] Naofumi Kitsunezaki et al. "Kinect applications for the physical rehabilitation". In: *Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on*. IEEE. 2013, pp. 294–299.

[46] *Resistance Band Tubing and Instruction Manual*. `http://www.thera-band.com/userfiles/file/resistance_band-tubing_instruction_manual.pdf`. Accessed: 2015-01-26.

[47] Carlos Palma, Augusto Salazar, and Francisco Vargas. "Real time Kinect evaluation of therapeutical gestures". In: *Signal Processing, Images and Computer Vision (STSIVA), 2015 20th Symposium on*. IEEE. 2015, pp. 1–7.

[48] Ross A Clark et al. "Validity of the Microsoft Kinect for assessment of postural control". In: *Gait & posture* 36.3 (2012), pp. 372–377.

[49] Chien-Yen Chang et al. "Towards pervasive physical rehabilitation using Microsoft Kinect". In: *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*. IEEE. 2012, pp. 159–162.

[50] Jing Tong et al. "Scanning 3d full human bodies using kinects". In: *Visualization and Computer Graphics, IEEE Transactions on* 18.4 (2012), pp. 643–650.

[51] HwanIk Jo et al. "Motion Tracking System for Multi-User with Multiple Kinects". In: *International Journal of u-and e-Service, Science and Technology* 8.7 (2015), pp. 99–108.

[52] Nur Aziza Azis, Ho-Jin Choi, and Youssef Iraqi. "Substitutive skeleton fusion for human action recognition". In: *Big Data and Smart Computing (BigComp), 2015 International Conference on*. IEEE. 2015, pp. 170–177.

[53] Kwok-Yun Yeung, Tsz-Ho Kwok, and Charlie CL Wang. "Improved Skeleton Tracking by Duplex Kinects: A Practical Approach for Real-Time Applications". In: *Journal of Computing and Information Science in Engineering* 13.4 (2013), p. 041007.

[54] *Skeletal Joint Smoothing White Paper*. `https://msdn.microsoft.com/en-us/library/jj131429.aspx`. Accessed: 2015-02-17.

[55] E Machida et al. "Human motion tracking of mobile robot with Kinect 3D sensor". In: *SICE Annual Conference (SICE), 2012 Proceedings of*. IEEE. 2012, pp. 2207–2211.

[56] Jody Shu, Fumio Hamano, and John Angus. "Application of extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates". In: *Journal of Engineering Mathematics* 88.1 (2014), pp. 161–175.

[57] Wenbing Zhao et al. "Rule based realtime motion assessment for rehabilitation exercises". In: *Computational Intelligence in Healthcare and e-health (CICARE), 2014 IEEE Symposium on*. IEEE. 2014, pp. 133–140.

[58] Carlos Palma, Augusto Salazar, and Francisco Vargas. "HMM Based Evaluation of Physical Therapy Movements Using Kinect Tracking". In: *Advances in Visual Computing*. Springer, 2015, pp. 174–183.

[59] Ahmed Nabil Mohamed. "A Novice Guide towards Human Motion Analysis and Understanding". In: *arXiv preprint arXiv:1509.01074* (2015).

[60] Hiroaki Sakoe and Seibi Chiba. "Dynamic programming algorithm optimization for spoken word recognition". In: *Acoustics, Speech and Signal Processing, IEEE Transactions on* 26.1 (1978), pp. 43–49.

[61] C Su, J Huang, S Huang, et al. "Ensuring Home-based rehabilitation Exercise by Using Kinect and Fuzzified Dynamic Time Warping Algorithm". In: *Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference*. 2012, pp. 884–895.

[62] Manuel P Cuellar et al. "An approach for the evaluation of human activities in physical therapy scenarios". In: *Mobile Networks and Management*. Springer, 2015, pp. 401–414.

[63] Ryan Staab. "Recognizing specific errors in human physical exercise performance with Microsoft Kinect". In: (2014).

[64] Padhraic Smyth. "Markov monitoring with unknown states". In: *Selected Areas in Communications, IEEE Journal on* 12.9 (1994), pp. 1600–1612.

[65] Zheping Yan, Dongnan Chi, and Chao Deng. "An Outlier Detection Method with Wavelet HMM for UUV Prediction Following". In: (2013).

[66] Junan Zhu, Zhiqiang Ge, and Zhanfeng Song. "HMM-driven Robust Probabilistic Principal Component Analyzer for Dynamic Process Fault Classification". In: (2015).

[67] *DTWAlgorithm.* `http://www.psb.ugent.be/cbd/papers/gentxwarper/DTWAlgorithm.ppt`. Accessed: 2015-02-17.

[68] Lawrence R Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[69] Carlos Palma, Augusto Salazar, and Francisco Vargas. "Automatic Detection of Deviations in Human Movements Using HMM: Discrete vs Continuous". In: *International Symposium on Visual Computing.* Springer. 2016, pp. 534–543.

[70] Chen Wang, Hongzhi Lin, and Hongbo Jiang. "Trajectory-based multidimensional outlier detection in wireless sensor networks using Hidden Markov Models". In: *Wireless Networks* 20.8 (2014), pp. 2409–2418.

[71] Anisa Allahdadi, Ricardo Morla, and Jaime S Cardoso. "Outlier detection in 802.11 wireless access points using Hidden Markov Models". In: *Wireless and Mobile Networking Conference (WMNC), 2014 7th IFIP*. IEEE. 2014, pp. 1–8.

[72] Su Yang and Weihua Liu. "Anomaly detection on collective moving patterns: A hidden markov model based solution". In: *Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing.* IEEE. 2011, pp. 291–296.

[73] Yingfeng Cai et al. "Trajectory-based anomalous behaviour detection for intelligent traffic surveillance". In: *Intelligent Transport Systems, IET* 9.8 (2015), pp. 810–816.

[74] Yuan Yuan et al. "Continuous Change Detection and Classification Using Hidden Markov Model: A Case Study for Monitoring Urban Encroachment onto Farmland in Beijing". In: *Remote Sensing* 7.11 (2015), pp. 15318–15339.

[75] J Du Toit and JH Van Vuuren. "Semi-automated maritime vessel activity detection using hidden Markov models". In: *Proceedings of the 43rd Annual Conference of the Operations Research Society of South Africa, Parys*. 2014, pp. 71–78.

[76] Kevin Murphy and Dunham Matt. *'Probabilistic Modeling Toolkit'*. `https://github.com/probml/pmtk3`. Dec. 2011.