University of Cincinnati

Date: 6/27/2014

I, **Benson Isaac**, hereby submit this original work as part of the requirements for the degree of Master of Science in Mechanical Engineering.

It is entitled:

Inverse Kinematics and Extended Kalman Filter based Motion Tracking of Human Limb

Student's name: **Benson Isaac**

This work and its defense approved by:

Committee chair: Manish Kumar, Ph.D.

Committee member: Hongdao Huang, Ph.D.

Committee member: David Thompson, Ph.D.

10581

Inverse Kinematics and Extended Kalman Filter based Motion Tracking of Human Limb

A thesis submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Master of Science

in the Department of Mechanical and Materials Engineering
of the College of Engineering and Applied Sciences

by

Benson Isaac

Bachelor of Technology
College of Engineering Trivandrum
May 2009

Committee Chair: Dr. Manish Kumar

Abstract

The advent of portable video and image sensing devices such as Microsoft Kinect have provided computer scientists and engineers an opportunity to exploit data obtained from these devices to help with monitoring different aspects of human activities. However, human motion tracking has always been a challenging area of research. The tracking process has an additional dimension of complexity when the results are solely dependent on the vision based sensors and when the tracked user is not wearing any contact sensors such as accelerometers, gyroscopes, etc. This research addresses the problem of motion tracking of human skeleton system using non-wearable vision based sensors. The application areas of such a system will be in patient care and nursing homes where the focus is on having passive sensors that do not affect the day to day activities of the individuals living in such homes.

The proposed approach combines synergistic paradigms of image processing, kinematics of rigid bodies and Extended Kalman Filtering scheme to estimate the motion of a human skeletal system which may be studied as a system of kinematic linkages. The unique aspect of this research is that, this approach solely depends on the measurement obtained from the vision sensors without involving any wearable or inertial sensors to measure the motion parameters. In this research we propose fusion of two filtering schemes- the optical flow equations applied to raw images obtained from the Microsoft Kinect and extended Kalman filter for human skeleton considered as a system of kinematic linkages. The novel approach proposed in this research presents a an effective

estimation algorithm which is tested with the help of experiments performed using the Microsoft Kinect sensor and compared using accurate tracks obtained from 24-Camera Opti-track motion capture system.

Acknowledgement

I owe a great deal of gratitude to several people for completion of my thesis. First of all, I am eternally indebted to my advisor, Dr. Manish Kumar, working with whom I have learned so much. Under his constructive and valuable guidance, I have developed as a strong and competent scholar, a finer professional and altogether a better person.

I am grateful to my thesis committee members, Dr. Samuel Huang and Dr. David Thompson, whose valuable insights and technical guidance have helped me during my research work. I also thank Prof. Dan Humpert, who has helped me with his inputs and support. His enterprising collaborations with College of Nursing and Maple Knoll Village have helped me understand the impact of human-centered approach to research. Dr. Kelly Cohen has been a source of inspiration for me during different stages of my thesis work.

I would like to acknowledge the lively environment maintained by Alireza Nemati, Baisravan Homchaudhuri, Balaji Sharma, Gaurav Mukherjee, Ruoyu Tan and Sushil Garg, my fellow researchers at Cooperative Distributed Systems Lab. I also thank my department and Larry Schartman for all the support. I thank my roommates and friends for always being there when I needed anything. My family has always been supportive and encouraging. I am blessed to have their prayers, confidence and loving care in every aspect of my life.

Table of Contents

Abstract	i
Acknowledgement	iv
1. Introduction.....	1
1.1 Motivation	2
1.2 Study Goals and Methodology	3
1.3 Research Objectives.....	5
1.4 Thesis Outline	6
2. Background	7
2.1 Existing Methods in Human Motion Tracking.....	9
2.2 Different Sensors used for Motion Tracking	12
2.3 Human Motion Kinematics and Activity Recognition	17
3. Algorithms used in Motion Tracking	20
3.1 Extended Kalman Filter Algorithm	23
3.2 Optical Flow Equations	26
3.3 Decision Trees	28
4. Problem Formulation and Approach	31
4.1 The Kinematic Model.....	31
4.2 Kalman Filter for Optical Flow Constrained Based Estimation	35
4.2.1 The Kalman Estimation.....	35
4.2.2 The measurement data and equations.....	36
4.2.3 The prediction and correction equation.....	36
4.3 Extended Kalman Filter for Determining the Position Coordinates	40
4.3.1 The Kalman Estimation.....	40
4.3.2 Measurement Data.....	41
4.3.3 The time and measurement update.....	42
5. Simulation Results and Discussions	46

5.1	Experimental Setup.....	46
5.2	Simulation Results.....	47
5.3	Discussions.....	54
6.	Conclusion and Future Work	57
6.1	Summary and Conclusions	57
6.2	Future Research Directions and Work.....	59
	Bibliography	60

Figure 1.1: Diagram of the motion tracking process	5
Figure 2.1: “Objects” identified in a scene [24].....	8
Figure 2.2 Vicon motion capture suit showing markers [43]	12
Figure 2.3: OptiTrack™ motion capture cameras Flex 3 and Flex 13.....	13
Figure 2.4: OptiTrack™ camera system (Marker based)	14
Figure 2.5: Kinect for Windows: Sensor Components	16
Figure 3.1: Kalman Filter as a predictor-corrector algorithm	21
Figure 3.2 Extended Kalman Filter: A complete picture of the operation [17].....	25
Figure 4.1. 3d description of a single link.....	32
Figure 4.2. Description of human upper limb	33
Figure 4.3. Skeletal model showing different points tracked on the body	34
Figure 5.1. Kinect tracking and OptiTrack™ systems at Cooperative Distributed Systems Laboratory at University at Cincinnati (currently at University of Toledo).....	47
Figure 5.2. Position estimates of elbow x-coordinates.....	48
Figure 5.3. Position estimates of elbow y-coordinates.....	49
Figure 5.4. Position estimates of elbow z-coordinates.....	49
Figure 5.5. Position estimates of wrist x-coordinates.....	50
Figure 5.6. Position estimates of wrist y-coordinates.....	50
Figure 5.7. Position estimates of wrist z-coordinates.....	51
Figure 5.8 Comparison of OptiTrack™ data and the proposed Kalman Filter based algorithm	52
Figure 5.9 : Sample data obtained after filtering.....	53
Figure 5.10: Sample output from the decision tree classification: Comparing the labeled Class and the output from Classification Tree	54
Figure 5.11: A typical classification tree graph for sit and walk action recognition	55

Chapter 1

1. Introduction

Human motion tracking has been one of the most challenging and most researched problems in the field of computer vision. The applications of such a system are wide-ranging from surveillance systems, animation for games, sports science and biomechanics, to motion tracking of the elderly. The motion capture systems can be broadly classified into marker-based systems and marker-less systems. While marker-based optical capture systems still remain the state-of-the-art, the challenge remains to analyze the data received from a marker-less vision based sensor system, to meaningfully interpret the data and to estimate in real-time the configuration that best describes the motion of the geometry that is being tracked.

The human motion tracking systems have involved a variety of models that use different image processing techniques and solution approaches. Surveys on existing methods for human motion tracking especially in the medical field can be found in [4] and [5]. The methods used for visual tracking as mentioned above, may vary from wearable marker based systems such as OptiTrack™, Vicon systems to non-invasive

marker-less sensors such as Microsoft Kinect and Primesense Carmine. The marker free visual systems essentially rely on the visual data to extract useful information through 2-D or 3-D shape model analysis to predict the motion of the body under surveillance such as described in [4] and [5]. The latest advances in algorithms that are used in computer vision have triggered the introduction of marker-less motion capture systems such as Microsoft Kinect [7] that can predict human motion with higher levels of accuracy. The marker-less motion tracking systems have a high potential in applications such as fall detection system in a smart home, where it is essential to have passive sensors that do not affect the daily activities of the inhabitants. With the availability of portable sensors that can be used in home-friendly environments, such systems will have crucial role to play in improving the quality of life in nursing and health care. Fall prevention and detection is one such promising application of this technology.

1.1 Motivation

In 2010, falls among older adults cost the U.S. health care system \$30 billion in direct medical costs and is a leading cause of injury and death in individuals over the age of 65 (Centers for Disease Control and Prevention, NCIPC)[34]. There is a lack of technology that implements fall prevention or detection without wearable sensors. Hence an effective portable and cost effective sensor that can track human motion and thereby help us

identify, detect and predict different activities within a non-invasive home environment becomes an immediate need of the hour given that the population of the elderly is steadily increasing with lack of adequate care-givers. The motivation for this work arises from the observation that commercially available sensors such as Microsoft Kinect are affordably priced, portable, non-invasive while at the same time have reliable tracking mechanism built into itself using computer vision based algorithms. However, the computer vision based techniques carry a substantial degree of uncertainty caused by factors such as occlusion, loss of marker data, or improper analysis of images obtained from the camera. This area has been under research for over thirty years and still the need remains to develop an effective tracking system. The fact that human body can be considered as a kinematic system of linkages with fixed length promises to add a new dimension to the computer vision based algorithms and at the same time help in better prediction of motion patterns.

1.2 Study Goals and Methodology

The approach discussed in this research is a precursor to finding an effective approach to predict human motion dynamics using non-invasive and non-wearable vision based sensor systems that can be applied in areas that require personalized care, e.g., a home care system for the elderly, where the focus is to provide a care-giving experience without

the use of sensors on the body of the person involved. The approaches proposed in the past for marker-less motion tracking involve use of different features of a continuous image data. The images from consecutive frames are analyzed as silhouettes [6], or by using blob detection algorithm [53] among other techniques. These techniques do not essentially incorporate the kinematics of human body in these studies. The goal of this research is to formulate this problem as a study of kinematic linkages in conjunction with computer vision based algorithms to achieve higher accuracy in human motion tracking.

This research focuses on use of frame to frame motion estimation as mentioned in [2] using non-wearable vision system such as Microsoft Kinect. An articulated motion algorithm that iteratively solves a linearly constrained least square system is used on the data collected from Microsoft Kinect. To further reduce the error introduced by occlusions, drift and environmental variations we introduce an additional layer of filtering based on Extended Kalman Filter[17]. This scheme exploits the fact that human body is a non-rigid system of rigid linkages connected through ball and hinge joints. The current state-of-the-art focuses on use of wearable sensors such as inertial or magnetic sensors ([15], [20]-[21]). Our approach incorporates the best of both worlds by fusing the two approaches thus achieving a non-wearable and marker-less model for human motion capture and tracking.

The final part of the thesis focuses on developing decision tree algorithm that can accurately predict between different activities. In this research we investigate two

activities namely, random walk and sit position which can be extended to the bigger problem of detecting different daily activities in humans.

1.3 Research Objectives

As we mentioned in the previous sections the goal of this research is to successfully fuse the computer vision based algorithms and the kinematics of human motion.

FIGURE 1.1: Diagram of the motion tracking process

More specifically the objectives of this research are,

Specific Aim 1: Collect raw image data from Kinect and apply Optical Flow based filtering of points to localize the configuration points on human skeletal model and develop a sustained human motion tracking model.

Specific Aim 2: To use the skeletal model obtained from Specific Aim 1 to apply kinematics based constraints to obtain an accurate skeletal tracking algorithm.

Specific Aim 3: To classify between two actions – “random walk” and “sit” based on the data collected from the preceding two steps.

1.4 Thesis Outline

Chapter 2 discusses the background study on current methods used in human motion tracking, a discussion on different sensors used for tracking and a study and a discussion on merits and demerits of each of the methods. Chapter 3 elucidates Kalman Filtering scheme and the optical method implemented in this research. The chapter also discusses the decision tree algorithms used in this research for classification of activities. Chapter 4 presents the proposed model based on optical flow and Extended Kalman Filter. Chapter 5 contains simulation results from our proposed model. Chapter 6 concludes with general inferences from this work, as well as discusses future research directions based on this work.

Chapter 2

2. Background

Human motion analysis is an engrossing and highly interdisciplinary science that has been explored ever since the early centuries of human mankind. The first scientific book on analysis of human motion can be attributed to Aristotle's book 'De Motu Animalium' – On the Movement of Animals [24]. He considers animal bodies as mechanical systems and investigates questions such as involuntary and voluntary movements in animals, physiological features and symmetry in animal motion.

More than two thousand years after Aristotle's book was written, study of human kinematics and study of human motion is still a challenging and continuously evolving area of research. Contributions of pioneers in science of mechanics, statistics, digital signal processing like Isaac Newton, Jacob and Daniel Bernoulli, Leonhard Euler, Thomas Bayes, Johan Carl Friedrich Gauss, Rudolf E Kalman and others of equal fame are highly noteworthy. Modern inquest into computer vision can be traced back to 1966 when Marvin Minsky at MIT hired a first-year undergraduate student and assigned him a project for the summer to connect a television camera to a computer, and have the machine describe what it sees. Guzman in [25] describes a program that analyzes a three

dimensional scenes that is presented in the form of line drawings. If a scene has a set of surfaces, the program partitions the scene into appropriate subsets of three dimensional bodies or “objects”.

FIGURE 2.1: “Objects” identified in a scene [24]

In the 1970s and 1980s, computer scientists focused their research on interpretation of still images and geometry. A real-time algorithm for facial recognition using both, physiology and information technology, was developed by Turk and Pentland [26] using statistical analysis. The trend of using statistical analysis has been in vogue ever since. With recent introduction of sensors and camera systems, powerful computational capabilities in digital technologies and large annotated datasets, a more comprehensive

video processing of real time data has been made viable in computer vision applications. Today vision sensors are used in varied applications such as mobile robotics, medical imaging, motion capture, vision based interactions, and gaming industry among others. Keeping in tune with the focus of this research, we will limit our discussion to vision based motion tracking and classification of predetermined actions performed during the motion tracking.

2.1 Existing Methods in Human Motion Tracking

Motion detection and tracking using sensors have been studied in literature for more than three decades and there still persists the challenge to find an effective tracking methodology. We may broadly classify the motion tracking sensors into wearable and non-wearable sensors. Vision based sensors such as Microsoft Kinect and Primesense Carmine are non-wearable vision sensors while OptiTrack™ system is a wearable motion tracking system. Most tracking algorithms use magnetic, inertial or other sensors in conjunction with the vision based sensors to develop a robust model.

Interest in exploiting human motion kinematics in image processing applications has been prevalent ever since the work of Korein [11]. Korein in his work analyzed the human arm and discussed the restriction on the elbow position based on the geometric analysis of human arm joints. Tolani et al. [12] discussed an analytical and numerical approach

that takes into consideration the inverse kinematics problems such as the position and orientation, among other parameters in a human limb. The modern approaches that implement *procrustes analysis* have evolved from approach discussed by Horn [13]. More recently the work presented by Aristidou and Lasenby, [14] gives a near optimal solution for marker based tracking system. They have used the Variable Turn Model within the framework of an unscented Kalman Filter. This work successfully takes into consideration the under-determined positions even from a single camera (given that the camera system usually needs data from a minimum of two sets of cameras to accurately describe a position), marker data, and Centre of Rotation and remains a bench-mark for multi-camera marker based detection systems.

The Microsoft research team in [7] uses a single depth image to accurately estimate the locations of body joints without temporal information. They use a simple per-pixel classification for pose estimation. Forsyth [8] and Bergen [9] present estimation algorithms that form the fundamental base of image processing and computer vision algorithms such as the study of affine transformation. In the algorithm presented by Sheikh et al. in [2] and [3], motion tracking is achieved by applying optical flow equations in consecutive frames of images which is combined with the per-frame body detection algorithm. The per-frame body detection is essentially a shape analysis algorithm. Hence essentially the algorithm consists of a Kalman Filter scheme with per-frame detection, frame-to-frame motion estimation using affine transformations followed by a covariance

propagation and update. This paper forms the basis of first part of the research presented in this research.

H. Martins Schepers et al., in [20], present a fusion of magnetic and inertial sensing for ambulatory human motion tracking. The fusion filter based on EKF from magnetic sensor is applied only when change in position and orientation from inertial sensor exceeds a certain threshold. Nadia B. F. et al. [33], discuss an algorithm that takes into consideration center of rotation and axis of rotation for position estimation. The VICON Motion Capture System is used for initial identification of parameters. The HumanEva [28] database presents Synchronized video and motion capture dataset and baseline algorithm that can be utilized to evaluate the articulated motion in humans. Jurgen Gall et al. [30] propose a multilayer approach in human motion capture that combines filtering, stochastic optimization and local optimization on HumanEva-II dataset. Zhou and Hu, in [15], discuss a motion tracking framework that fuses data from the on-board accelerometers and gyroscopes that is placed on the wrist position of the body. This data is used as measurement data for the extended Kalman Filter and is compared with an inverse kinematics model of the human upper limb. Hence both inertial and vision sensor based data are incorporated to achieve the solution to the problem of tracking human body upper limb.

2.2 Different Sensors used for Motion Tracking

In this section, we shall review general state-of-the-art approaches on motion tracking and motion capture with a focus on human motion tracking as an application. The sensors used in human motion tracking algorithms can be either based on image processing or sensors placed on the body or a combination of both. Hence, the motion capture systems can be broadly classified as wearable and non-wearable sensors [32].

a) Wearable sensors

Wearable sensors are placed on various parts of the human body for motion tracking. Main application of wearable sensors is in biomechanical analysis. Inertial sensors, accelerometers, gyroscopes, force sensors, electromyography are some of the few commonly used wearable sensors. These sensors have to be worn on the body and so do not offer an ideal solution to in-home or smart home based patient care.

Figure 2.2 Vicon motion capture suit showing markers [43]¹

¹ Reproduced with permission from ©Vicon Motion Systems Ltd. UK

Marker based optical tracking systems consist of one or more optical sensors. In general, optical sensing systems have configurations with a wide range of variety and customizability based on the application. Vision sensors are dependent on the light source. The source of light may be one of the following:

- Structured light that may project known pattern pixels onto a target. A common application is the structured light 3D scanner, where we can calculate the surface information and the depth of the scene.
- Directed light, which is used along with retro-reflective markers that return light to its source with minimum scattering and high level of accuracy.
- Ambient light
- Active lighting, the one emitted by the markers or the targets themselves.

FIGURE 2.3: OptiTrack™ motion capture cameras Flex 3 and Flex 13²

² Reproduced with permission from © 2014 NaturalPoint, Inc.

The type of markers used in marker based system would vary with the type of light used for tracking purpose. For validation purposes in this research, we use the OptiTrack™ available at the Cooperative Distributed Systems Laboratory. Vicron Motion Capture system and BTS systems are other examples of marker based motion tracking systems. While these motion capture systems are highly accurate, they may not a cost effective solution for applications such as smart home monitoring. Also the individuals living in these environments may prefer non-wearable sensors over marker based tracking.

3

FIGURE 2.4: OptiTrack™ camera system (Marker based)

³ Reproduced with permission from © 2014 NaturalPoint, Inc. OptiTrack™

b) Non-wearable sensors

Non-wearable sensors are primarily used in controlled environments such as laboratories or smart homes where data acquisition and data retrieval can be made possible from camera, audio, laser or pressure sensors. Recent developments in depth measurements have enabled systems such as Microsoft Kinect [7] to be actively used in research with reasonable accuracy in 3-d tracking applications. Vision sensors are most widely used sensors in human motion tracking.

One of the common applications of these vision sensors is in Surveillance and Monitoring, which is a continuous monitoring provided by a distributed system of hundreds of security cameras; Smart homes and hospitals, for patients in hospitals, older adults in nursing homes or loved ones in home environment who need constant care; Telepresence systems, for remote access and remote control such as in Tele-health systems. The emphasis in all the mentioned applications is on use of smart algorithms that provide information about the person needing care, for example in situation that result in an unusual behavior in a patient or an emergency situation such as fall in among the older adults.

The choice of Kinect in this research can be primarily attributed to that fact that the sensor is inexpensive; portable while at the same time has high levels of accuracy with its in-built motion tracking capabilities, when there are no occlusions. Recent studies

have validated use of Kinect sensors in biomechanical studies such as gait analysis [36]-[37] and render credibility to the decision to pick this sensor to test our algorithm.

FIGURE 2.5: Kinect for Windows: Sensor Components

Different sensors that the Kinect has are shown in Figure 2.5 shows. Kinect has an RGB camera and the color sensor with a resolution of 1280x960 can capture and store three channel data. The depth information is obtained using the infrared emitter and infrared depth sensor [35]. The depth map is constructed using a speckle pattern of infrared laser light, a technology that Microsoft licensed from Primesense. The other features such as audio sensors and accelerometer are mentioned in detail in Table 1

TABLE 1: Kinect: Specifications [35]

Kinect	Array Specifications
Viewing angle	43° vertical by 57° horizontal field of view
Vertical tilt range	$\pm 27^\circ$
Frame rate (depth and color stream)	30 frames per second (FPS)
Audio format	16-kHz, 24-bit mono pulse code modulation (PCM)
Audio input characteristics	A four-microphone array with 24-bit analog-to-digital converter (ADC) and Kinect-resident signal processing including acoustic echo cancellation and noise suppression
Accelerometer characteristics	A 2G/4G/8G accelerometer configured for the 2G range, with a 1° accuracy upper limit.

2.3 Human Motion Kinematics and Activity Recognition

Kinematic parameterization is an active area of research in human motion tracking. Human motion is an articulated system of rigid linkages and overall motion of the human body can be classified as a complex combination of angular and linear motions. In this research we implement an inverse kinematics based model which is described in detail in Chapter 4. Human activity recognition is closely associated with the Kinematic data that is obtained through motion tracking.

Data mining is one of the important and significant applications of Machine Learning. Most of the algorithms used in data mining can be broadly classified into supervised and unsupervised learning. The algorithms used for recognition of activities can be generally classified into the following categories [45]:

- a) Dynamic Time Warping (DTW): In Dynamic Time warping temporal sequences are analyzed for similarity. In [46], Sampena et al. present a DTW approach to recognize actions such as punching, waving, clapping, etc. This algorithm is fast and convenient for real time data analysis but it tends to become computationally intensive when the temporal templates that need to be matched are of larger size. Hence the approach is suitable for templates with limited attributes.
- b) Generative Models: Statistical Learning algorithms such as Hidden Markov Models, Dynamic Bayesian Networks that learn joint probability distribution over observation and label sequences. Generative models are completely probabilistic models.
- c) Discriminative models: In cases where the overall goal is data categorization, discriminative models are used. The Support Vector Machines (SVMs), Artificial Neural Networks (ANNs) are commonly used discriminative models
- d) Logic based algorithms: Decision trees and rule based classifiers can be classified as logic based algorithms. In this work we apply a decision tree classifier as the

kinematic attributes were inter related and a decision tree would take into consideration multiple factors to yield a result.

The various algorithms used in this research to accomplish the specific aims mention in Chapter 1 are described in detail in the two chapters that follow.

Chapter 3

3. Algorithms used in Motion Tracking

This chapter provides a description of algorithms that have been used for motion tracking in this research. The concepts of Kalman Filter and Optical Flow have been actively applied in this research. The Kalman Filter is one of the most important data fusion algorithms in use today even though it is over fifty years old. Ever since it's most famous application in navigation computer for the Apollo mission that took the first man to the moon, the algorithm is in use extensively in almost every aspects of digital computing including communication systems, navigation systems and even computer games. This chapter justifies the use of Kalman Filter based algorithms for this research.

Kalman Filter can estimate the general state for a discrete-time process which is driven by a linear stochastic difference equation. It may be considered as a feedback control mechanism for estimation, i.e., the filter initially estimates the state of the process at a specific time instance and then the measurements, which may be noisy, act as the feedback loop. Essentially, the equations involved in Kalman filter can be classified into two steps. The two steps involved in Kalman Filter are:

a) Time update or prediction equations: The *a priori* estimates of current state and error covariance are obtained from the previous time step without any observation input from current time-step.

b) Measurement update or correction equations: The objective of measurement update equations is to attain an improved *a posteriori* estimate by incorporating a new measurement into the *a priori* estimate.

The two sets of equations are called predictor and corrector equations, respectively, for time update and measurement update. Hence the algorithm functions as a cycle of predictor-corrector equations for solving numerical problems.

Figure 3.1: Kalman Filter as a predictor-corrector algorithm

The basic Kalman filter is confined to linear stochastic difference equations. More complex systems such as the articulated motion of human body considered in this research, satellite navigation systems, brain-computer interface, inertial guidance systems, computer vision applications, etc. involve non-linear equations and the following methods are most widely implemented in such cases:

a) Extended Kalman Filter:

In Extended Kalman Filter (EKF), the state transition model and the observation model may be non-linear functions and not necessarily linear. They are different from Kalman Filter as KF deals with linear equations. Linearizing dynamics in Extended Kalman Filter is based on the current estimate of mean and covariance.

b) Unscented Kalman filter

For highly non-linear state transition and observation models—that is, the extended Kalman filter may give inconsistent performance. The reason for inconsistency may be that the covariance is propagated through linearization of the underlying non-linear model. On the contrary the unscented Kalman filter (UKF) captures the true mean and covariance in a more precise manner. The method used in UKF is a deterministic sampling technique [54] known as the unscented transform. This method picks a minimal set sigma points around the mean. These are then propagated through the non-linear

functions, so that the mean and the covariance of the estimate are retrieved from them [22].

Other variants of Kalman Filter such as Discrete-time Kalman Filter, Kalman-Bucy Filter and Hybrid Kalman Filter are used in literature for specific applications. For the purpose of explaining the algorithm that has been used in this research we shall limit our discussion to Extended Kalman Filter.

3.1 Extended Kalman Filter Algorithm

G. Welch and G. Bishop in [17] present a detailed description of Kalman Filter and Extended Kalman Filter (EKF). This section will describe some of the key features that would help understand the working of EKF. While Kalman Filter is applied to linear systems, EKF is the non-linear version of Kalman Filter. Let us consider that a process has state vector, $x \in \mathbb{R}^n$ and that, at the current time step, it is governed by a non-linear differential equation for the state,

$$x_k = f(x_{k-1}, u_{k-1}, v_{k-1}) \quad (3.1)$$

With a measurement $z \in \mathbb{R}^m$,

$$z_k = h(x_k, v_k) \quad (3.2)$$

where w_k is the random variable that represents process noise and v_k is the random variable that represents measurement noise at the current time step k and u_k is the control input. The process and measurement noise are assumed to be white, with normal probability distributions and independent of each other with process noise covariance matrix, Q and measurement noise covariance matrix, R , which may change at each time step.

$$p(w) \approx N(0, Q) \tag{3.3}$$

$$p(v) \approx N(0, R) \tag{3.4}$$

If $\hat{x}_k$ is the *a posteriori* estimate of the state vector, and after the state and measurement vector approximation we obtain the *a priori* estimate, $\hat{x}_k^-$ and the EKF time and measurement update equations can be summarized in the following figure.

Figure 3.2 Extended Kalman Filter: A complete picture of the operation [17]

where,

P_k^- is the *a priori* estimate error covariance at time step k

P_k is the *a posteriori* estimate error covariance at time step k

A, W, H and V are the Jacobian matrices of partial derivatives of f with respect to x, w, x, v , respectively, at each time step, that is

$$A_{[i,j]} = \frac{\partial f_{[i]}}{\partial x_{[j]}}(\hat{x}_{k-1}, u_{k-1}, 0),$$

$$W_{[i,j]} = \frac{\partial f_{[i]}}{\partial w_{[j]}}(\hat{x}_{k-1}, u_{k-1}, 0),$$

$$H_{[i,j]} = \frac{\partial h_{[i]}}{\partial x_{[j]}}(\hat{x}_k^-, 0),$$

$$V_{[i,j]} = \frac{\partial h_{[i]}}{\partial v_{[j]}}(\hat{x}_k^-, 0),$$

K_k is the Kalman gain at time step k

The Extended Kalman filter is the *de facto* standard in Global Positioning Systems and navigation systems [52] and the reason why we chose for the purpose of this research. However, if the modelling of the process is inaccurate, and if the initial estimate of the state is erroneous the filter tends to diverge.

3.2 Optical Flow Equations

In visual experience, motion is an innate property and it is a rich source of data that helps us understand our world through scene analysis, object recognition and 3D shape acquisition [39]. The objective of optical flow equations is to understand and study this spatio-temporal image intensity data to estimate the motion between two image frames. These equations are differential equations and can be typically termed as the local Taylor series approximations of spatio-temporal image coordinates. For an image with spatio-temporal coordinates (x, y, t) with intensity $I(x, y, t)$ that moves by $(\Delta x, \Delta y, \Delta t)$ between

two image frames that follows the brightness consistency constraint can be assumed to conserve intensity between consecutive frames [39]:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3.5)$$

Taylor series approximation of the equation gives:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \text{H. O. T.} \quad (3.6)$$

Further solving the above equation we can finally arrive at the gradient constraint equation of the form:

$$\frac{\partial I}{\partial x} \frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y} \frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t} \frac{\Delta t}{\Delta t} = 0 \quad (3.7)$$

The above equation in two unknowns can be resolved using different algorithms and additional constraints based on the application. In this research we have used differential method of estimating optical by using partial derivatives of image coordinates. Of all differential methods used for optical flow estimation, Lucas-Kanade method [38] is the most widely used. The usual assumption of a pure translation (constant flow) motion model for pixels in a large window such as the one obtained from Kinect may be unreasonable. In this research, Lucas-Kanade approach is generalized to affine motion model by introducing a warp function ([2], [8], [42]).

3.3 Decision Trees

In machine learning, the attributes in various instances in a dataset can be identified based on a set of features. The unique properties that these features represent for each of the classes can be used to label the outputs. Such a learning algorithm is called supervised learning. The decision tree classifiers use tree like models or structures to map observations and outcomes to the different classes. The decision tree classifiers are most researched classifiers in machine learning. Decision tree classifiers have been actively used in activity recognition research [46][47]. Our choice of this classifier is primarily because of its applicability and its inter-usability with unsupervised learning algorithms such as Hidden Markov Models (HMMs) to build a predictive model. The advantages of decision tree algorithms are its simplicity to understand and interpret, logical and intuitive evaluation of different possible scenarios and the flexibility to combine with other decision techniques. However, the limitations of decision trees include bad predictions if the data is split too many times to obtain classes, which is called over-fitting or if the data is skewed, the decision module tends to favor one class over the other. The common algorithms used for decision tree learning are Iterative Dichotomiser 3 (ID 3), C4.5 and Classification and Regression Trees (CART).

In this work, we have used the C4.5 algorithm which is an extension to the Iterative Dichotomiser 3 algorithm developed by Ross Quinlan [51]. Since the data that we acquired during our experiments involves predefined activities, this algorithm

provides a good basis for activity recognition. The measures applied to derive the best split is based on degree of impurity in the child nodes. Information entropy is usually the measure

$$Entropy = - \sum_{i=0}^{c-1} p(x_i) \log_2 p(x_i) \quad (3.8)$$

Where $p(x_i)$ is the fraction of instances belonging to class x and c is the number of classes.

The pseudocode for decision tree algorithms can be summarized into following steps [47]:

- › Check for base cases
- › Find the information gain for every attribute
- › Next denote a_{best} , which will be the attribute that has the highest normalized information gain
- › A decision *node* is formulated that splits on a_{best}
- › Recursively split on a_{best} , and add those nodes as child *nodes*

The splitting criteria to determine the best attribute to split the database in C4.5 algorithm is obtained using Information gain, which is the difference of the global

entropy and the local entropy at the child node. Some of the algorithms also use Gain ratio as splitting criteria.

$$\Delta = Entropy_{Global} - Entropy_{Local} \quad (3.9)$$

The C4.5 algorithm can handle attributes that are missing, continuous or discrete and also prunes the decision tree to counter over-fitting and increase the accuracy of prediction.

CHAPTER 4

4. Problem Formulation and Approach

In this chapter we shall discuss the algorithm used in this research in detail. The Kinematic model considered to describe human body is based on inverse kinematics obtained by describing the different joints and configuration points on human body in terms of Euler angles. First of all the input stream of data in image coordinates is filtered using optical flow based method. The filtered data describing human skeletal system is then considered as a measurement input for the kinematic model and the position data of different joints is filtered to reduce the error that may have been introduced from image data. Our approach is described in detail in the following sections.

4.1 The Kinematic Model

In this research, a two-joint upper limb system is initially taken into consideration and then the approach is extended to more joints on the body. The kinematic chain described in Figure 4.2. , consists of shoulder and elbow joints and the analysis includes the wrist, elbow, and shoulder positions. The inverse kinematics of human arm is derived by reversing the Denavit-Hartenberg convention.

Figure 4.1. 3d description of a single link

Let the world position co-ordinates of the shoulder, elbow and wrist be denoted by $X_s(x_s, y_s, z_s)$, $X_e(x_e, y_e, z_e)$ and $X_w(x_w, y_w, z_w)$ and the lengths of the upper and lower limbs be L_1 and L_2 , respectively as shown in Figure 4.2. These positions are obtained using the optical flow method as described in the next section. The optical flow method is used on the raw images obtained from Microsoft Kinect to estimate the corrected 2D locations of the skeletal points on the body. . It may be noted that Kinect also has infra-red sensor that provides the depth map corresponding to its visual image.

Figure 4.2. Description of human upper limb

Since the positional data is obtained in three dimensions, the problem can be divided into projected coordinates in three orthogonal planes. For this analysis, it is assumed that shoulder co-ordinates are known. Since the measurement data is in the form of position co-ordinates, it may be converted into the corresponding angular data by using simple trigonometric properties. This will yield us angular measurement along the three orthogonal planes.

The Euler angles of the shoulder position along x, y, and z directions can be expressed as (as shown in Figure 4.1):

$$\theta_x^s = \tan^{-1}(z_s - z_e)/(y_s - y_e) \quad (4.1)$$

$$\theta_y^s = \tan^{-1}(z_s - z_e)/(x_s - x_e) \quad (4.2)$$

$$\theta_z^s = \tan^{-1}(y_s - y_e)/(x_s - x_e) \quad (4.3)$$

Similarly, the Euler angles between the wrist and the elbow positions $(\theta_x^e, \theta_y^e, \theta_z^e)$ are determined. The Euler angles thus obtained are used in an Extended Kalman Filter to estimate the positions of the wrist and elbow.

Figure 4.3. Skeletal model showing different points tracked on the body

4.2 Kalman Filter for Optical Flow Constrained Based Estimation

The optical flow based approach used is inspired by the algorithm described by Sheikh et al. [2]. The algorithm used in our analysis builds a probabilistic framework around the problem. The aim of the algorithm is to develop an accurate representation of the state of the system, given the information about the current state of the system obtained using sensors such as the Kinect. The probabilistic method used in this research is based on Kalman filtering applied to spatio-temporal image data obtained from successive frames of gray-scale images at successive time instances.

4.2.1 The Kalman Estimation

Given the initial state of the system, i.e., the image co-ordinates of the selected image features or joints on the human body, X_t the estimation of the final configuration of the system can be given as:

$$X_t^+ = X_t^- + \sigma_t^- (\sigma_t^- + \overline{\sigma}_t)^{-1} (\overline{X}_t - X_t^-), \quad (4.4)$$

where, (X_t^-, σ_t^-) are the predicted position and the error covariance matrix of the prediction. $\overline{X}_t, \overline{\sigma}_t$ are the input measurement data and the noise covariance matrix associated with the measurement.

The corrected covariance is given by,

$$\sigma_t^+ = (I - \sigma_t^-(\sigma_t^- + \overline{\sigma}_t)^{-1})\sigma_t^-, \quad (4.5)$$

4.2.2 The measurement data and equations

The measurement data is collected from the Kinect at predetermined points on the body from a fixed frame in time to estimate $(\overline{X}_t, \overline{\sigma}_t)$ for that image. To estimate the variance in the measurement data we use a mean compensated singular decomposition of the input vector at the time instance t , X_t using Matlab which is of the form,

$$X_t = WDV^T \quad (4.6)$$

where W and V are unitary rotation matrices and D is the scaling matrix.

The covariance of the detection is then computed as,

$$\overline{\sigma}_t = VV^T \quad (4.7)$$

4.2.3 The prediction and correction equation

The motion induced between two views of a plane can be expressed in terms of affine transformation A_t , such that the corrected equation of configuration of the system can be represented as,

$$X_t' = A_t X_t \quad (4.8)$$

Where, the affine transformation matrix is given by,

$$A_t = \begin{bmatrix} a1 & a2 & a3 \\ a4 & a5 & a6 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.9)$$

Since $X_t \in R^2$, a nine point system of motion (corresponding to nine points being tracked) would give a set of eighteen equations. To estimate the affine transformation parameters we use a Squared Sum of Difference (SSD) minimization as proposed in [3] and [8]. Applying Gauss-Newtonian estimation for first order approximation yields an optical flow constraint equation,

$$\nabla I_x(x' - x) + \nabla I_y(y' - y) + \nabla I_t = 0 \quad (4.10)$$

where, ∇I_x and ∇I_y are spatial gradients and ∇I_t is the temporal gradient of the image and $X_t = [x \ y]$ and $X_t' = [x' \ y']$. Further simplification of Equations (4.8) - (4.10) would yield a linear least squares minimization problem of the form,

$$\tau_t(\nabla I_x, \nabla I_y, \nabla I_t)A_t = b(\nabla I_x, \nabla I_y, \nabla I_t) \quad (4.11)$$

where τ and b are a function of the spatio-temporal gradients, ∇I_x , ∇I_y and ∇I_t and the spatial configuration coordinates, X_t and X_t' and A_t is the affine transformation estimate at current time instance. This least square minimization problem can be solved using any

standard optimization tool, e.g., the 'lsqlin' in MATLAB. The solution would give us the affine transformation matrix, A_t . As mentioned in [9], the covariance matrix for Kalman correction for affine transformation is given by,

$$\sigma_t^- = J_x \sigma_x J_x^T + J_A \sigma_A J_A^T \quad (4.12)$$

where,

$$J_x = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \quad (4.13)$$

$$J_A = \begin{bmatrix} \frac{\partial u}{\partial a_1} & \cdots & \frac{\partial u}{\partial a_6} \\ \frac{\partial v}{\partial a_1} & \cdots & \frac{\partial v}{\partial a_6} \end{bmatrix} \quad (4.14)$$

where $u = (x' - x)$, $v = (y' - y)$, σ_x is the covariance at the previous time instance and σ_A is the inverse of information matrix, $\tau^T \tau$ (obtained in Equation 4.11),

$$\sigma_A = [(\tau^T \tau)^T (\tau^T \tau)] \quad (4.15)$$

The method described in this section is used to obtain the 2-D locations of the points of human limbs chosen for the tracking purpose. Once, the 2-D locations are obtained, the depth map obtained from the Kinect sensor is used to determine the 3rd dimension of the points. Once 3-D locations of these points are obtained, a tracking algorithm based on Extended Kalman Filter, as described below, is used to track these points as these points move in 3-D space. The major motivation for using this Extended Kalman Filter comes from the fact that the constraints provided by the rigid body motion of points (considering limbs to be linkages of rigid bodies) can be exploited to improve the tracking of these points in events of noise and occlusions.

4.3 Extended Kalman Filter for Determining the Position Coordinates

The Kalman Filter is used to estimate the states of a dynamic system described by a linear stochastic difference equation and is extensively used in tracking and estimation problems [16],[17]. An Extended Kalman filter finds application in the cases where the system process is non-linear such as the problem discussed in this research. Since the model discussed in this work deals with angular dimensions and involves non-linearity, we have implemented an Extended Kalman Filter.

4.3.1 The Kalman Estimation

The positions $X_s(x_s, y_s, z_s)$, $X_e(x_e, y_e, z_e)$ and $X_w(x_w, y_w, z_w)$ are obtained from the previous step and passed as measurement data. The Euler angles at the shoulder and elbow positions $(\theta_x^s, \theta_y^s, \theta_z^s, \theta_x^e, \theta_y^e, \theta_z^e)$ can be determined from the equations described in Equation (4.1) – (4.3). The angular velocity $(\dot{\theta}_x^s, \dot{\theta}_y^s, \dot{\theta}_z^s, \dot{\theta}_x^e, \dot{\theta}_y^e, \dot{\theta}_z^e)$ and the angular acceleration $(\ddot{\theta}_x^s, \ddot{\theta}_y^s, \ddot{\theta}_z^s, \ddot{\theta}_x^e, \ddot{\theta}_y^e, \ddot{\theta}_z^e)$ can be defined by the angular kinematics motion model for time-frame k . The state vector $x \in R^n$ describing the motion of the arm can be represented as described below:

$$x_k = [\theta_x^s, \dot{\theta}_x^s, \ddot{\theta}_x^s, \theta_y^s, \dot{\theta}_y^s, \ddot{\theta}_y^s, \theta_z^s, \dot{\theta}_z^s, \ddot{\theta}_z^s, \theta_x^e, \dot{\theta}_x^e, \ddot{\theta}_x^e, \theta_y^e, \dot{\theta}_y^e, \ddot{\theta}_y^e, \theta_z^e, \dot{\theta}_z^e, \ddot{\theta}_z^e]_k^T \quad (4.16)$$

The transformation along the projected planes is assumed to be a constant angular motion as the time period, dt is very small between consecutive time frames k and $k+1$. In this study the step time dt is chosen as 0.03 seconds. Inertial sensors have not been used in this analysis and hence the model works on pure kinematics. Let us define:

$$dF = \begin{bmatrix} 1 & dt & \frac{dt^2}{2} \\ 0 & 1 & dt \\ 0 & 0 & 1 \end{bmatrix} \quad (4.17)$$

and

$$F = \begin{pmatrix} dF & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & dF \end{pmatrix} \quad (4.18)$$

The non-linear stochastic difference equation governing the kinematic relationship of the angular quantities with state vector $x \in R^n$ at the current step $k + 1$ can be represented as:

$$x_{k+1} = Fx_k + w_k \quad (4.19)$$

where w_k is the process noise.

4.3.2 Measurement Data

The measurement equation is given by:

$$z_{k+1} = f(x_{k+1}) + v_k \quad (4.20)$$

where $f(x_{k+1})$ is a function of the state vector, x_{k+1} expressed in terms of the Euler angles $(\theta_x, \theta_y, \theta_z)$ for shoulder (denoted by s) and elbow (denoted by e), respectively and v_k is

the observation noise. The measurements for human upper limb system analyzed here would be the position coordinates of the elbow and wrist X_e and X_w . Hence, $z_{k+1} = [x_e, y_e, z_e, x_w, y_w, z_w]_{k+1}$

The relationship between the Euler angles and the coordinates is given in Equations (4.1-4.3).

The process and measurement noise w_k and v_k , respectively, can be assumed to be independent of each other, white and Gaussian by distribution. The equations for process and measurement noise covariance matrices, Q and R , respectively, can be given as,

$$p(w) \approx N(0, Q) \quad (4.21)$$

$$p(v) \approx N(0, R) \quad (4.22)$$

4.3.3 The time and measurement update

Let x_{k+1}^- be *a priori* estimate at the current step $k+1$ and x_{k+1} be *a posteriori* estimate.

The time-update equations are as described in the following equations:

$$x_{k+1}^- = F x_{k+1} \quad (4.23)$$

$$P_{k+1}^- = F P_k F^T + Q \quad (4.24)$$

The measurement-update equations can be represented as:

$$K_{k+1} = P_{k+1}^- H^T (H P_{k+1}^- H^T + R)^{-1} \quad (4.25)$$

$$x_{k+1} = x_{k+1}^- + K_{k+1} (z_{k+1} - f(x_{k+1})) \quad (4.26)$$

$$P_{k+1} = (I - K_{k+1} H) P_{k+1}^- \quad (4.27)$$

where, F is the transformation matrix and P_k and P_{k+1}^- are *a priori* and *a posteriori* estimate error covariance matrix respectively.

K_{k+1} is the Kalman gain function that helps minimize error by weighing the relative certainty of the measurements and current state estimate and H is the Jacobian function of the measurement function and is given by:

$$H = \frac{\partial f(x(\theta_{x,y,z}^p))}{\partial \theta_j^{p_i}} \quad (4.28)$$

where p_i is the i^{th} joint, $p = s, e$ and $j = x, y, z$

Once the results are obtained from the Extended Kalman filter for the current step, as described above, the angular measurements are converted back to their corresponding world coordinates. As mention in [15], the H matrix is calculated using the knowledge of the shoulder position and the elbow position as described below:

If, $\theta_y^s = \pm \frac{\pi}{2}$, then

$$x_e = x_s \quad (4.29)$$

Else if, $-\frac{\pi}{2} < \theta_y^s < \frac{\pi}{2}$, or $\frac{3\pi}{2} < \theta_y^s < 2\pi$,

$$x_e = x_s - L_1 \sqrt{\frac{\cos^2 \theta_z^s + \sin^2 \theta_x^s \sin^2 \theta_z^s}{\sin^2 \theta_x^s \sin^2 \theta_z^s + \frac{1}{\cos^2 \theta_y^s}}} \quad (4.30)$$

Else if, $\frac{\pi}{2} < \theta_y^s < \frac{3\pi}{2}$, or $-\frac{3\pi}{2} < \theta_y^s < -\frac{\pi}{2}$,

$$x_e = x_s - L_1 \sqrt{\frac{\cos^2 \theta_z^s + \sin^2 \theta_x^s \sin^2 \theta_z^s}{\sin^2 \theta_x^s \sin^2 \theta_z^s + \frac{1}{\cos^2 \theta_y^s}}} \quad (4.31)$$

Once the position estimates of x-coordinates are obtained, the y and z coordinates can be determined by:

$$y_e = y_s - \sqrt{L_1^2 - (x_e - x_s)^2} \cos \theta_x^s \quad (4.32)$$

$$z_e = z_s - \sqrt{L_1^2 - (x_e - x_s)^2} \sin \theta_y^s \quad (4.33)$$

Similarly after determining the positions of the elbow joint, the same approach can be used to determine the position of the wrist $X_w(x_w, y_w, z_w)$ with elbow position as the reference. This approach can be extended to the skeleton model obtained from Kinect by analyzing different combinations of joints by changing the reference joint each time.

Chapter 5

5. Simulation Results and Discussions

This chapter describes the performance of the proposed Kalman filter using angular kinematic model. The results obtained using Microsoft Kinect is verified against 18 camera OptiTrack™ motion capture systems which is highly accurate. In the following sections we shall discuss our experimental setup and results thus obtained.

5.1 Experimental Setup

Microsoft Kinect is used to obtain the measurements and the results are compared to the commercially available OptiTrack™ motion tracking system. The 24 camera OptiTrack™ system at Cooperative Distributed Systems Laboratory uses a 100 frames/sec (FPS) marker tracking JPEG on-camera processing with a 640×480 resolution. We utilized only 18 cameras as part of this study. The camera system is highly accurate with mean error of 2-3 mm. For the purpose of analysis, the motion of a human body is acted out with markers placed at the wrist, elbow and shoulder positions. The data is collected by the OptiTrack™ system at a sampling rate of 100Hz. The Microsoft Kinect supports 10-

30 FPS depending upon the resolution of the image. The data was collected real-time and the analysis of results was performed using MATLAB. The data visualization and analysis for activity classification was carried out using open source statistical software, Orange.

Figure 5.1. Kinect tracking and OptiTrack™ systems at Cooperative Distributed Systems Laboratory at University at Cincinnati (currently at University of Toledo)

5.2 Simulation Results

Analysis has been performed incorporating two different causes of error in vision sensor based systems. The first kind of error takes into consideration measurement data that has Gaussian white noise which is usually prevalent in most sensor systems. The

second type of error is one that involves random noise that is caused by occlusions and loss of marker data. The shoulder position is assumed to be the reference position for this analysis as the shoulder movement is fairly uniform and does not lose tracking in normal circumstances. In order to demonstrate the robustness of the proposed algorithm to larger amount of noise, noises are introduced into data obtained from Kinect to simulate both of these cases (Gaussian noise and occlusions modeled by spiking noise).

Figure 5.2. Position estimates of elbow x-coordinates

Figure 5.3. Position estimates of elbow y-coordinates

Figure 5.4. Position estimates of elbow z-coordinates

Figure 5.5. Position estimates of wrist x-coordinates

Figure 5.6. Position estimates of wrist y-coordinates

Figure 5.7. Position estimates of wrist z-coordinates

Figure 5.8 Comparison of OptiTrack™ data and the proposed Kalman Filter based algorithm

(Top row: X-coordinates of elbow (left) and wrist (right),
Second row: Y-coordinates of elbow (left) and wrist (right) and Bottom row: Z-coordinates of elbow (left) and wrist (right))

1	Px	Py	Pz	Vx	Vy	Vz	Ax	Ay	Az	Joint	Time	Class
2	412.52	216.99	2451.2	2744.9	574.04	1054	45063	39705	14659	1	126	Walk
3	483.56	5.3521	2481.1	1838.7	-834.4	1305	-5829	2462.1	-26687	2	126	Walk
4	347.9	4.041	2572.3	1511.1	-118.5	402.66	-28280	29650	-70797	3	126	Walk
5	619.22	6.6633	2389.8	2166.3	-1550	2207.3	16623	-24726	17423	6	126	Walk
6	449.9	-170.6	2428.5	928.47	-889.7	913.59	-15134	-8077	-12941	9	126	Walk
7	336.75	-347.2	2429.4	-401.4	-571.5	215.01	-39607	-3864	-20347	10	126	Walk
8	495.72	-345.7	2322.4	437.94	-1319	829.42	-9270	-33369	21955	13	126	Walk
9	455.19	215.41	2483.3	1280.2	-47.35	962.26	-43940	-18641	-2753	1	127	Walk
10	522.06	3.4689	2494.7	1154.9	-56.5	409.46	-20513	23338	-26865	2	127	Walk
11	416.48	3.8102	2610.2	2057.5	-6.924	1137.1	16392	3348.4	22034	3	127	Walk
12	627.64	3.1277	2379.2	252.39	-106.1	-318.2	-57418	43328	-75765	6	127	Walk
13	509.12	-182.1	2483.4	1776.6	-345.8	1648.6	25444	16318	22050	9	127	Walk
14	422.63	-367.4	2552.6	2576.3	-604.9	3697.7	89333	-1002	104482	10	127	Walk
15	569.73	-367.9	2391.7	2220.3	-665.3	2077.6	53470	19597	37447	13	127	Walk
16	491.65	224.32	2511.2	1093.8	267.32	837.93	-5592	9440	-3730	1	128	Walk
17	551.07	-0.884	2515.7	870.35	-130.6	630.19	-8537	-2223	6621.9	2	128	Walk
18	440.07	5.5427	2626.7	707.62	51.977	495.75	-40495	1767	-19241	3	128	Walk
19	662.07	-7.311	2404.7	1033.1	-313.2	764.63	23421	-6213	32485	6	128	Walk
20	530.65	-188.5	2506.2	645.86	-192.6	681.85	-33923	4595.1	-29002	9	128	Walk
21	441.94	-372.2	2564.9	579.22	-143.2	368.62	-59913	13852	-99874	10	128	Walk
22	578.51	-380.1	2428.3	263.53	-366.2	1098.4	-58702	8973.9	-29378	13	128	Walk
23	538.5	220.16	2543.8	1405.4	-125	977.32	9346.3	-11771	4181.6	1	129	Sit
24	582	-6.636	2543.8	927.99	-172.5	841.96	1729.1	-1259	6353	2	129	Sit
25	472.24	2.3525	2654.5	965.03	-95.71	832.37	7722.4	-4431	10098	3	129	Sit
26	691.77	-15.62	2433.1	890.95	-249.4	851.54	-4264	1913.4	2607.5	6	129	Sit
27	550.17	-195.3	2527.5	585.69	-203.6	641.27	-1805	-330	-1217	9	129	Sit
28	450.31	-378.4	2579.9	251.31	-186.2	449.67	-9837	-1291	2431.5	10	129	Sit
29	586.36	-389.5	2442.7	235.46	-283.2	431.51	-842.1	2488.4	-20007	13	129	Sit
30	575.2	209.84	2564.3	1101	-309.5	613.61	-9131	-5534	-10911	1	130	Sit
31	606.56	-15.69	2567	736.59	-271.7	695.58	-5742	-2974	-4391	2	130	Sit
32	489.95	-5.126	2673.2	531.5	-224.4	560.46	-13006	-3859	-8157	3	130	Sit
33	723.16	-26.26	2460.8	941.67	-319	830.71	1521.8	-2088	-625.1	6	130	Sit
34	564.85	-199.1	2539.4	440.36	-112.8	356.38	-4360	2725.1	-8547	9	130	Sit
35	453.34	-376.1	2575.4	90.901	68.715	-133.8	-4812	7648.4	-17505	10	130	Sit
36	592.94	-388.7	2448.3	197.36	23.451	168.19	-1143	9199.7	-7900	13	130	Sit
37	594.6	209.72	2573.1	582	-3.517	263.91	-15570	9178.2	-10491	1	131	Sit

Figure 5.9 : Sample of training data obtained after filtering

Where x, y, z denote the axes on the coordinate system that places Kinect at the origin and $(P_x, P_y, P_z), (V_x, V_y, V_z)$ and (A_x, A_y, A_z) denote position, velocity and acceleration, respectively along these axes.

	Px	Py	Pz	Vx	Vy	Vz	Ax	Ay	Az	Joint	Time	Class	Classification Tree
1240	21.8543...	641.2512...	2260.89...	666.654846...	35.482177...	123.2226...	3708.6926...	1427.34375...	3347.5341...	1	138	Sit	Sit
1241	75.0152...	372.9462...	2310.46...	432.645721...	58.765869...	101.4624...	-2860.280...	-3519.9645...	1758.0322...	2	138	Sit	Sit
1242	-101.07...	354.1102...	2329.08...	356.834564...	105.03295...	207.9418...	-5823.818...	-7016.9677...	3271.0693...	3	138	Sit	Sit
1243	251.106...	391.7823...	2291.85...	508.456878...	12.498779...	-5.01708...	103.25775...	-22.988891...	244.99511...	6	138	Sit	Sit
1244	99.7226...	140.4472...	2308.92...	363.516754...	1.0180664...	76.14990...	685.65673...	-1078.4042...	-629.5166...	9	138	Sit	Sit
1245	21.7072...	-103.039...	2318.25...	139.802337...	-42.69790...	122.5341...	3656.0507...	-303.35311...	-2148.046...	10	138	Sit	Walk
1246	227.152...	-81.0637...	2296.53...	448.972778...	-70.76156...	-20.8593...	4807.1230...	3029.68603...	-3886.083...	13	138	Sit	Walk
1247	44.4323...	635.8073...	2260.53...	677.339843...	-163.3172...	-10.7446...	320.55014...	-5963.9829...	-4019.018...	1	139	Sit	Sit
1248	100.089...	368.7684...	2310.35...	752.227050...	-125.3338...	-3.36181...	9587.4394...	-5522.9916...	-3144.726...	2	139	Sit	Sit
1249	-68.557...	352.6333...	2332.97...	975.547485...	-44.30511...	116.6015...	18561.386...	-4480.1420...	-2740.209...	3	139	Sit	Sit
1250	268.736...	384.9036...	2287.73...	528.906555...	-206.3616...	-123.325...	613.49029...	-6565.8144...	-3549.243...	6	139	Sit	Sit
1251	122.061...	137.5475...	2309.23...	670.162109...	-86.99020...	9.221191...	9199.3603...	-2640.2480...	-2007.861...	9	139	Sit	Sit
1252	35.3340...	-104.072...	2322.69...	408.805633...	-30.99632...	133.3007...	8070.0981...	351.047515...	322.99804...	10	139	Sit	Sit
1253	252.732...	-83.2736...	2293.54...	767.389038...	-66.29745...	-89.6923...	9552.4882...	133.923339...	-2064.990...	13	139	Sit	Sit
1254	71.6080...	602.2930...	2240.07...	815.272155...	-1005.428...	-613.674...	4137.9697...	-25263.335...	-18087.89...	1	140	Walk	Walk
1255	124.192...	347.4557...	2305.27...	723.094055...	-639.3823...	-152.373...	-873.9898...	-15421.454...	-4470.336...	2	140	Walk	Walk
1256	-44.827...	330.4302...	2333.88...	711.917236...	-666.0937...	27.39257...	-7908.906...	-18653.658...	-2676.269...	3	140	Walk	Walk
1257	293.212...	364.4812...	2276.66...	734.271240...	-612.6727...	-332.138...	6160.9404...	-12189.331...	-6264.404...	6	140	Walk	Walk
1258	147.531...	117.7854...	2306.47...	764.097534...	-592.8652...	-82.7124...	2818.0617...	-15176.252...	-2758.007...	9	140	Walk	Walk
1259	68.4017...	-122.206...	2325.02...	992.030883...	-544.0086...	69.95361...	17496.757...	-15390.369...	-1900.415...	10	140	Walk	Walk
1260	273.338...	-101.563...	2290.34...	618.170959...	-548.6874...	-96.0571...	-4476.543...	-14471.699...	-190.9423...	13	140	Walk	Walk
1261	89.4956...	560.7586...	2215.68...	536.625854...	-1246.032...	-731.967...	-8359.390...	-7218.1274...	-3548.803...	1	141	Walk	Walk
1262	154.806...	314.9643...	2294.73...	918.414245...	-974.7427...	-316.105...	5859.6059...	-10060.812...	-4911.987...	2	141	Walk	Walk
1263	-14.100...	283.6321...	2337.53...	921.794677...	-1403.942...	109.6215...	6296.3232...	-22135.472...	2466.8701...	3	141	Walk	Walk
1264	323.713...	346.2964...	2251.94...	915.033874...	-545.5426...	-741.833...	5422.8789...	2013.90380...	-12290.84...	6	141	Walk	Walk
1265	196.120...	88.66535...	2292.11...	1457.67053...	-873.6019...	-430.891...	20807.191...	-8422.0986...	-10445.36...	9	141	Walk	Walk
1266	156.778...	-152.595...	2309.93...	2651.29296...	-911.6585...	-452.871...	49777.863...	-11029.497...	-15684.74...	10	141	Walk	Walk
1267	318.090...	-122.672...	2269.05...	1342.56127...	-633.2626...	-638.481...	21731.712...	-2537.2580...	-16272.72...	13	141	Walk	Walk

Figure 5.10: Sample output (test data) from the decision tree classification: Comparing the labeled Class and the output from Classification Tree

5.3 Discussions

Figure 5.2 - Figure 5.7 show the plots of estimated X, Y and Z positions of wrist and elbow versus the measurements obtained from the Kinect sensor. FIGURE 5.8 shows the plots of these positions as compared to that obtained from very accurate OptiTrack™

system. It is however to be noted that Y-position data obtained from OptiTrack™ for the elbow joint has high amount of error. As seen from the position estimates of elbow and wrist, the Kalman filter removes significant noise from the kinematic model. As observed, the results obtained after the correction are nearly accurate and consistent with the ground truth (as shown via comparison with OptiTrack™ data). It is our belief that this approach will suit systems that have highly noisy image data input from vision sensors.

Figure 5.11: A typical classification tree graph for sit and walk action recognition

For the purpose of this work, we collected twenty five instances each of random walk to sit and sit to random walk transitions from three different individuals in a controlled

environment. This classification only considers two classes, sit and walk. The data collected from the three individuals was divided into training data and test data. The training data (Figure 5.9) comprises of data from the first two individuals and the test data (Figure 5.10) is obtained from the third individual. Each of the test and training data are labeled into the two above mentioned classes by analyzing the video data. The seven different attributes that were considered to build the decision trees are the velocity and acceleration along the three axes and the joint configuration information of the following joint/configuration points: the head, shoulder right, center and left, hip center, right and left. These points are chosen as these are fairly stable points during motion of the body that would have less erratic motion when compared to points such as wrist and the elbow. This analysis yields a classification accuracy of 92.2% and as well as a precision of 92.2%. Sample results of the classification are shown in FIGURE 5.10. Another important observation is that the characteristics of motion of the elbow joint as described in [11] and [12] is accurately described in this kinematic model. Korein in [11] observed that the movement of the elbow joint is restricted due to the constraint in movement induced by the shoulder joints.

Chapter 6

6. Conclusion and Future Work

In this chapter we summarize the results and conclusions obtained from the motion tracking algorithms introduced in this research and the results obtained through classification of the data into two different activities.

6.1 Summary and Conclusions

The research presents a method that exploits the kinematic motion of rigid bodies to track motion of human limbs. The method utilizes vision based sensors and hence is independent of external inertial sensors. The model is tested by comparing the results obtained by the proposed method applied on Kinect data with the data obtained from the commercially available highly-accurate OptiTrackTM motion tracking system. The result indicates that the model is fairly accurate. The advantage of this approach is its applicability to sensors that are purely based on data obtained from the vision system without requiring data from the external inertial sensors or the markers that are placed on the body. This research successfully fuses the kinematics based approach with image processing techniques to build an accurate model of human limb/ skeleton system. The usual limitations of a vision based system that depend on the environment such as lighting apply to this system as well.

In this research, initially we apply the optical flow constraint based Kalman correction on image data received from the Kinect. We then implement an inverse kinematic model similar to the work described by Zhou and Hu in [15]. However, these are the salient differences in this research:

1) The analysis in this research is based on the angular measurements and not the position data which enables us to develop a model based on the angular kinematics of the joints,

2) We do not incorporate data from external inertial sensors, accelerometers or gyroscopes in this work and hence the proposed Kalman Filter works based on the angular kinematic model rather than relying on the data from inertial sensors,

3) We base our simulations on the real-time tracking data obtained from a human actor or subject and so the model accurately analyses a human motion pattern, and

4) The optical flow equations are applied using spatio-temporal image constraints on raw image data and then the Kalman filtering scheme is applied to the Kinematic model.

Finally, a decision tree classification yields results with 92.2% accuracy that classifies two different activities, namely, walk and sit position, which is performed in lab environment.

6.2 Future Research Directions and Work

This research intends to develop an algorithm that can work with the latest marker-less motion capture systems in the market such as Microsoft Kinect where the input is primarily from a single camera. The target of this work was to build towards an ultimate goal of developing algorithms that can be applied to medical and in-home health care systems where the subjects can be monitored without requiring them to wear any external sensors on their body while at the same time monitoring their motion and behavioral patterns.

There is more scope to fuse the inverse kinematics approach to detection algorithms in computer vision systems to obtain more robust solution to tackle the problem of effective human motion tracking. The tracking model also can be extended to other motion capture systems. The availability of sequential data over a time series also implies that this work can be further expanded to analyze and classify human activities to build systems such as that for fall detection in smart homes. Our work analyses two activities of significance, and an extension to the work would be to incorporate more activities that are closely associated with day-to-day events of individuals living in nursing homes, especially to classify instances where there is emergency such as falls.

Bibliography

- [1] Moeslund, Thomas B., Adrian Hilton, Volker Krüger, and Leonid Sigal. *Visual Analysis of Humans*. Springer, 2011.
- [2] Sheikh, Yaser Ajmal, Ankur Datta, and Takeo Kanade. "On the sustained tracking of human motion." In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pp. 1-7. IEEE, 2008.
- [3] Datta, Ankur, Yaser Sheikh, and Takeo Kanade. "Linearized motion estimation for articulated planes." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33, no. 4 (2011): 780-793.
- [4] Zhou, Huiyu, and Huosheng Hu. "Human motion tracking for rehabilitation—A survey." *Biomedical Signal Processing and Control* 3, no. 1 (2008): 1-18.
- [5] Gavrilu, Dariu M. "The visual analysis of human movement: A survey." *Computer vision and image understanding* 73, no. 1 (1999): 82-98.
- [6] Rosenhahn, Bodo, Uwe Kersting, Smith Andrew, Thomas Brox, Reinhard Klette, and Hans-Peter Seidel. *A silhouette based human motion tracking system*. CITR, The University of Auckland, New Zealand, 2005.
- [7] Shotton, Jamie, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. "Real-time human pose recognition

- in parts from single depth images." *Communications of the ACM* 56, no. 1 (2013): 116-124.
- [8] Forsyth, David A., and Jean Ponce. *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference, 2002.
 - [9] Bergen, James R., Patrick Anandan, Keith J. Hanna, and Rajesh Hingorani. "Hierarchical model-based motion estimation." In *Computer Vision—ECCV'92*, pp. 237-252. Springer Berlin Heidelberg, 1992.
 - [10] Hartley, Richard, and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
 - [11] Korein, James U. *A geometric investigation of reach*. MIT press, 1986.
 - [12] Tolani, Deepak, Ambarish Goswami, and Norman I. Badler. "Real-time inverse kinematics techniques for anthropomorphic limbs." *Graphical models* 62, no. 5 (2000): 353-388.
 - [13] Horn, Berthold KP. "Closed-form solution of absolute orientation using unit quaternions." *Journal of the Optical Society of America A* 4, no. 4 (1987): 629-642.
 - [14] Aristidou, Andreas, and Joan Lasenby. "Real-time marker prediction and CoR estimation in optical motion capture." *The Visual Computer* 29, no. 1 (2013): 7-26.

- [15] Zhou, Huiyu, and Huosheng Hu. "Kinematic model aided inertial motion tracking of human upper limb." In *Information Acquisition, 2005 IEEE International Conference on*, pp. 6-pp. IEEE, 2005.
- [16] Kumar, Manish, and Devendra P. Garg. "Intelligent multi-sensor fusion techniques in flexible manufacturing workcells." In *American Control Conference, 2004. Proceedings of the 2004*, vol. 6, pp. 5375-5380. IEEE, 2004.
- [17] Bishop, Gary, and Greg Welch. "An introduction to the kalman filter." *Proc of SIGGRAPH, Course 8*, no. 27599-23175 (2001): 41.
- [18] Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." *Journal of Fluids Engineering* 82, no. 1 (1960): 35-45.
- [19] Wren, Christopher Richard, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. "Pfinder: Real-time tracking of the human body." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19, no. 7 (1997): 780-785.
- [20] Schepers, H. Martin, Daniel Roetenberg, and Peter H. Veltink. "Ambulatory human motion tracking by fusion of inertial and magnetic sensing with adaptive actuation." *Medical & biological engineering & computing* 48, no. 1 (2010): 27-37.
- [21] Baak, Andreas, Thomas Helten, Meinard Müller, Gerard Pons-Moll, Bodo Rosenhahn, and Hans-Peter Seidel. "Analyzing and evaluating markerless motion tracking using inertial sensors." In *Trends and Topics in Computer Vision*, pp. 139-152. Springer Berlin Heidelberg, 2012.

- [22] Julier, Simon J., and Jeffrey K. Uhlmann. "A new extension of the Kalman filter to nonlinear systems." In *Int. symp. aerospace/defense sensing, simul. and controls*, vol. 3, no. 26, pp. 3-2. 1997.
- [23] Wan, Eric. "Sigma-Point Filters: An Overview with Applications to Integrated Navigation and Vision Assisted Control." In *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE*, pp. 201-202. IEEE, 2006.
- [24] Nussbaum, Martha Craven, ed. *Aristotle's De Motu Animalium: Text with translation, commentary, and interpretive essays*. Princeton University Press, 1985.
- [25] Guzmán, Adolfo. "Decomposition of a visual scene into three-dimensional bodies." In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pp. 291-304. ACM, 1968.
- [26] Turk, Matthew, and Alex Pentland. "Eigenfaces for recognition." *Journal of cognitive neuroscience* 3, no. 1 (1991): 71-86.
- [27] Lee, Jung Keun, and Edward J. Park. "3D spinal motion analysis during staircase walking using an ambulatory inertial and magnetic sensing system." *Medical & biological engineering & computing* 49, no. 7 (2011): 755-764.
- [28] Sigal, Leonid, Alexandru O. Balan, and Michael J. Black. "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion." *International journal of computer vision* 87, no. 1-2 (2010): 4-27.

- [29] Yuan, Qilong, and I-Ming Chen. "Kinematic Uncertainties in Human Motion Tracking and Interaction." In *Advances on Theory and Practice of Robots and Manipulators: Proceedings of Romansy 2014 XX CISM-IFTToMM Symposium on Theory and Practice of Robots and Manipulators*, pp. 491-499. Springer International Publishing, 2014.
- [30] Gall, Juergen, Bodo Rosenhahn, Thomas Brox, and Hans-Peter Seidel. "Optimization and filtering for human motion capture." *International journal of computer vision* 87, no. 1-2 (2010): 75-92.
- [31] (June 11, 2014) OptiTrack™ System, <http://www.naturalpoint.com/optitrack/systems/>, last accessed on June 11, 2014
- [32] Muro-de-la-Herran, Alvaro, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications." *Sensors* 14, no. 2 (2014): 3362-3394.
- [33] Figueroa, Nadia B., Florian Schmidt, Haider Ali, and Nikolaos Mavridis. "Joint origin identification of articulated robots with marker-based multi-camera optical tracking systems." *Robotics and Autonomous Systems* 61, no. 6 (2013): 580-592.
- [34] (June 27, 2014) Falls Among Older Adults: An Overview <http://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html>, last accessed on June 27, 2014

- [35] (June 11, 2014) Microsoft Kinect Sensor Components and Specifications, <http://msdn.microsoft.com/en-us/library/jj131033.aspx>, last accessed on June 11, 2014
- [36] Clark, Ross A., Yong-Hao Pua, Adam L. Bryant, and Michael A. Hunt. "Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining." *Gait & posture* 38, no. 4 (2013): 1064-1066.
- [37] Gabel, Moshe, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. "Full body gait analysis with Kinect." In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 1964-1967. IEEE, 2012.
- [38] Lucas, Bruce D., and Takeo Kanade. "An iterative image registration technique with an application to stereo vision." In *IJCAI*, vol. 81, pp. 674-679. 1981.
- [39] Fleet, David, and Yair Weiss. "Optical flow estimation." In *Handbook of Mathematical Models in Computer Vision*, pp. 237-257. Springer US, 2006.
- [40] Gaschler, Andre. "Real-time marker-based motion tracking: application to kinematic model estimation of a humanoid robot." *Master's thesis, Technische Universität München, Germany* (2011).
- [41] Barron, John L., David J. Fleet, and Steven S. Beauchemin. "Performance of optical flow techniques." *International journal of computer vision* 12, no. 1 (1994): 43-77.
- [42] Baker, Simon, and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework." *International journal of computer vision* 56, no. 3 (2004): 221-255.

- [43] Roetenberg, Daniel. *Inertial and magnetic sensing of human motion*. University of Twente, 2006.
- [44] (June 11, 2014) Vicon camera systems using markers, <http://www.vicon.com/System/Markers>, last accessed on June 11, 2014
- [45] Ke, Shian-Ru, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. "A review on video-based human activity recognition." *Computers* 2, no. 2 (2013): 88-131.
- [46] Sempena, Samsu, Nur Ulfa Maulidevi, and Peb Ruswono Aryan. "Human action recognition using dynamic time warping." In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, pp. 1-5. IEEE, 2011.
- [47] Kotsiantis, Sotiris B., I. D. Zaharakis, and P. E. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 3-24.
- [48] Bouchard, Kevin, Bruno Bouchard, and Abdenour Bouzouane. "A new qualitative spatial recognition model based on Egenhofer topological approach using C4. 5 algorithm: experiment and results." *Procedia Computer Science* 5 (2011): 497-504.
- [49] Ramirez-Amaro, Karinne, Eun-Sol Kim, Jiseob Kim, Byoung-Tak Zhang, Michael Beetz, and Gordon Cheng. "Enhancing Human Action Recognition through Spatio-temporal Feature Learning and Semantic Rules." *Humanoid Robots* (2013).

- [50] Ravi, Nishkam, Nikhil Dandekar, Preetham Mysore, and Michael L. Littman. "Activity recognition from accelerometer data." In *AAAI*, vol. 5, pp. 1541-1546. 2005.
- [51] Quinlan, John Ross. *C4. 5: programs for machine learning*. Vol. 1. Morgan kaufmann, 1993.
- [52] Wang, Wei, Zong-yu Liu, and Rong-rong Xie. "Quadratic extended Kalman filter approach for GPS/INS integration." *Aerospace Science and Technology* 10, no. 8 (2006): 709-713.
- [53] Zhao, Tao, and Ramakant Nevatia. "Tracking multiple humans in crowded environment." In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II-406. IEEE, 2004.
- [54] Wan, Eric A., and Rudolph Van Der Merwe. "The unscented Kalman filter for nonlinear estimation." In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 153-158. IEEE, 2000.