

Analysis of Motor Skills in Subjects with Down's Syndrome using Computer Vision
Techniques

by

Jeremy Paul Svendsen
B.Sc., University of Victoria, 2007

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTERS OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

© Jeremy Svendsen, 2009
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by
photocopying
or other means, without the permission of the author.

Analysis of Motor Skills in Subjects with Down's Syndrome using Computer Vision
Techniques

by

Jeremy Paul Svendsen
B.Sc., University of Victoria, 2007

Supervisory Committee

Dr. A. Branzan Albu, Supervisor
(Department of Electrical and Computer Engineering)

Dr. P. Agathoklis, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. J. H. Weber-Jahnke, Outside Member
(Department of Computer Science)

Supervisory Committee

Dr. A. Branzan Albu, Supervisor
(Department of Electrical and Computer Engineering)

Dr. P. Agathoklis, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. J. H. Weber-Jahnke, Outside Member
(Department of Computer Science)

ABSTRACT

Computer vision techniques for human motion analysis have the potential to significantly improve the monitoring of motor rehabilitation processes. With respect to traditional marker-based techniques, computer vision offers both portability and low-cost. This thesis describes methods that have been designed for the analysis of the motor skills of subjects with Down's syndrome. More specifically, the motion of interest is weight-shifting; this motion plays an important role in the safety of locomotory activities, as well as of other daily actions.

From a theoretical viewpoint, the thesis proposes several new concepts for human motion analysis and describes their algorithmic implementation, as well as their applicability to the detection and description of several motion primitives.

The thesis introduces the concept of curved bounding box, which is an extension of the rectangular bounding box that is typically used for detection and tracking of rigid motion. This concept is successfully applied to the detection of deformable motion, such as arm, knee and upper body motions.

A new technique for identifying subject-representative patterns of motion is also proposed. This technique is based on Motion History Images, which hold both analytical and visualization power.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Acknowledgements	xi
Dedication	xii
1 Introduction	1
1.1 Motor characteristics of patients with Down’s Syndrome	2
1.2 Technical Challenges Facing Video-Based Human Motion Analysis . .	3
1.3 Contributions	6
2 Literature Review on Motion Description from Video Data	8
2.1 Part-based motion description	8
2.2 Joint-based motion description	9
2.3 Multiple Cameras	11
2.3.1 Visual Hull	11
2.3.2 Stereo Vision	12
2.4 Action Estimation	12
2.5 Motion History Images	13
3 Proposed Approach	15
3.1 Basic Assumptions	16

3.2	Discussion of Parameters Involved in Motion Measurements	18
3.3	Background Subtraction	18
3.4	Generation of Curved Bounding Boxes	20
3.4.1	Two Spline Bounding Box	22
3.4.2	Four Spline Bounding Box	24
3.4.3	Parabolic Bounding Box	25
3.5	Detection of Arm Motions	26
3.5.1	Detection of Knee Leans	29
3.5.2	Detection of Hip and Upper Body Motions	29
3.6	Identification of Specific Patterns of Motion	31
4	Results	35
4.1	Experimental Database	35
4.2	Running Time	36
4.3	Experimental Validation of Motion Detectors Based on Bounding Box Information	37
4.3.1	Arm Motion Detection Using a Two Spline Bounding Box	38
4.3.2	Knee Motion Detection	40
4.3.3	Hip and Upper Body Motion Analysis	41
4.4	Identification of Specific Patterns of Motion	49
5	Conclusion and Future Work	57
5.1	Conclusion	57
5.2	Future Work Directions	58
A	The Natural Cubic Spline Function	60
B	The Linear Least Squares Method for a Parabola	62
	Bibliography	64

List of Tables

Table 3.1	List of the geometric primitives used to generate the bounding boxes.	21
Table 3.2	List of the types of bounding boxes and the motions they analyze.	21
Table 4.1	The measured running times of the different methods.	37
Table 4.2	Experimental validation of the arm motion detector for the left side.	39
Table 4.3	Experimental validation of the arm motion detector for the right side.	39
Table 4.4	Experimental validation of the arm motion detector.	39
Table 4.5	Experimental validation for the knee lean motion detector for the left side.	41
Table 4.6	Experimental validation for the knee lean motion detector for the right side.	41
Table 4.7	Experimental validation of the knee motion.	41
Table 4.8	Motion counts and times for subject 1.	50
Table 4.9	Motion counts and times for subject 2.	51
Table 4.10	Motion counts and times for subject 3.	52
Table 4.11	Motion counts and times for subject 4.	53
Table 4.12	Motion counts and times for subject 5.	54
Table 4.13	Motion counts and times for subject 6.	55
Table 4.14	Motion count for subject 7.	56

List of Figures

Figure 3.1	Modular diagram of the proposed approach.	17
Figure 3.2	(a) A rectangular bounding box around a subject. (b) A bounding box exactly fitting the shape of the subject.	22
Figure 3.3	Examples of bounding boxes around all the subjects. Several different motions are represented here. a) Standing. b) Bending over to the left with arm out. c) Both arms out. d) Leaning left with arms in. e) Leaning left with left arm in and right arm partially out. f) Bending over forward with hands on hips. g) Moving only the hips to the left.	23
Figure 3.4	The arrows indicate the locations of where splines are joined. a) Two spline bounding box. b) Four spline bounding box.	24
Figure 3.5	An example of the bounding box which uses 4 splines on each side instead of 2.	25
Figure 3.6	a) The first subject standing upright with the parabolas shown. In this case the parabolas are both convex. b) An example of a person over-extending their left leg. In this case the parabola on the right is concave.	27
Figure 3.7	The contour of subject showing how the points are taken.	27
Figure 3.8	The radial coordinates of the bounding box relative to the subjects centroid when the subjects arms are at their side. The pose which generated this plot is also presented.	28
Figure 3.9	The radial coordinates of the bounding box relative to the subjects centroid when the subjects arms are reaching out. The pose which generated this plot is also presented.	29
Figure 3.10	An example of a knee lean being performed by subject 5	30
Figure 3.11	a) An example of a histogram generated by analysis of the curvature. For the histogram only slight leans occurred. b) Key frame of the motion which generated this histogram.	31

Figure 3.12a) A second example of a histogram. b) The subject which the histogram represents.	32
Figure 3.13 The arrows indicate the boundaries of the upper, middle, and lower divisions for the motion history images.	34
Figure 3.14a) Motion history image of subject 3 putting her arm out and leaning to the other side. b) Motion history image of subject 7 performing a pronounced lean to the left side.	34
Figure 4.1 a) The first frame of the video sequences of the subjects with Down's syndrome. b) An example frame from the video sequence watched by the subjects.	36
Figure 4.2 a) frame where a missed detection of the arm occurs; b) frame where a false detection of the arm occurs.	40
Figure 4.3 Histograms representing typical subjects standing upright with slight leans to either side. All the histograms in the left column correspond to the left side of a subject and all the histograms in the right column correspond to the right side of the side of a subject. Rows a) and b) correspond to the top subject and rows c) and d) correspond to the bottom subject.	43
Figure 4.4 Histograms representing typical subjects leaning sharply to one side and compensating by leaning their upper body in the opposite direction. These were performed by typical subjects for 40 seconds. The left column corresponds to the left side of the subject and the right column corresponds to the right side. Rows a) and c) are the same subject leaning left and then right and rows b) and d) are a different subject leaning left then right. . .	45
Figure 4.5 Histograms representing typical subjects performing slight leans with their arms out at either side. The left column is the left side of the subject and the right column is the right side. Rows a) and b) correspond to sequences by the top subject and rows c) and d) correspond to sequences performed by the bottom subject.	46
Figure 4.6 The histograms for subjects 1 to 4. The left column is curvature of the left side of the subject and the right column is the right side of the subject. At the end of each row the subject responsible for each set of histograms is shown in a typical pose.	47

Figure 4.7	The histograms for subjects 5 through 7. The left column is curvature of the left side of the subject and the right column is the right side of the subject. At the end of each row the subject responsible for each set of histograms is shown in a typical pose.	48
Figure 4.8	a) An example of the most frequent by subject 1. b) The motion history image corresponding to the motion seen in a). This motion is a mid level motion on the left side. c) An example of the most common motion of subject 1 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is a lower motion using the right side.	50
Figure 4.9	a) An example of the most common motion by times performed by subject 2. b) The motion history image corresponding to the motion seen in a). This motion is a middle level motion on the right side. c) An example of the most common motion of subject 2 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is an upper level motion using the right side.	51
Figure 4.10a)	An example of the most common performed by subject 3. This frame was one of the frames used to generate the motion history image seen in b). This is a middle level motion using both sides.	52
Figure 4.11a)	An example of the most common motion by frequency and time length by subject 4. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a lower level motion using the left side.	53
Figure 4.12a)	An example of the most frequent motions and the most time consuming motions performed by subject 5. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a middle level motion using the right side.	54

Figure 4.13a) An example of the most common motion by times performed by subject 6. b) The motion history image corresponding to the motion seen in a). This motion is a middle level motion on the right side. c) An example of the most common motion of subject 6 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is a middle level motion using both sides. 55

Figure 4.14a) An example of the most frequent motions and the longest motions performed by subject 7. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a lower level motion using the left side. 56

Figure A.1 An example of a cubic spline function. 61

ACKNOWLEDGEMENTS

I would like to thank:

my family, for supporting me,

my friends, for entertaining me,

Dr. Alexandra Branzan Albu, for mentoring, support, encouragement, and patience.

Dr. Naznin Virji-Babul for providing the database for this study as well as valuable information in the field of Down's syndrome and motor rehabilitation.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny...'

Isaac Asimov

DEDICATION

To my family.

Chapter 1

Introduction

Human motion analysis represents a new research direction in computer vision. Indeed, as shown by Moeslund and colleagues [1, 2], the current state of the art in computer-vision based human motion analysis is mostly represented by surveillance applications, therefore it targets activity and subject identification. This prior work can not be used as a starting basis for motion analysis, since the latter is different in scope than motion recognition.

Human motion analysis aims at generating a detailed motion description, which can be used either for categorization purposes or for visual inspection by an expert. Human activity recognition is strictly based on categorization and therefore uses limited motion descriptions. For instance, in medical applications, the main issue is to be able to describe how a subject moves, as opposed to classifying his/her motions as normal/abnormal. Moreover, in medical studies, there is a whole range of abnormalities and the interest is not in determining whether a motion is normal or abnormal, but in describing the type of abnormalities present and how they affect the performance of a certain action.

Our research focuses on the development of a new computer vision-based approach for the study of motion characteristics in human subjects suffering from motor impairment. More specifically, our study focuses on describing the motion of patients with Down's syndrome. The main goal of our study consists in identifying subject-representative patterns of motion and in providing a quantitative and qualitative description of these patterns. The following subsection provides concise medical background information about motor abnormalities typically observed in Down's syndrome patients.

1.1 Motor characteristics of patients with Down's Syndrome

Down's syndrome (DS) is one of the most common causes of mental retardation as it occurs in approximately 1 of every 650 live births. It is caused by the presence of an extra copy of chromosome 21. DS produces motor impairments of varying degrees, which can be caused either by low muscle tone [3], or by an impairment located at the level of perceptual-motor coupling [4, 5].

Regardless of the cause responsible for the motor impairments, recent studies have shown that individuals with DS are able to learn to develop adaptive motor strategies that optimize safety and stability [6]. The learning environment must be very carefully designed in order to avoid the development of fixed patterns of motor behavior.

Virtual reality technology provides new opportunities for developing realistic task

scenarios and is thus rapidly gaining acceptance in the rehabilitation research community.

One important open question is related to how to measure motor progress within a virtual reality environment without cumbering the subject with wired motion captors. Our study aims at finding an answer to this question by using computer vision technology for human motion analysis. This technology offers important advantages such as pervasiveness, low-cost and portability with respect to traditional technologies of marker-based motion capture.

1.2 Technical Challenges Facing Video-Based Human Motion Analysis

Analyzing human motion based only on visual information poses several challenges which were first summarized by Sminchisescu and Triggs [7] from a motion recognition perspective. The list below presents these challenges from the viewpoint of human motion analysis rather than recognition.

1. Voluntary human motions are performed by the musculoskeletal system. Accurate models of human musculoskeletal systems are complex and require at least 30 different joint parameters [8]. Furthermore, the skeletal structure is subject to non-linear joint limits and non-self-intersecting constraints (ie. body parts cannot intersect each other). Visual information alone is not sufficient for building a high accuracy and high detail model of human motion. While such models can be obtained via non-visual measurements, computer vision based

solutions offer the advantage of being non-intrusive, markerless and low-cost.

The challenge is identifying the right level of detail and accuracy in video-based human motion analysis.

Many researchers focus on detecting and tracking body parts separately, then assembling them into a model. For instance, Forsyth and Fleck [9] proposed a general methodology of body plans for finding humans in images. They used generalized cylinders for body part detectors. Felzenszwalb and Huttenlocher [10] used colour-based part detectors and dynamic programming in order to assemble body parts into pictorial structures. Micilotta et al. [11] use the RANSAC (Random Sample Consensus) method for estimating the pose from body parts detected with Adaboost.

In human motion analysis, the degree of accuracy to which motion is modeled depends on the task at hand. In some cases, the motion of interest is confined at the level of a particular joint (for instance, we are interested in quantifying how much the subject bends his/her knee). In other cases, the motion of interest involves more than one joint (ie. lower body bending involves both knees and hips, and we are interested in observing the coordination between the motion of these two joints). The motion representation is specific to the type of motion being looked for.

2. Even if all joint locations are known in a given frame, ambiguities still exist. Since one does not know depth information, there are two solutions for as many

as 10 different joint configurations. This means there can be over a thousand different possible solutions for a given point configuration. This restriction is lifted by those which use multiple cameras (section 2.3). The challenge consists in the high computational complexity of a joint-based human motion model (section 2.2).

3. Loose clothing has a significant impact on the shape of the human motion. *The challenge consists of finding motion descriptors which are invariant to variable clothing.* Body part detectors, such as the ones mentioned above, fail under loose clothing circumstances. Therefore, silhouette-based methods and deformable whole body contours are more appropriate for describing human motion in such conditions. Our proposed approach (see chapter 3) employs a silhouette-based method for motion identification and measurement.
4. Self occlusion restricts observations from a single viewpoint. About one third of the over 30 degrees of freedom (joints) are hidden from view at all times. Motions that are typically self-occluded occur in the plane orthogonal to the the image plane. Self occlusion is partially solved by using multiple cameras (section 2.3) but are not performing stereo vision (section 2.3.2).
5. Human actions occur as a continuous flow of events. There is no natural breaks between different the action being performed. *The challenge consists in the segmentation of spatiotemporal events (motions, actions) without a priori information about their structure.* There are surprisingly few studies to address

this problem. Branzan Albu et al. [12] perform the temporal segmentation of cyclic human activities based on tracking changes in periodicity. Ali and Aggarwal [13] segment continuous human activity into distinct actions using changes in angles formed by their body parts (torso, upper leg, lower leg) with the vertical axis.

1.3 Contributions

The main contribution of this research is to provide a series of algorithms which help understand the full body motions of subjects. In specific, the algorithms were designed for and tested on subjects with Down's Syndrome. A proper understanding about a subjects full body motions can help correct improper motions.

Unlike most previous attempts to detect and categorize human motion, these methods do not look for each body part individually to construct a motion. Instead shapes are fitted over the subject to reduce complexity. This allows a computer to analyze a mathematical curve instead of the more complex shape of a person.

Although the algorithms were tested on subjects with Down's Syndrome, they also work for other subjects as well. In this case the algorithms assess the motions of healthy people or people with other disabilities. The benefits of this include improving ones ability to perform at various sports, workplace injury avoidance for physical labor jobs, and simply moving around. Even people who would normally be considered to be very healthy could benefit from these algorithms.

The applications of these algorithms extend beyond human motion analysis to

articulated motion analysis. This is useful if the shape of the articulated object is variable or are not easily retrievable. This approach does not require a predefined model of the observed shape and so it can easily be adapted to other uses.

Chapter 2

Literature Review on Motion Description from Video Data

In the last two decades, a substantial amount of work has been done in the area of human motion detection. Typically, human motion recognition involves the following steps: detection of the human body (or its parts) performing the motion, tracking, motion description, and classification. The most relevant step for human motion analysis is the motion description. The remainder of this section provides a taxonomy of the motion descriptions proposed in literature and discusses the most relevant contributions to this topic.

2.1 Part-based motion description

The methods in this category take a bottom-up approach by first identifying body parts, and then assembling those for pose estimation. The motion is then described as a sequence of poses that are estimated on a frame-by-frame basis.

Early work by Ju et al [14] represents a persons limbs via a set of planar patches;

they assume that these patches correspond to body parts with rigid motion.

Shortly afterward Forsyth and Fleck [9] introduced the concept of body plans to represent people and animals as a structured set of parts. Body plans are learned from image data using statistical methods.

Felzenszalb and Huttenlocher [10] use a similar concept called pictorial structures for representing the human body in motion as a collection of parts arranged in a deformable configuration. Once detected, the parts are matched together in a spring-like manner by using a energy minimization algorithm.

Roberts et al [15] use a top-down approach for the probabilistic matching of body parts and thus are able to handle partial occlusion. Their matching approach is based upon learning view-dependent probabilistic models of body part shapes that take into account intra and inter person variability (in contrast to rigid geometric primitives).

Another method for matching body parts onto people is to use AdaBoost. The AdaBoost algorithm, first described by Freund and Schapire [16], determines the best location of the body parts. This method was used by Micilotta et. al. [11].

2.2 Joint-based motion description

Since joints are typically represented in the image space as points, a popular joint-based representation of the human body in motion is the stick-figure. The motion description directly associated to it consists in a set of spatio-temporal trajectories corresponding to each tracked joint as in Jean et al [17].

The stick-figure is widely used for gait detection and recognition, due to the unique

signature of gait obtained from the spatio-temporal trajectories of joints. Stick figures are typically estimated using spatio-temporal information from tracking as in Jean et al [18]. They can also be estimated on each frame using only spatial information, either via the medial axis transform [19] or via the distance transform [20]. Regardless of the way in which they are estimated, stick figures represent noisy versions of the human skeleton.

Another motion representation integrating information about extreme joint locations is the star skeleton proposed by Fujiyoshi [see reference 8 from Aggarwal and Yu]. The star skeleton is obtained by connecting the centroid of the human body to a predefined number of extreme points located on its contour.

Aggarwal and Yu extend this model to contain two stars, one centered on the centroid of body, the other centered on the highest contour point of the body. They represent motion information via a codebook where symbols are represented by feature vectors derived from the two-star skeleton.

Methods which are classified as model based have very distinct a priori model. This involves putting constraints on the detected pose of the subject based on the predefined model. Most of these methods try and create a 3D skeletal frame of the subjects pose. Several models also put the constraint that the individual body parts cannot intersect each other.

When using a single camera it is more difficult to obtain 3D pose information. To overcome this, Bregler et. el. [21] have devised a method in which a skeletal frame is used to provide constraints. Since the possible twists and motions of people are

constrained, one can determine the actual directions of motion. In specific, they made a geometric model and a twist motion model, and then combined the two. They then go on to track the twists and turns of individual body parts in order to track a person while they walked.

2.3 Multiple Cameras

2.3.1 Visual Hull

One common method for reconstructing the pose of a subject is to create a visual hull. A visual hull is the apparent pose of a subject based on multiple different camera angles. This is done by knowing the location of the camera and using point correspondence on the subject. The visual hull is the reconstructed 3D representation of the subject which is constructed using voxels. This method was performed by [22, 23, 24]. A common method for creating a virtual hull is to first create a joint skeleton of the person.

The concept of visual hulls is also useful in the field of computer graphics and computer animation. In computer animation it is often desired to create very realistic representations of people or animals from an arbitrary point of view. In order to do this, images of people or animals are taken from many angles. Gortler [25] describes a method for calculating a visual hull using silhouette image data. Since one of their goals was to produce visual hulls in real time, their algorithm is designed to be very cost effective.

In addition to producing a visual hull, Shakhnarovich et. al. [26] go further and

by also performing gait analysis and facial recognition. They create a visual hull by piecing together silhouettes from multiple different angles. The gait analysis is done by partitioning a silhouette into 7 regions and then the centroid of each region is calculated. The centroids are tracked and a feature vector is generated which is matched using a Gaussian model and a nearest neighbor classifier to determine who the cameras were observing. The researchers also performed facial recognition by creating a virtual face from the virtual hull. This worked since face detection and/or recognition works best if the face is viewed directly from the front.

2.3.2 Stereo Vision

One advantage of using stereo vision is that it is much easier to detect self occlusion. This was done by Plankers and Fau [27] to obtain a full 3D pose of the subject even with partial occlusion.

2.4 Action Estimation

Some methods for detecting human motion do not look for all the body parts of the subject and then reconstruct a shape. Many computer vision methods only observe motion and some just look for specific shapes. The method proposed by Yu and Aggarwal [28] is designed to detect jumping over a fence by observing the overall shape of the subject. When a subject jumps over a chain link or iron bar fence, they form a very distinctive shape which can be detected via computer vision. This shape is distinct since the person is using their arms and one leg for support with the

other leg dangling there is a star like shape produced. The algorithm goes further and breaks down action into 4 different stages. The stages are walking, climbing up, crossing over the top, and climbing down. They then went on to analyze the action using the Viterbi algorithm of the Hidden Markov Model decoding problem.

There are many papers written about interactions of people with either other people or other objects. When there are multiple people in the scene the algorithm must now be able to distinguish between the different people. In a paper by Park and Aggarwal [29] a Hierarchical Bayesian method is used to detect violent interactions between two people.

2.5 Motion History Images

Motion history images are a powerful tool for human motion analysis. This can be used to compile an entire motion into a much more compact, manageable representation. Branzan Albu and Beugeling [30] analyze human motion of the elderly using a 3D motion history representation. Their 3D Motion History Image had two spatial dimensions and time. This method was used to look for sways in the subjects motion as they did common tasks like stepping up on a platform or sitting down into a chair.

In a work by Bradski and Davis [31], the concept of timed motion history images is presented. The concept they present is to add a floating point timestamp to each previous frame so that motion information can be obtained. The result is an image in which the intensity gradient can be used to determine motion velocity. In a later paper by Davis [32], image pyramids are used to analyze the motion in the motion

history image at different scales. The end result is a flow chart describing the motions of the subject.

Yilmaz and Shah [33] expand on the concept of a 3D motion history image. In addition to layering the frames in the temporal direction they connect them using point correspondence. This generates what they call a spatiotemporal volume (STV). The paper analyzes the motions by looking at the shape of the contours generated. The shape of the contours conveys the speed as well as the type of action being performed. The method by Gorelick and Blank [34] also analyzes motions as space-time shapes by using the Poisson equation.

Chapter 3

Proposed Approach

Our proposed approach focuses on detecting and analyzing weight shifting patterns in the motion of Down's syndrome subjects. As shown in section 1.1, the ability of performing correct weight shifting is essential for many activities of daily living (ADL), including locomotion.

In order to be able to isolate weigh-shifting patterns from other motor activities, Down's syndrome subjects were filmed while controlling a virtual snowboard and navigating through predefined obstacles in a virtual reality game. Additional information about the experimental procedure is found in chapter 4. A typical weight-shifting pattern involves knee flexing and hip translation. Abnormal weight shifting occurs when a subject overcompensates and uses the upper body (arms, trunk) to shift weight. Therefore, in order to characterize a weight-shifting pattern, attention needs to be paid to arm, trunk, hips, and knee motions.

Our approach analyzes whole body motion and does not use a joint-based kinematic model, due to the variable appearance of the subjects (loose clothing). More specifically, we introduce the concept of curved bounding boxes (spline-based and

parabolic-based) and use it in order to identify knee, arm, hip and the upper body motions, as summarized in the diagram presented in figure 3.1.

Moreover, in order to characterize and visualize the weight-shifting, patterns that correspond to a specific subject, we generate motion history images, which are further classified into classes of patterns (see figure 3.1).

The remainder of this chapter is structured as follows. Section 3.1 outlines the basic assumptions used by the proposed approach. Section 3.2 describes the background subtraction procedure. Next, the generation of curved bounding boxes is described in section 3.4. The final section describes how the motion history images are produced and used to describe the weight-shifting pattern that is specific for each subject.

3.1 Basic Assumptions

The proposed approach is based on three application-specific assumptions which in the video sequences of our database hold true for almost all frames. In the frames where these assumptions are not true, there is a small error in the results of the algorithms.

1. The subject's feet are stationary throughout the entire video sequence. This assumption held true for almost all frames except for one instance where a subject jumped and one where a subject changed their stance. If this assumption is broken, the analysis of the bounding box involving knee motions is incorrect as well as the segmentation of the motion history images.

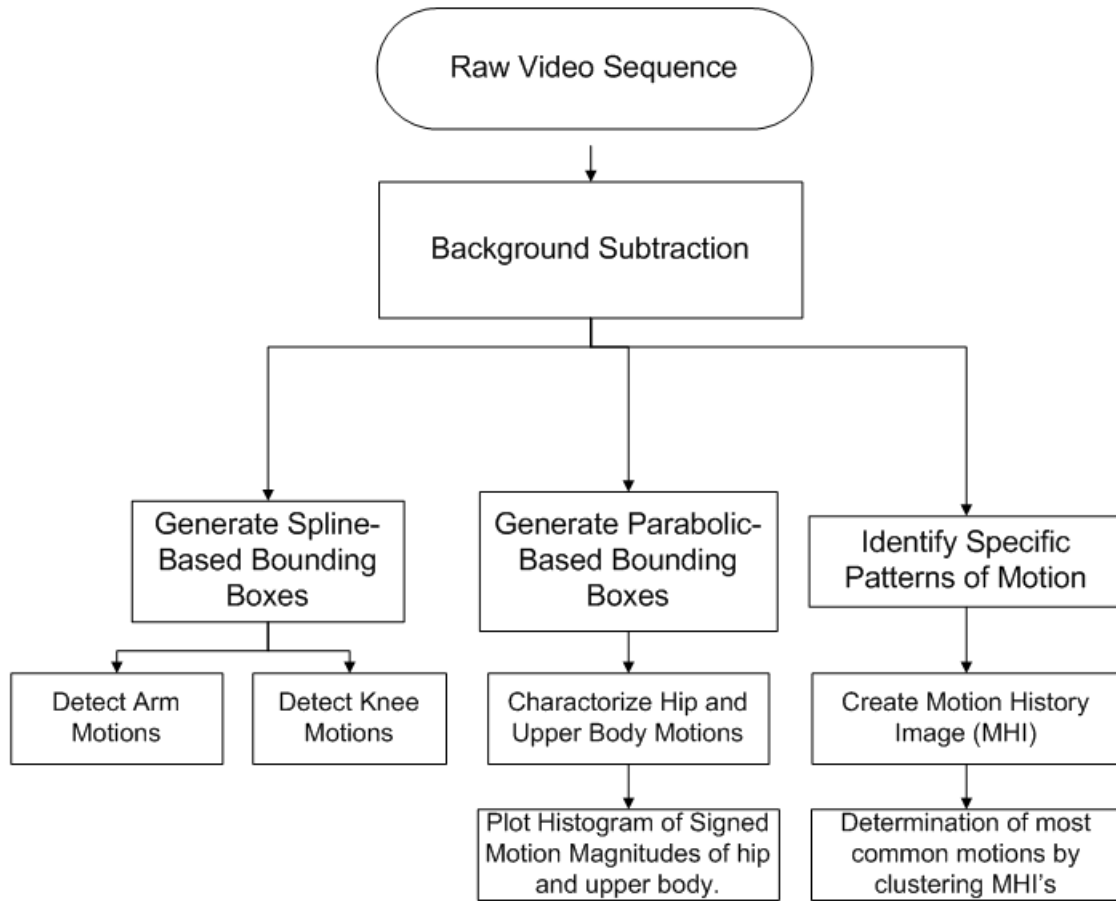


Figure 3.1: Modular diagram of the proposed approach.

2. The subject's head is located at the topmost part of their body. This assumption holds true for all frames except a few where the subjects put their hands above their heads. We need this assumption in order to detect the head as the uppermost region of the body.
3. The motion is completely performed in a plane that is parallel to the image plane. This assumption holds true because the subjects interact with the virtual obstacle course that is projected onto a screen located at their front.

3.2 Discussion of Parameters Involved in Motion Measurements

In order to be able to compare motions of different subjects, one must first take into account the variability in the shape of the subject. To account for shape variability, each subject was compared to an upright stationary image of themselves. In this research the upright posture was considered to be the reference posture and everything else was compared to it. An upright posture near the beginning of each subject's motions was manually selected so that all other measurements could be compared to it. This method allows the experimenter to get the subject's height and width in pixels, and therefore to express motion-induced displacements in meters as well. The upright reference frame is also used as reference for determining whether the subject is in an upright posture or not.

Subjects performed actions in different ways, involving different parts of the body to different extents. All of them had some leaning but some of them had almost no arm motions and some had a lot of arm motions. For both leans and arm motions, thresholds were needed. For example, how far to one side must a subject be before the motion is considered a lean.

3.3 Background Subtraction

Background subtraction implies removing all pixels which are not a part of the subject. There are several different algorithms for removing the background from a video sequence. It is not uncommon to have background subtraction in mind prior to

capturing the video. For example, one common method is chroma keying. This method uses a constant colour background and then removes all pixels with that colour. This method runs into problems if the user is wearing the same colour as the background.

Another technique which is often used is to make a short video sequence of the background only. In this case the each pixel of the analyzed video is compared using statistical methods to the same pixel in the the background video. This method has a few limitations. The first is that once the background sequence is made, the camera cannot be moved. The second limitation is that the background must be completely static and the lighting must be uniform. Even with uniform lighting, reflective surfaces will show up.

The video sequence used in this experiment was captured against a quasi-homogeneous green background, but without a corresponding background-only sequence. Thus, a background frame was obtained by manually outlining each subject in the first frame. Subtracting the background frame from the every current frame removed almost all of the background. A pixel is considered background if equation 3.1 is true. In equation 3.1 the i th background pixel is subscripted with a ib . A pixel is considered a green pixel if it follows equation 3.2.

$$|R_{ib} - Ri| + |G_{ib} - Gi| + |B_{ib} - Bi| < 50 \quad (3.1)$$

$$G > R + 50 \wedge G > B + 50 \quad (3.2)$$

3.4 Generation of Curved Bounding Boxes

By definition a bounding box for a set of objects is a closed area which completely contains the union of the objects in the set.

One of the most common uses of bounding boxes has been to locate the subjects in an image as in [35, 36, 37, 38]. This is useful for tracking moving targets in the surveillance domain. In this case, the bounding box is just to show the user the location of an object of interest. Bounding boxes have been used before in applications like the cardboard model [14] where each different body part was located. In this case, the rectangles were used to show their algorithm's accuracy at locating individual body parts.

The majority of previous work has focused on using rectangular bounding boxes to indicate the location of a moving object. Rectangular bounding boxes are defined by the smallest rectangle which completely encompasses the region of interest. In order to describe deformable human motion, we introduce a new concept of a curved bounding box. We will then show how a curved bounding box can be used to describe human actions such as weight shifting.

Many factors must be considered when choosing a shape for a bounding box. The shapes can range from a simple rectangle as in figure 3.2a to a complete outline as in figure 3.2b. In some cases, as in figure 3.2b, the bounding box may have a deformable contour. In general, the more complex the shape is, the higher the computational complexity and running time for the bounding box generation algorithm will be. In other words, it is quick and easy to analyze the motion of a rectangle, yet slow

to analyze the motions of a complex shape. A complex shape can also yield more information than a simple shape. Rectangular bounding boxes are limited to only measuring translation, rotation, and a change in aspect ratio. Deformable bounding boxes can also estimate the motions of the parts within the objects.

In this study we are using bounding boxes to measure human motion. Unlike previous studies, the bounding boxes in this project were used for more than just providing information about the location of the subject. Instead of a rectangular bounding box, a deformable one which changes shape as the subject moves is used. This means the shape of the bounding box is tied to the pose of the subject. The shapes that are being used to describe the bounding are shown in table 3.1. In table 3.2 the motions observed by each different bounding box type are listed.

Body Side	Geometric Primitive
Right	Multiple Cubic Splines or a single Parabola
Bottom	Line
Left	Multiple Cubic Splines or a single Parabola
Top	Half Ellipse

Table 3.1: List of the geometric primitives used to generate the bounding boxes.

Bounding Box	Types of Motion Analyzed
Two Splines per Side	Arm Motions
Four Splines per Side	Knee Motions
Parabolas	Hip Motions and Trunk

Table 3.2: List of the types of bounding boxes and the motions they analyze.

The following subsections describe the three types of bounding boxes that were used in the approach, namely the two spline bounding box, the four spline bounding box, and the parabolic bounding box. In general, the more splines used the more

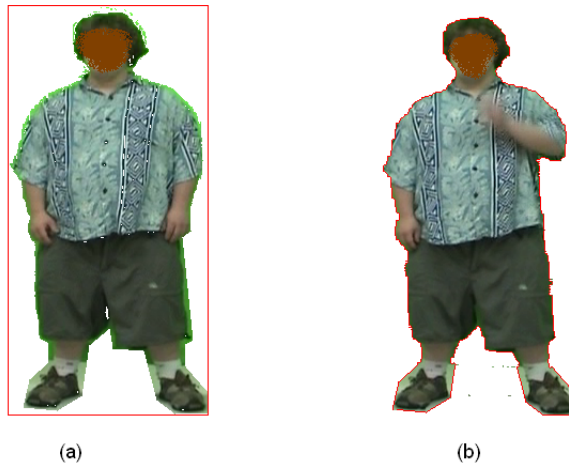


Figure 3.2: (a) A rectangular bounding box around a subject. (b) A bounding box exactly fitting the shape of the subject.

information available. The downside of adding more splines consists in the increase in the computational complexity.

3.4.1 Two Spline Bounding Box

The bounding box is pieced together by using several different mathematical curves. Since the feet of the subjects were relatively stationary, the bottom of the bounding box is just a simple line. The top of the head was drawn as half of an ellipse.

The sides of the bounding box are more complex than a line or ellipse. Each side is drawn using two natural cubic splines. Additional information on the natural cubic spline can be found in appendix A.

The bounding boxes were drawn such that the intersection point between the two splines was chosen to be the outermost point on either side of the subject. For example, if the subject had their arms out the intersection would be at the ends of

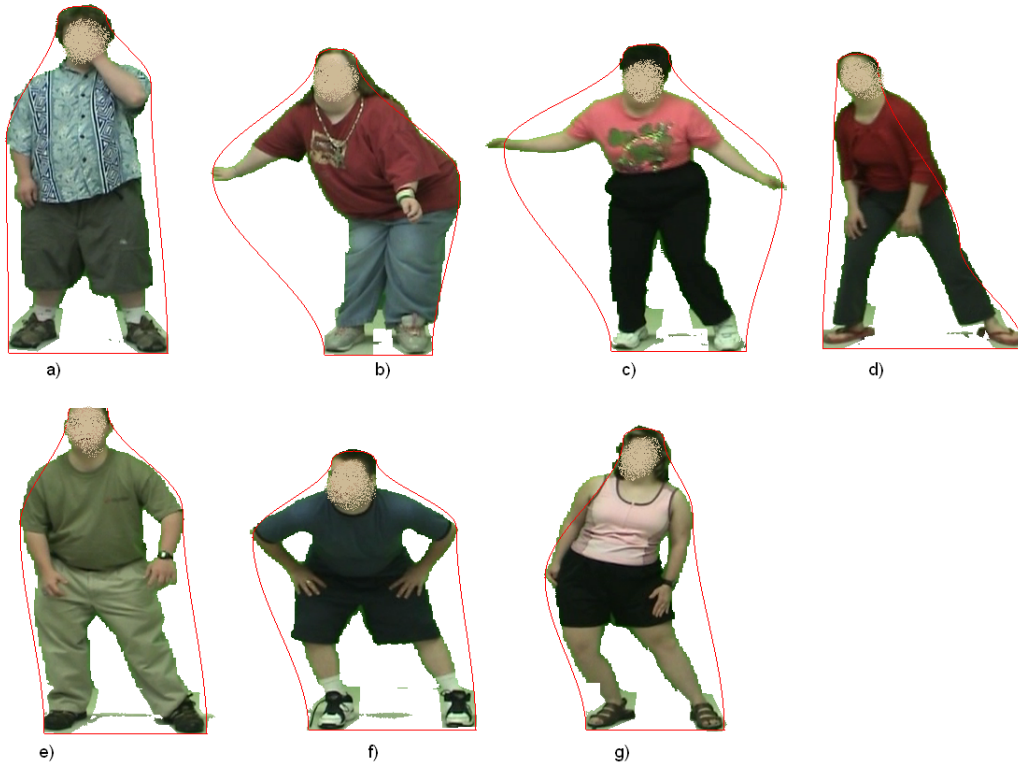


Figure 3.3: Examples of bounding boxes around all the subjects. Several different motions are represented here. a) Standing. b) Bending over to the left with arm out. c) Both arms out. d) Leaning left with arms in. e) Leaning left with left arm in and right arm partially out. f) Bending over forward with hands on hips. g) Moving only the hips to the left.

the arms. This means that when the subjects stick their arms out at length, the bounding box stretches out. Figure 3.4a shows an example of a two spline bounding box where arrows are used to indicate where the two splines are joined. Figure 3.3 shows examples of poses that can be analyzed using the 2 spline bounding box. The contour of the curved bounding box is superimposed on all these poses in order to show which information is preserved about the relative location of the body parts.

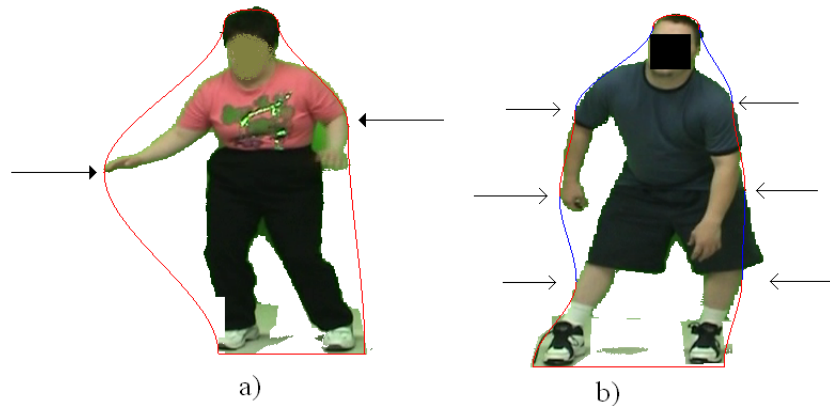


Figure 3.4: The arrows indicate the locations of where splines are joined. a) Two spline bounding box. b) Four spline bounding box.

3.4.2 Four Spline Bounding Box

Bounding boxes consisting of two splines do not convey information about the knee motion. Thus a different style of bounding box was used. This new bounding box used 4 cubic splines, which have the same height on each side (see figure 3.5).

The four spline bounding box is a particular result of a generic method. This method allows for creating bounding boxes with an arbitrary number of splines per sides, by keeping the same vertical height for each spline. It was decided to implement this method with four splines since knees are roughly located at 25% of the body height. In each frame, the vertical height of every spline is the same. This is done by splitting the subject into equal sections by height. One spline is then fitted on each side for each section. The outermost points at the height where the sections meet is where the bounding boxes join. Figure 3.4b shows the locations where the splines are joined. One may note that the splines are joined at the same height on both sides.

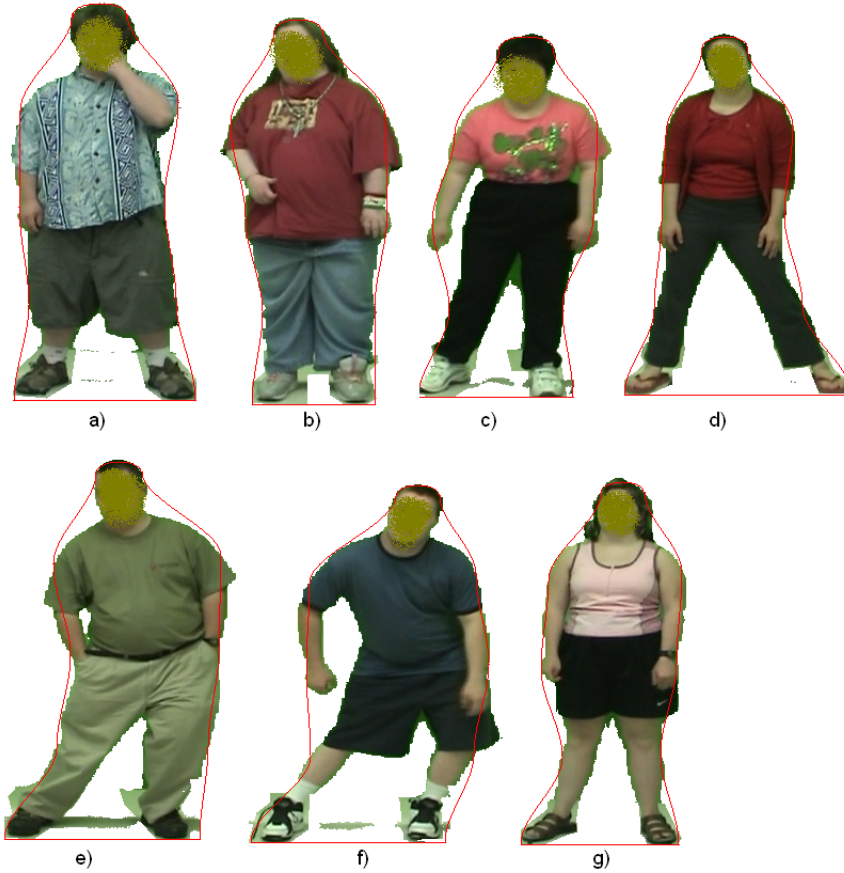


Figure 3.5: An example of the bounding box which uses 4 splines on each side instead of 2.

3.4.3 Parabolic Bounding Box

Unlike spline-based bounding boxes, the parabolic bounding box aims at capturing information about the global curvature of the human body in motion. This information is necessary for detecting abnormal motions such as trunk-based overcompensation (weight-shifting from the trunk as opposed from the legs) and knee hyper-extension. An example of a knee hyper-extension is shown in figure 3.6b. One may observe from figure 3.6b that a concave parabola can be associated to abnormal motion, while a convex one is likely to correspond to normal motion.

One method which is able to detect the concavity is to curve fit a concave shape onto the side of the subject. In the normal upright position, the parabolas will both be convex as one can see from figure 3.6a. When the subject leans to one side by overcompensating with the trunk the parabolas become concave as can be seen in figure 3.6b. In this study, a parabola was chosen since it is the most simple parametric curve. Parabolas are fitted onto both sides using the least squares algorithm (see appendix B). An example of a fitted parabola can be seen in figure 3.6b. In this case, the functions are plotted as $x(y)$ where y is the vertical axis and x is the horizontal axis. For convenience, the coefficients used for the parabola will be a , b , and c such that:

$$f(y) = ay^2 + by + c \quad (3.3)$$

Once the parabola has been fitted onto the side of the subject, the curvature of the parabola is analyzed as shown in section 3.5.2.

3.5 Detection of Arm Motions

Arm motions stretch the bounding box as shown in figure 3.3c. To describe arm motions using the 2 spline bounding box we use a radial representation of the bounding box.

This representation is created by traveling around the bounding box and by recording the polar coordinates of each pixel in the bounding box contour. The radial values are taken relative to the centroid of the subject and the angles are relative to the pos-

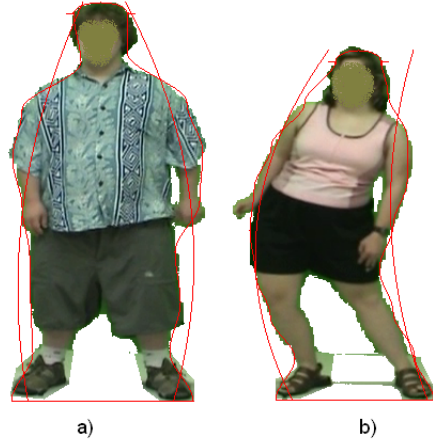


Figure 3.6: a) The first subject standing upright with the parabolas shown. In this case the parabolas are both convex. b) An example of a person over-extending their left leg. In this case the parabola on the right is concave.

itive x axis. An image showing how the silhouette values (r, θ) are obtained can be seen in figure 3.7. A plot of these points can be seen in figure 3.8. Collecting these points allows one to observe changes of the bounding boxes. For example, to observe the arm motions, one would look at the points around 0 degrees and around 180 degrees.

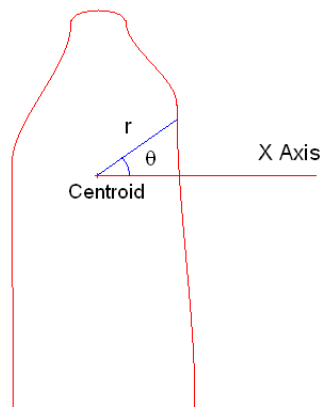


Figure 3.7: The contour of subject showing how the points are taken.

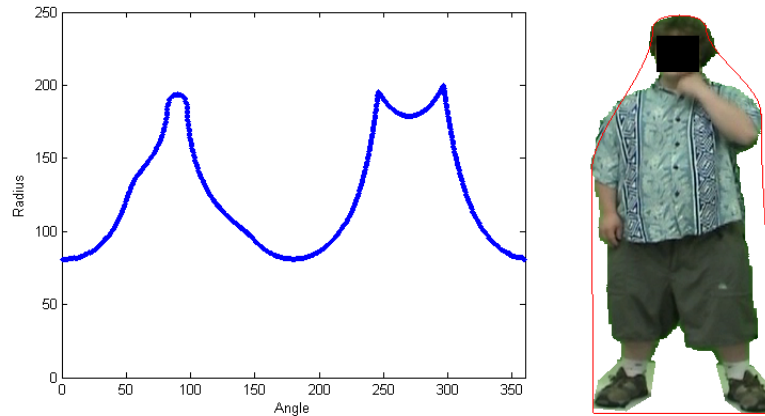


Figure 3.8: The radial coordinates of the bounding box relative to the subjects centroid when the subjects arms are at their side. The pose which generated this plot is also presented.

The radial contour graph in figure 3.8 has three peaks. The first peak from the left corresponds to the head and the two peaks to the right correspond to the feet. This graph corresponds to the same subject seen in figure 3.3a. There are no peaks around 0 degrees or 180 degrees as the subject had his arms held in. When a subject has their arms out, two more peaks can be observed as in figure 3.9.

For detecting arm motions from the polar representation of the bounding box, we adopt a threshold-based approach. Points between 330 degrees and 20 degrees were considered for the right arm and points between 130 degrees and 180 degrees were considered for the left arm. These degree ranges were split up into 10 degree intervals. Within these intervals the average values of the points was taken. If the average value was greater than the upright frame by at least 10 percent it was assumed that the subject was moving their arms. This method is represented mathematically by formula 3.4.

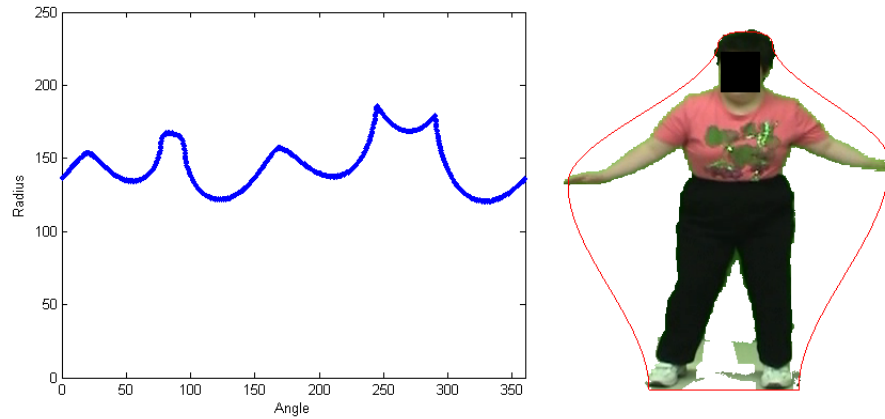


Figure 3.9: The radial coordinates of the bounding box relative to the subjects centroid when the subjects arms are reaching out. The pose which generated this plot is also presented.

$R_s(\theta)$ = average radial distance of the upright frame.

$R_c(\theta)$ = average radial distance of the current frame.

$$\text{If } R_c(\theta) > 1.10 * R_s(\theta) \text{ then arm motion detected.} \quad (3.4)$$

3.5.1 Detection of Knee Leans

The process of detecting knee leans was performed by analyzing the slope of the bottom spline. The slope of the lowest spline at a given frame was compared to the slope of the lowest spline of the reference frame.

3.5.2 Detection of Hip and Upper Body Motions

Once the parabolas were fitted onto the subjects, the parabola coefficients can be analyzed. The most important coefficient in equation 3.3 is a . This value is the parameter that determines the concavity of the parabola.



Figure 3.10: An example of a knee lean being performed by subject 5

The values of a at each frame were binned and a histogram was produced. Before any results can be obtained, a bin size must be chosen for the histogram. If a small bin size is chosen there will be more noise in the resulting histogram. If a large bin size is chosen too much information will be lost. The chosen bin size is 25 times the subjects height.

Since the subjects each performed many actions the results are a superposition of several motions performed with different body parts, and therefore they are hard to interpret. To facilitate their interpretation, a baseline was created. This baseline consists of two typical subjects performing one type of motion per video sequence. Examples of a histograms can be observed in figures 3.11a and 3.12. This example is actually for a very simple motion, standing in one place and swaying a little to each side. An example frame which generated the histogram in 3.11a is visible in figure 3.11b.

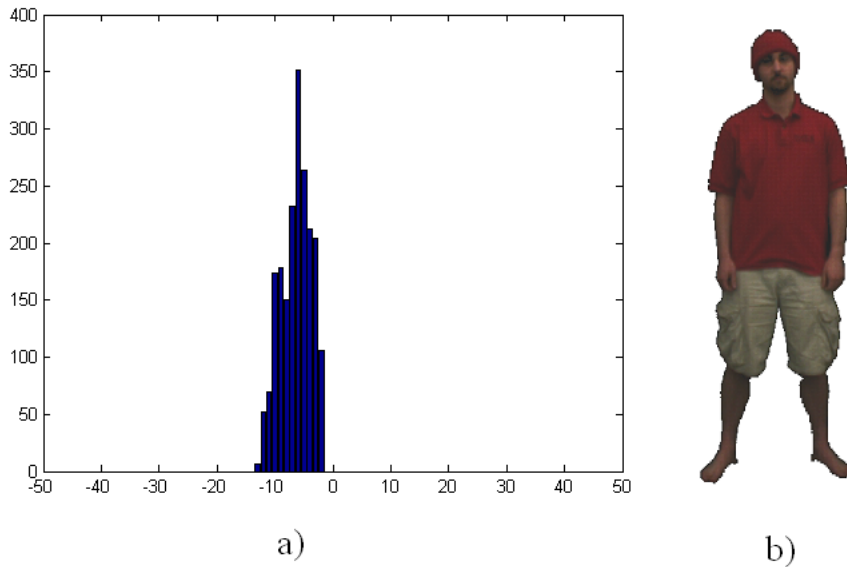


Figure 3.11: a) An example of a histogram generated by analysis of the curvature. For the histogram only slight leans occurred. b) Key frame of the motion which generated this histogram.

3.6 Identification of Specific Patterns of Motion

The detection of arm, knee, and upper body motions from curved bounding boxes provides a detailed account of what motions have occurred on a frame by frame basis. However, this is not sufficient for discovering the distinctive patterns of motions that characterize each subject. Therefore we need a method to integrate information about motions performed with specific body parts into a spatio-temporal template suitable for both quantitative evaluation and visualization purposes. Our method is based upon motion history images, which offer both integrative power and visualization capabilities.

For a given subject performing a sequence of actions in response to the visual stimulus of the virtual reality game, actions are first segmented using upright poses

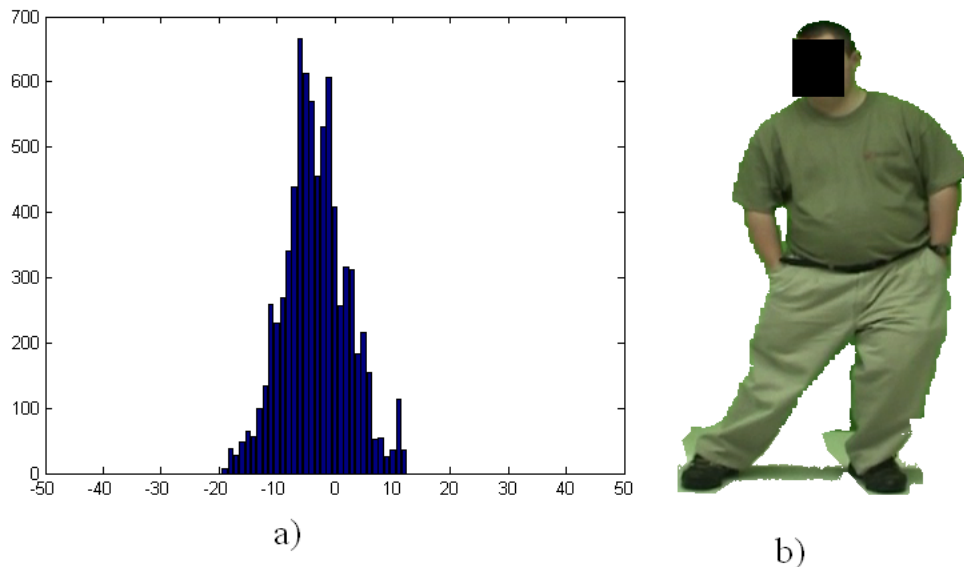


Figure 3.12: a) A second example of a histogram. b) The subject which the histogram represents.

as action separators. Next, a motion history image is created for each action by considering all the frames composing it, as follows:

$$MHI(x, y) = \begin{cases} C(t_i) & \text{if } D(x, y, t_i) = 1 \text{ AND } \left\{ \begin{array}{l} D(x-1, y, t_i) = 0 \\ \text{OR } D(x+1, y, t_i) = 0 \\ \text{OR } D(x, y-1, t_i) = 0 \\ \text{OR } D(x, y+1, t_i) = 0 \end{array} \right\} \text{ for } t_i = t_s \dots t_e \\ 0 & \text{elsewhere} \end{cases} \quad (3.5)$$

Where $D(x, y, t_i)$ is the spatial occupancy function and $C(t_i)$ is the colour allocated to the contour of the subject at frame i . Considering t_s and t_e , the start and end frames of a given action, the motion history image is built by stacking all contours of the silhouette performing the action represented in pseudo-colour mode. The action

evolution is captured by the colour transition specified in equation 3.6.

$$\{C(t_i) = \{C : \{t_s, \dots, t_e\} \rightarrow \{645nm \dots 380nm\}\} \quad (3.6)$$

Among the motion history images that are built for every action, the most representative template needs to be chosen. This choice was performed according to each of the two hypothesis below:

1. The most representative action is the one performed most often.
2. The most representative action is the one taking the maximum number of frames (on a cumulative basis, by counting all instances of the action).

In addition to finding the most representative action, the motion history images that represent the same action at different time instances during the video sequence need to be grouped together. This is performed by measuring the displacement of the subject at different heights during the current action. Each subject is segmented into upper, middle, and lower sections as shown in 3.13 and the global displacements are computed for each section. A group of actions is defined as having the most displacements concentrated in one of the sections (upper, middle, and lower). Thus, three groups of actions are formed for each side of the body, as well as for actions involving both sides. In total, there are nine groups of actions that are representable via MHIs.

Figure 3.14 shows several examples of representative motion history images which were created for each subject.

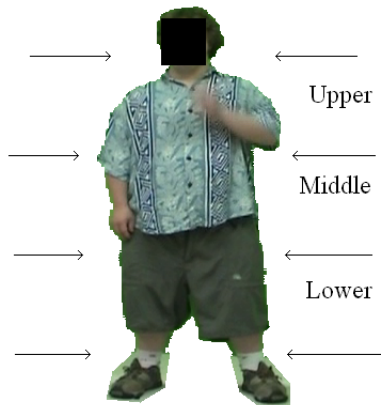


Figure 3.13: The arrows indicate the boundaries of the upper, middle, and lower divisions for the motion history images.

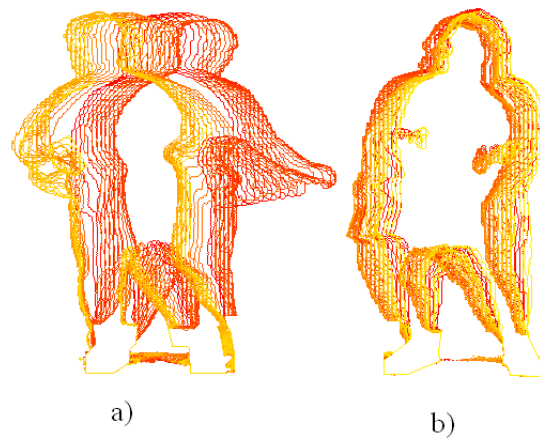


Figure 3.14: a) Motion history image of subject 3 putting her arm out and leaning to the other side. b) Motion history image of subject 7 performing a pronounced lean to the left side.

Chapter 4

Results

This chapter starts by describing the experimental database in section 4.1 and the running time. Next, it presents the validation of the motion analysis algorithms based on bounding box information (section 4.3). The last section presents the subject-representative patterns of motion that were generated by using motion history images.

4.1 Experimental Database

The database was provided by the Down's syndrome Research Foundation in Vancouver. It consists of seven different subjects who performed activities while playing a virtual reality game. The game consisted in controlling a virtual snowboard by lateral weight shifting motions in order to navigate through obstacles. The obstacle course, as well as the avatar representing the subjects location were displayed on a screen located in front of the subject (see figure 4.1b). The game content was identical for all subjects, in order to be able to compare their motions. All motions were performed by the subjects in a standing position with their feet stationary. They were facing the camera so that a frontal view of their motions was acquired as shown in figure

4.1a.

In summary, the database for this study consists of seven video sequences (one per subject) each with a length of 2 minutes and 7 seconds. The videos were filmed at a resolution of 480 by 720 pixels and a frame rate of 30 frames per second.



a)



b)

Figure 4.1: a) The first frame of the video sequences of the subjects with Down's syndrome. b) An example frame from the video sequence watched by the subjects.

4.2 Running Time

One of the advantages of these algorithms over others is that they are relatively fast. On a fast enough computer and a low enough resolution they can be run in real time. This is for a resolution of 720 by 480 and 30 frames per second. The analysis for this project was done on videos and so the remainder of this section assumes that the algorithm is running in this manner.

The running time of all the algorithms is linear with the resolution and total number of frames to analyze. The program however relies more on the hard disk speed than the processor speed. This is because the algorithms are reading and writing large amounts of raw image data to disk.

In order to calculate the running time, the actual running time for each algorithm was found. The testing was done on a Dell Optiplex 745 with a Intel Core 2 6600 CPU. The results are found in table 4.1. These were calculated by timing the algorithm for the entire dataset and then dividing by the total frames. The time for the parabolic bounding box includes the time it takes for Matlab to plot the histograms. The time for the motion history images includes the time it takes to save each motion history image to disk. From table 4.1, it is apparent that the running time for the bounding box methods is very similar. This shows how important the read and write speed of the hard disk is to running time.

Algorithm	Average Running Time (seconds/frame)
2 Spline BB	0.0638
4 Spline BB	0.0641
Parabolic BB	0.0642
Motion History Images	0.0918

Table 4.1: The measured running times of the different methods.

4.3 Experimental Validation of Motion Detectors Based on Bounding Box Information

The arm, knee, and upper body motion detections described in chapter 3 were validated against ground truth. For the arm and knee, the ground truth was obtained via manual inspection of 500 frames.

4.3.1 Arm Motion Detection Using a Two Spline Bounding Box

Experimental validation results summarized in tables 4.2 and 4.3 show the number of correct detections, false detections, correct rejections, and missed detections for a 500 frame sample for each subject.

There were a few body positions which regularly returned a false result. One position which was performed by more than one subject is a lean to one side and an extension of their arm on the other side. The arm motion is not recognized because the knee lean introduces a motion of the centroid. This inhibits the detection of the arm as a local maxima on the radial contour representation of the silhouette. This type of motion is depicted in figure 4.2. This motion was responsible for most of the missed detections of subjects 5 and 6.

False detections of the arm motion systematically occur when subjects shift their weight by overcompensating with their upper body. A key posture from such a motion is shown in figure 4.2b). The performance of the arm detector is summarized using the following measure of performance:

$$\text{Correct Decisions} = \frac{\text{Correct Detections} + \text{Correct Rejections}}{\text{Total Frames}} \quad (4.1)$$

The performance measures for each subject are shown in table 4.4. Based on these results, we can conclude that the technique proposed for arm motion detection has good performance although this performance varies a lot from one subject to another.

Subject	Correct Detections	False Detections	Correct Rejection	Missed Detections
1	0	0	500	0
2	420	40	0	0
3	390	97	11	2
4	67	149	284	0
5	0	13	484	3
6	36	4	336	124
7	0	272	228	0

Table 4.2: Experimental validation of the arm motion detector for the left side.

Subject	Correct Detections	False Detections	Correct Rejection	Missed Detections
1	0	0	500	0
2	0	405	95	0
3	361	15	113	11
4	93	91	316	0
5	120	20	308	52
6	19	34	425	22
7	0	283	217	0

Table 4.3: Experimental validation of the arm motion detector for the right side.

Subject	1	2	3	4	5	6	7	All Subjects
Correct Decisions Left Side (%)	100	84.0	80.2	70.2	96.8	74.4	45.6	78.7
Correct Decisions Right Side (%)	100	19.0	94.8	81.8	85.6	88.8	43.4	73.3

Table 4.4: Experimental validation of the arm motion detector.

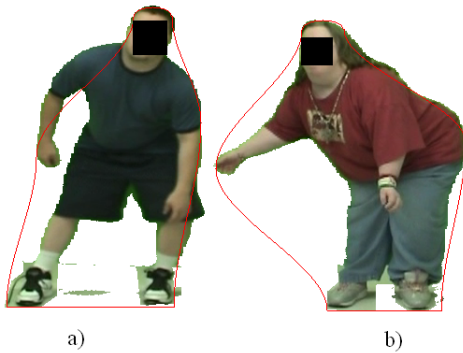


Figure 4.2: a) frame where a missed detection of the arm occurs; b) frame where a false detection of the arm occurs.

4.3.2 Knee Motion Detection

Table 4.5 summarizes the experimental validation of the knee motion detector. One problem which can cause errors is when the subject moves their feet, which contradicts one of our basic assumptions. This was the case of subject 3, who did not obey game playing instructions. If a subject moves their feet their center of balance is shifted. This change was not considered in either the background subtraction or this algorithm. The performance measures for each subject were measured with the same algorithm as that for the arm motions (equation 4.1) and are shown in table 4.7. Based on these results, we can conclude that the technique proposed for knee motion detection has good performance.

Subject	Correct Detections	False Detections	Correct Rejection	Missed Detections
1	75	0	425	0
3	113	10	304	83
4	17	0	431	52
5	57	0	389	27
6	204	2	289	5
7	188	47	265	0

Table 4.5: Experimental validation for the knee lean motion detector for the left side.

Subject	Correct Detections	False Detections	Correct Rejection	Missed Detections
1	58	34	408	0
3	98	45	336	21
4	135	79	286	0
5	280	11	209	0
6	227	45	228	0
7	195	12	291	2

Table 4.6: Experimental validation for the knee lean motion detector for the right side.

4.3.3 Hip and Upper Body Motion Analysis

Since the analysis of the upper body motion goes beyond the detection of motion compensation patterns, the generation of ground truth for calibration and validation purposes can not be done via manual labeling of frames. The ground truth that was generated consists in reference histograms. These histograms were created for three specific types of motions which were performed by two typical subjects. These

Subject	1	3	4	5	6	7	All Subjects
Correct Decisions Left Side (%)	100	83.4	89.6	89.2	98.6	90.6	91.6
Correct Decisions Right Side (%)	93.2	86.0	84.2	97.8	91.0	97.2	91.6

Table 4.7: Experimental validation of the knee motion.

specific motions were chosen in order to reflect the motions that were performed by Down syndrome subjects. This way, one can visualize the effect of a single typical motion on its corresponding histogram, and can further the effect of multiple different motions on their combined histogram. The three types of specific motions are as follows:

1. Slight lateral weight-shifting patterns.
2. Pronounced lateral weight-shifting patterns with upper body compensation.
3. Arm motions.

For the first type of motion, involving only slight leans, the curvature of the silhouette presents only slight variations ± 10 . Therefore, one can expect the corresponding histogram to be narrow and centered on the -10 . This is confirmed by our measurements, which are shown in figure 4.3.

The second type of simulated motions involves sharp leans of the lower body on one side simultaneous with leans of the upper body on the other side for overcompensation. This motion was selected for simulation since it was very frequently performed by DS subjects. The lateral concavity of the silhouette measured by parabola fitting was very pronounced in this type of motion. One can observe in the corresponding histograms (figure 4.4) that the peak is centered off the origin.

The third type of simulated motion consisted in arm motions from a quasi-upright stance. As with the previous two types, this motion was chosen since it was found as being frequently performed by DS subjects. The corresponding histograms for this

type of motion are shown in figure 4.5. One may note that these histograms have peaks centered close to zero, and are slightly broader than the histograms for the slight leans (type 1).

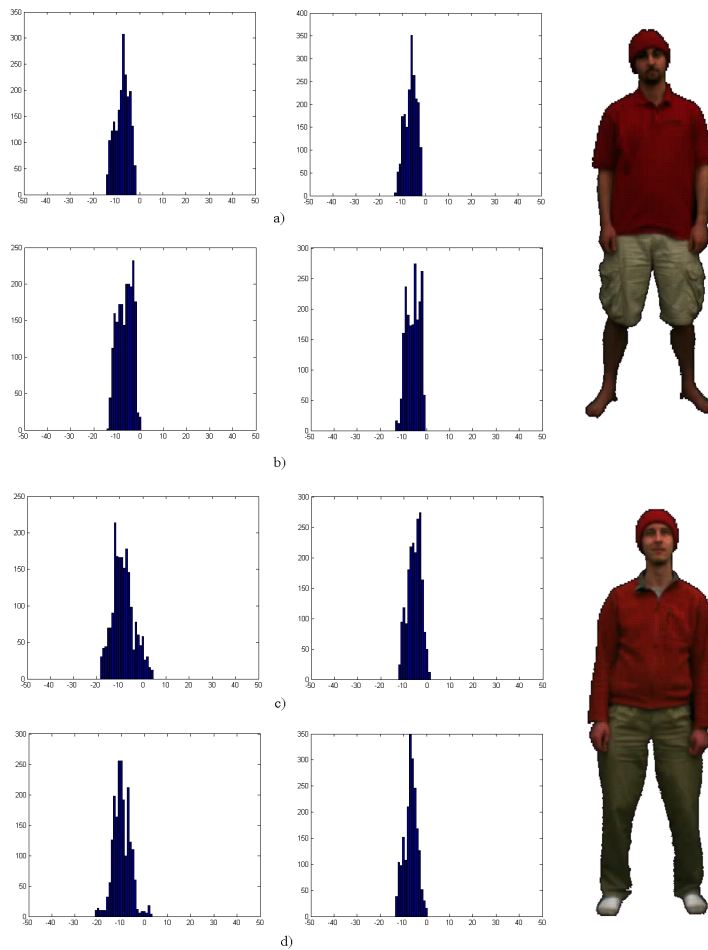


Figure 4.3: Histograms representing typical subjects standing upright with slight leans to either side. All the histograms in the left column correspond to the left side of a subject and all the histograms in the right column correspond to the right side of the side of a subject. Rows a) and b) correspond to the top subject and rows c) and d) correspond to the bottom subject.

Histograms Describing Motions Performed by DS Subjects

The histograms of the subjects with Down's syndrome are generated for the entire corresponding video sequence. This implies that the histogram integrates information about all the motions performed by the subject. It is therefore more difficult to interpret these histograms than the histograms of (much simpler) simulated motions. However, one may use the appearance of the histograms for simulated motions in order to detect whether upper body motion occurred as compensatory motion. These histograms can be seen in figures 4.6 and 4.7.

From figures 4.6 and 4.7 it is apparent that many of the Down's syndrome subjects performed compensation motions. The subjects have peaks which occur at locations other than the upright peak.

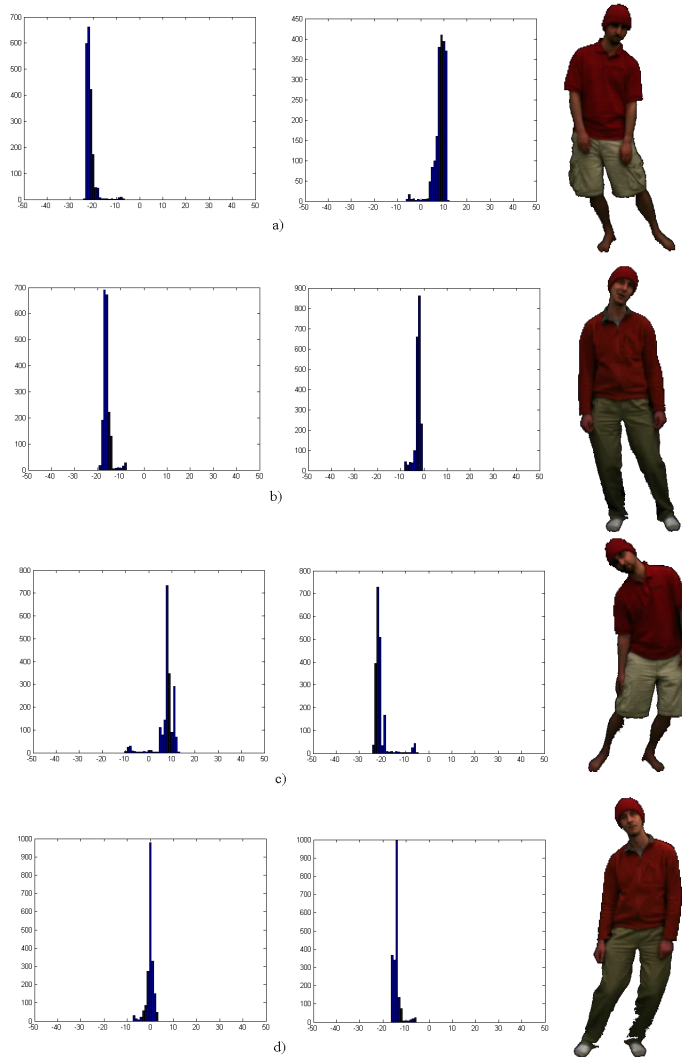


Figure 4.4: Histograms representing typical subjects leaning sharply to one side and compensating by leaning their upper body in the opposite direction. These were performed by typical subjects for 40 seconds. The left column corresponds to the left side of the subject and the right column corresponds to the right side. Rows a) and c) are the same subject leaning left and then right and rows b) and d) are a different subject leaning left then right.

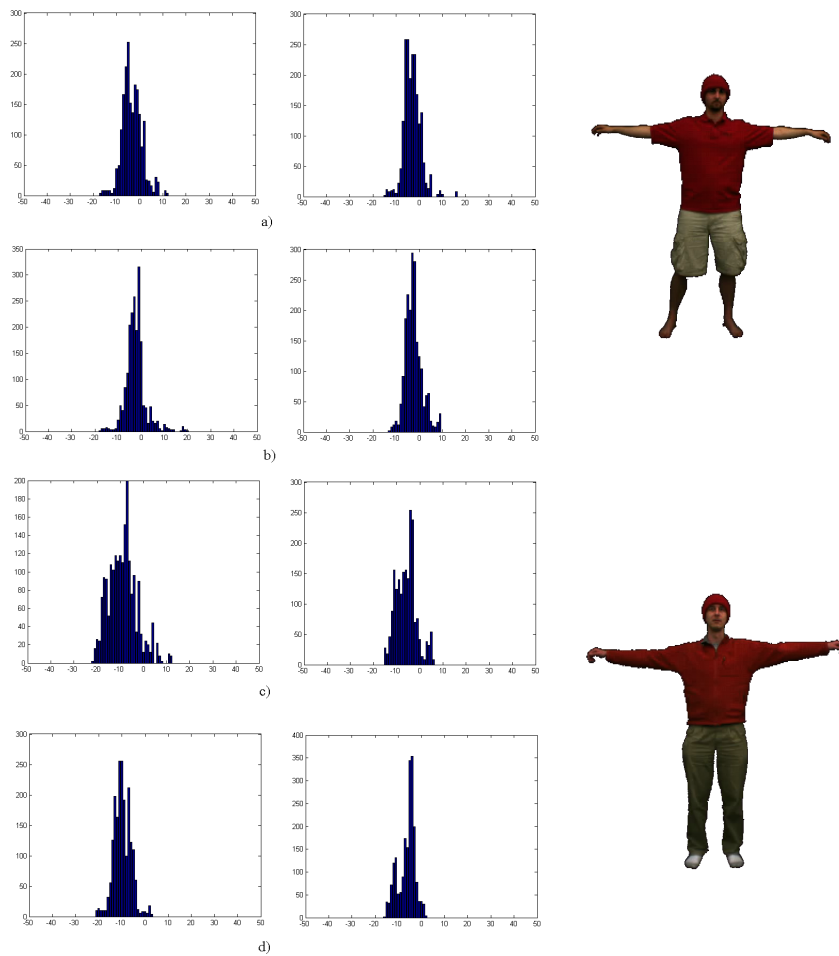


Figure 4.5: Histograms representing typical subjects performing slight leans with their arms out at either side. The left column is the left side of the subject and the right column is the right side. Rows a) and b) correspond to sequences by the top subject and rows c) and d) correspond to sequences performed by the bottom subject.

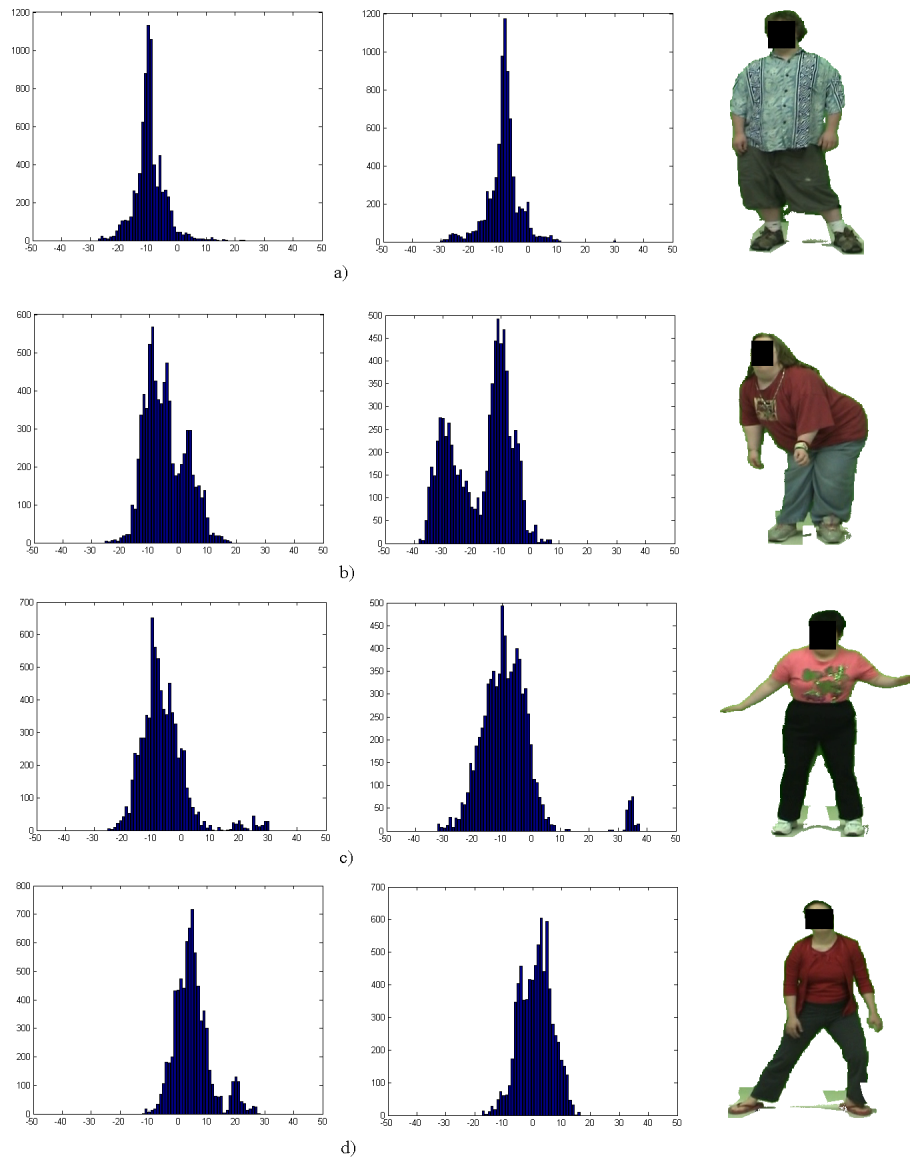


Figure 4.6: The histograms for subjects 1 to 4. The left column is curvature of the left side of the subject and the right column is the right side of the subject. At the end of each row the subject responsible for each set of histograms is shown in a typical pose.

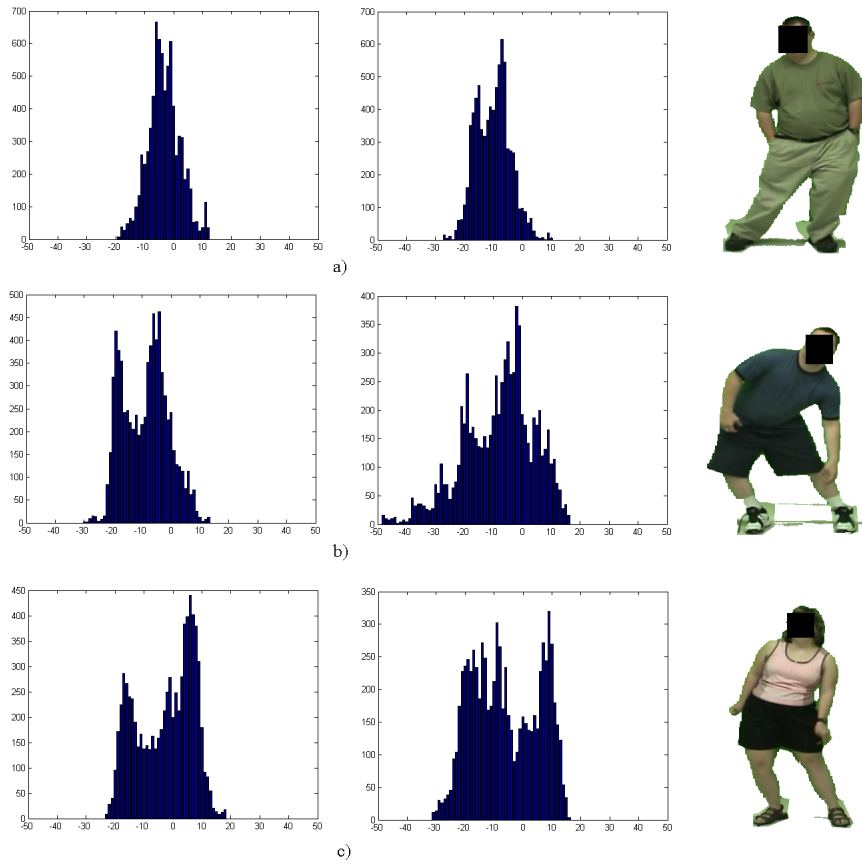


Figure 4.7: The histograms for subjects 5 through 7. The left column is curvature of the left side of the subject and the right column is the right side of the subject. At the end of each row the subject responsible for each set of histograms is shown in a typical pose.

4.4 Identification of Specific Patterns of Motion

As described in Chapter 3, all motions performed by each subject were segmented first, then described and visualized via motion history images. The purpose of this approach was to determine a specific pattern of motion for each subject.

For a given subject, his/her motion history images were classified according to the height they occurred. Next, the specific pattern of motion corresponding to that subject was set as the motion history image that represents the most common motion.

Two different ways of defining the most common motion were investigated. The first one defines the most common motion as the one with most occurrences in the subject-specific sequence (regardless of the duration of this occurrence). The second way defines the most common motion as the motion with the longest cumulative time duration (by summing the duration of all its occurrences throughout the video sequence).

The remainder of this section overviews the specific patterns of motion that were found for every subject.

Subject 1

Subject 1 performed 43 different actions, most of which were lower level. The most frequent motions are left and right middle section at 9 and 8 motions respectively. The longest motions were lower body motions as over 50% of his time was spent performing these motions. Key-frames representing the most common motions (from both frequency of occurrence and total duration standpoints), as well as the corresponding

patterns of motion are represented in figure 4.8.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	4	5.00	2	2.40	1	2.20
Middle Level	9	12.93	6	14.43	8	13.67
Lower Level	4	23.13	3	18.80	6	26.63

Table 4.8: Motion counts and times for subject 1.

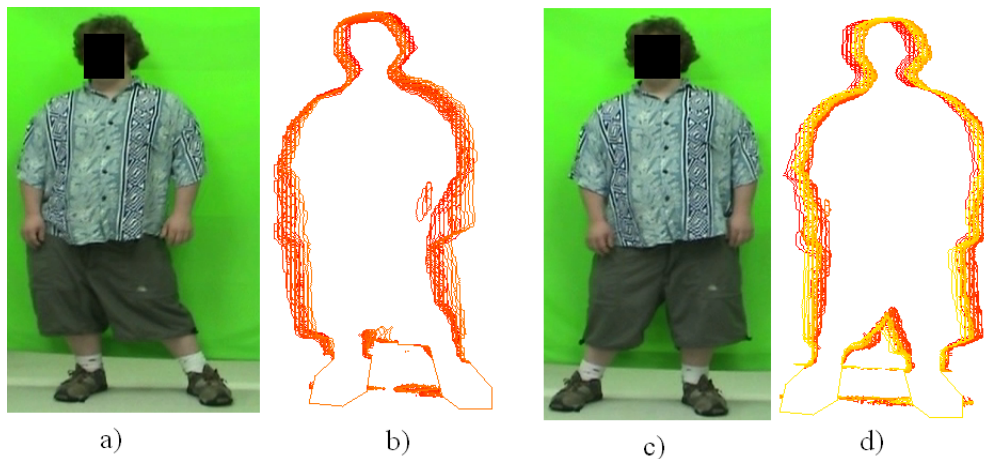


Figure 4.8: a) An example of the most frequent by subject 1. b) The motion history image corresponding to the motion seen in a). This motion is a mid level motion on the left side. c) An example of the most common motion of subject 1 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is a lower motion using the right side.

Subject 2

Subject 2 performed 29 different actions, most of these were middle level. This subject also had the most motions around the middle section. Although the most frequent motion is the on the right, she spent nearly 45% of her time performing motions in the upper level and both sides. One can see from the histogram in figure 4.9b

and recognizing that it is a motion to the right that this subject was performing a lot of compensation motions. An example of the most common motions and their corresponding motion history images is found in figure 4.9.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	0	0	6	56.83	0	0
Middle Level	5	7.27	6	20.73	12	13.93
Lower Level	0	0	0	0	0	0

Table 4.9: Motion counts and times for subject 2.

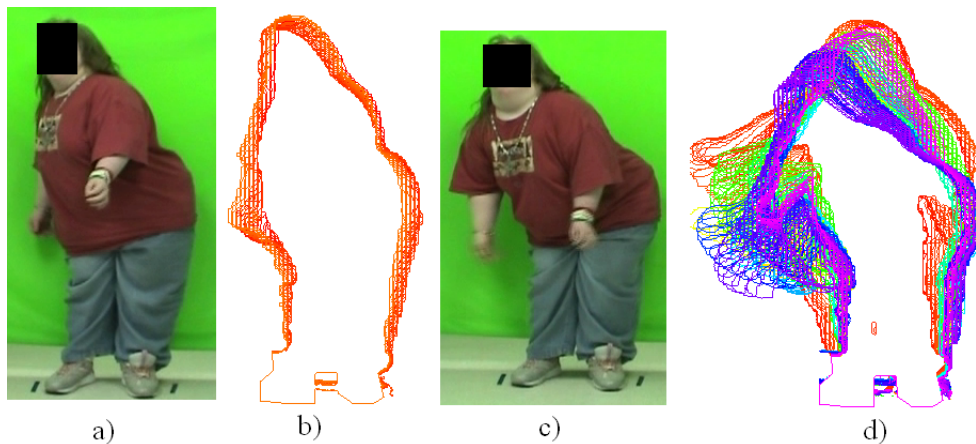


Figure 4.9: a) An example of the most common motion by times performed by subject 2. b) The motion history image corresponding to the motion seen in a). This motion is a middle level motion on the right side. c) An example of the most common motion of subject 2 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is an upper level motion using the right side.

Subject 3

Subject 3 performed 64 different motions, most of which occurred at the middle level.

This subjects most frequent motions were in the middle level and the right lower

level. An unusual observation is that there were a lot more lower body leans on the right side than the left side. The longest motions for this subject occurred at the in the middle level and used both sides. These motions took 44% of the subjects time.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	1	1.13	2	19.43	0	0
Middle Level	14	9.67	17	56.17	10	9.30
Lower Level	4	2.60	1	0.53	15	13.40

Table 4.10: Motion counts and times for subject 3.

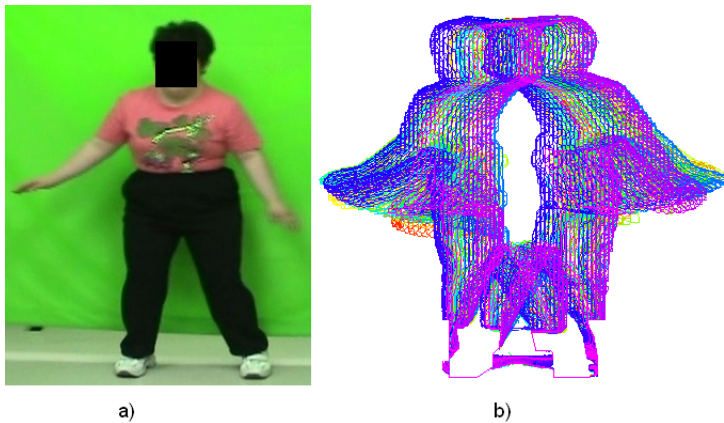


Figure 4.10: a) An example of the most common performed by subject 3. This frame was one of the frames used to generate the motion history image seen in b). This is a middle level motion using both sides.

Subject 4

Subject 4 performed 87 different motions, most of which were lower level. Due to the unusual stance, the majority of the motions occurred in the lower body region. This subjects spent 61% of her time performing these motions. Unlike subjects 2 and 3,

this subject performed many small motions and no long motions. This subject had more motions on their right side which is an indication that they may favor their left side. An example of the most common motion can be found in figure 4.11.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	1	0.37	0	0	1	1.43
Middle Level	3	1.63	3	7.63	1	0.70
Lower Level	30	19.0	4	5.33	44	53.03

Table 4.11: Motion counts and times for subject 4.

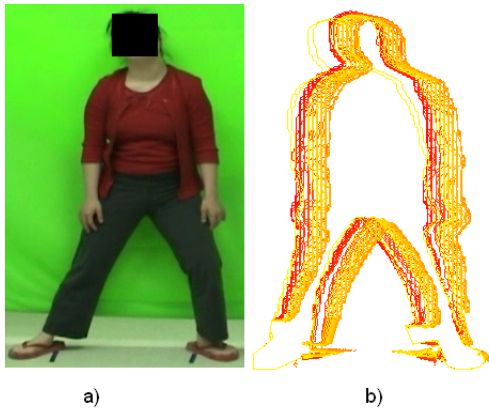


Figure 4.11: a) An example of the most common motion by frequency and time length by subject 4. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a lower level motion using the left side.

Subject 5

Subject 5 performed 34 different motions during their sequence. This subject really favored the right side (his left side) as there were a lot more motions on the right side than on the left side. This subject spent more time performing motions on the right

side than all other motions combined. Although it is not visible in table 4.12, this subject had three motions on the right side over 10 seconds in length.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	0	0	1	0.33	2	10.27
Middle Level	2	3.37	3	5.57	13	57.07
Lower Level	4	9.50	4	6.93	4	5.83

Table 4.12: Motion counts and times for subject 5.

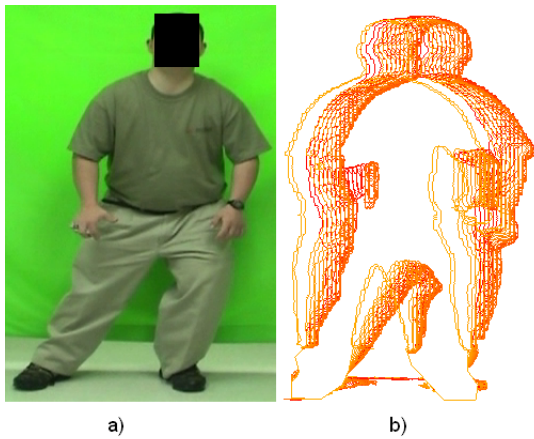


Figure 4.12: a) An example of the most frequent motions and the most time consuming motions performed by subject 5. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a middle level motion using the right side.

Subject 6

Subject 6 had only 5 different motions recorded. This is because the motions performed were very long motions. The most frequent motions for this subject were middle level motions to the right side. This subject however spent far more time

performing motions which used both sides and had the highest displacement at the middle level. This subject performed middle level motions which were 27 seconds and 75 seconds in length.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	0	0	0	0	0	0
Middle Level	0	0	2	101.47	3	6.37
Lower Level	0	0	0	0	0	0

Table 4.13: Motion counts and times for subject 6.

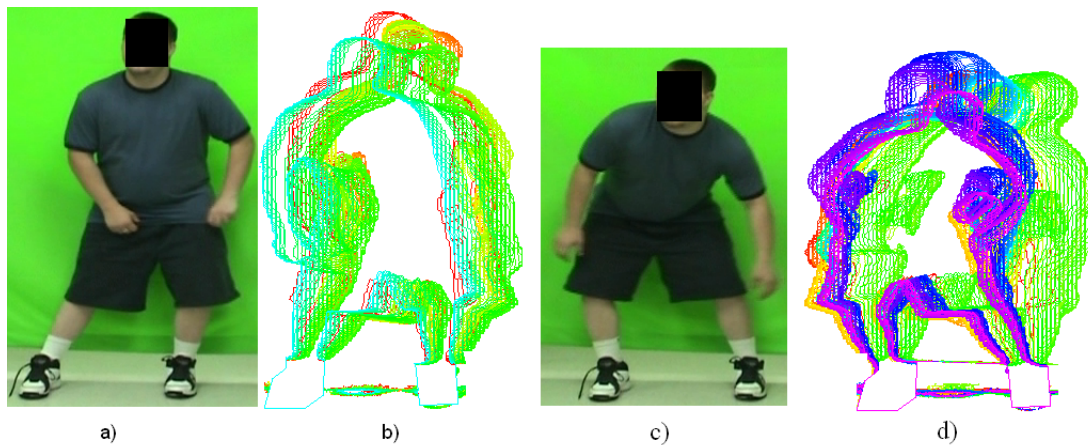


Figure 4.13: a) An example of the most common motion by times performed by subject 6. b) The motion history image corresponding to the motion seen in a). This motion is a middle level motion on the right side. c) An example of the most common motion of subject 6 measured by time. d) The motion history image corresponding to the motion seen in c). This motion is a middle level motion using both sides.

Subject 7

Subject 7 had 43 motions, most of which were lower level motions. All of the motions used the lower level as one will notice in table 4.14. This subject spent 78% of her time performing leans at the lower level. Most of these motions only used one side.

	Left Side		Both Sides		Right Side	
	Number	Time (s)	Number	Time (s)	Number	Time (s)
Upper Level	0	0	0	0	0	0
Middle Level	0	0	0	0	0	0
Lower Level	20	41.43	6	20.93	17	37.30

Table 4.14: Motion count for subject 7.

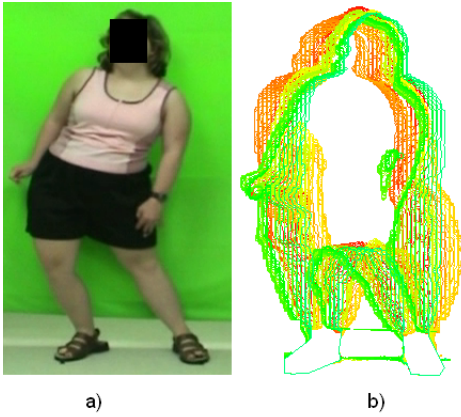


Figure 4.14: a) An example of the most frequent motions and the longest motions performed by subject 7. This frame was one of the frames used to generate the motion history image seen in b. This motion primarily a lower level motion using the left side.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis proposes a set of methods for the analysis of motor skills in DS subjects. More specifically, the proposed techniques were developed for the study of weight shifting patterns of motion. This type of motion is an essential component of locomotory activities, as well as of other daily actions.

The experimental database consists of a set of subject-specific video sequences acquired while subjects were playing a virtual reality game involving weight shifting motions.

From a theoretical viewpoint, this thesis proposes several new concepts and their corresponding algorithmic development. These are summarized below.

Curved bounding boxes provide new tools for representing and analyzing human motion. This thesis presents and demonstrates an arm motion detector based on the 2 spline bounding box, a knee motion detector based on the 4 spline bounding box, as well as a hip and upper body motion detector based on the parabolic bounding

box.

Next, subject-specific patterns of motion are determined from motion history images which allows for clustering of MHI that represent similar motions and for visualizing the most representative motion for a given subject. This concept is used to integrate both analytical and visualization components, which is one of its strengths.

5.2 Future Work Directions

Future work will focus on relaxing a few of the initial assumptions about the analyzed motion. This will be valuable for extending the application of the proposed methods towards the general field of motor rehabilitation. In particular, the static feet assumption needs to be relaxed so that motions other than weight shifting can be considered. This relaxation is possible via a more sophisticated background subtraction method. The experimental validation protocol also needs to be improved so that larger databases can be used. The manual ground truth method is a reliable basis for validation, however it is very difficult and time consuming to do this on a frame by frame basis. Therefore, marker-based validation protocols will be considered in future experiments.

Although the proposed algorithms were designed to observe the motions of subjects with Down's Syndrome, they are not limited to this use. They can also measure the motions of people with limited mobility or normal mobility.

The proposed methods could be taken further by using them to analyze motions other than full body motion. This would be articulated motion analysis. Thus, any-

thing that moves and has joints, could potentially be analyzed with these algorithms.

Appendix A

The Natural Cubic Spline Function

In this research, natural cubic spline functions were used in the formation of the bounding box. A natural cubic spline is different from an ordinary cubic spline function in that the slope at the end points has been set to zero. Setting the slope to zero forces the curve as well as its first derivative to be continuous where the two curves join. An example of a natural cubic spline can be seen in figure A.1. In order to plot the cubic spline one must evaluate equation A.1 for known end points (x_1, y_1) and (x_2, y_2) . In order to evaluate equation A.1, one must first evaluate formulas A.2 through A.5. The formulas are the standard cubic spline formulas except that x and y have been switched.

There are several reasons why natural cubic splines were chosen to form the bounding box. The first reason is that an ordinary cubic spline function would not work. Since only two points were used to create each spline, and there is four unknown variables (see equation A.1), there is not enough information to get only one solution. Since a cubic spline has the condition that its derivative must be zero at the end points, there is now two equations and two points which makes four unknowns and

four equations, enough to get a one and only one solution. Another reason why cubic splines were used is because they are continuous at the intersection point. A cubic spline is the lowest order polynomial in which can be made so that the end points have a slope of zero¹. Higher order polynomials can have a slope of zero at the end points, but they also have an infinite number of solutions. This means that the Natural Cubic Spline function is the lowest order polynomial which can have a slope of zero at the end points and the only polynomial which can be zero at the end points and have only one solution.

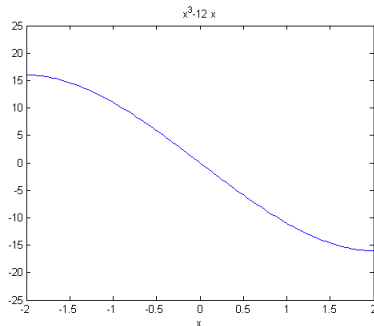


Figure A.1: An example of a cubic spline function.

$$x = ay^3 + by^2 + cy + d \quad (\text{A.1})$$

$$a = \frac{2(x_2 - x_1)}{y_1^3 - 3y_1^2y_2 + 3y_1y_2^2 - y_2^3} \quad (\text{A.2})$$

$$b = \frac{-3a(y_2 + y_1)}{2} \quad (\text{A.3})$$

$$c = 3ay_2y_1 \quad (\text{A.4})$$

$$d = x_1 - ay_1^3 - by_1^2 - cy \quad (\text{A.5})$$

¹For arbitrary end points

Appendix B

The Linear Least Squares Method for a Parabola

The first step is to find where equation B.1 is a minimum. In this case N is the number of points being used to fit the parabola. The result is long and so in order to simplify the math, a few variables were created. Recall that the objective is to find the values of a , b , and c . The first step is to find values for n , m , p , q , r , t , and u using equation B.2. In this derivation we are using $x(y)$.

$$\sum_{i=0}^N (x_i - c - by_i - ay_i^2) \quad (\text{B.1})$$

$$m = \sum_{i=0}^N y_i^2 \quad n = \sum_{i=0}^N y_i^3 \quad p = \sum_{i=0}^N y_i \quad q = \sum_{i=0}^N y_i^4 \quad r = \sum_{i=0}^N x_i y_i^2 \quad t = \sum_{i=0}^N x_i \quad u = \sum_{i=0}^N x_i y_i^2 \quad (\text{B.2})$$

Once the values have been found using equation B.2 one can then calculate the values for d and w . These can be found by evaluating equations B.3 and B.4. Once

d and w have been solved for, one can evaluate a, b, and c using equations B.5, B.6, and B.7.

$$d = t - \frac{um}{q} + \left(\frac{un - rq}{n^2 - mq}\right)\left(\frac{nm}{q} - p\right) \quad (\text{B.3})$$

$$w = \frac{m^2}{q} - N - \left(\frac{nm - pq}{n - mq}\right)\left(\frac{mn}{q} - p\right) \quad (\text{B.4})$$

$$a = \frac{u - bn - cm}{q} \quad (\text{B.5})$$

$$b = \frac{un - rq - c(nm - pq)}{n^2 - mq} \quad (\text{B.6})$$

$$c = \frac{-d}{w} \quad (\text{B.7})$$

Bibliography

- [1] T. B. Moeslund, A. Hilton, and V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” *Comput. Vis. Image Underst.*, vol. 104, no. 2, pp. 90–126, 2006.
- [2] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture,” *Computer Vision and Image Understanding*, vol. 81, pp. 231–268, 2001.
- [3] H. SR, “Neuromotor development of infants with down syndrome,” *Dev Med Child Neurol*, 1985.
- [4] H. SE, M. J, and F. U, “The motor deficit of down’s syndrome children: a problem of timing?,” *Journal of Child Psychology and Psychiatry*, 1981.
- [5] S. G, van der Kamp J, L. A, and P. T, “Information-movement coupling in children with down syndrome,” in *Weeks DJ, Chua R, Elliot D, eds. Perceptual-motor behaviour in Down syndrome.*, Human Kinetics Publishers, 2000.

- [6] L. ML, “Motor coordination in down syndrome: the role of adaptive changes.,” in *Weeks DJ, Chua R, Elliot D, eds. Perceptual-motor behaviour in Down syndrome.*, Human Kinetics Publishers, 2000.
- [7] C. Sminchisescu and B. Triggs, “Estimating articulated human motion with covariance scaled sampling,” *International Journal of Robotics Research*, 2003.
- [8] M. Murray, “Gait as a total pattern of movement,” *American Journal of Physical Medicine, No 1*, pp. 290–329, 1967.
- [9] D. A. Forsyth and M. M. Fleck, “Body plans,” *Computer Vision and Pattern Recognition*, June 1997.
- [10] P. Felzenszwalb and D. Huttenlocher, “Efficient matching of pictorial structures,” *Computer Vision and Pattern Recognition*, 2000.
- [11] A. S. Micilotta, E. J. Ong, and R. Bowden, “Detection and tracking of humans by probabilistic body part assembly,” *British Machine Vision Conference*, 2005.
- [12] A. B. Albu, M. Yazdi, and R. Bergevin, “Detection of cyclic human activities based on the morphological analysis of the inter-frame similarity matrix,” *Real-Time Imaging*, vol. 11, pp. 219–232, June 2005.
- [13] A. Ali and J. Aggarwal, “Segmentation and recognition of continuous human activity,” *Workshop on Detection and Recognition of Events in Video*, pp. 28–35, 2001.

- [14] S. Ju, M. Black, and Y. Yacoob, “Cardboard people: a parameterized model of articulated image motion,” *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, pp. 38–44, Oct 1996.
- [15] T. J. Roberts, S. J. McKenna, and I. W. Ricketts, “Human pose estimation using learnt probabilistic region similarities and partial configurations,” *European Conference on Computer Vision*, 2004.
- [16] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Computational Learning Theory: Eurocolt 95*, 1995.
- [17] F. Jean, R. Bergevin, and A. Albu, “Trajectories normalization for viewpoint invariant gait recognition,” *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, Dec. 2008.
- [18] F. Jean, R. Bergevin, and A. Albu, “Body tracking in human walk from monocular video sequences,” *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, pp. 144–151, May 2005.
- [19] A. Bharatkumar, K. Daigle, M. Pandey, Q. Cai, and J. Aggarwal, “Lower limb kinematics of human walking with the medial axis transformation,” *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pp. 70–76, Nov 1994.

- [20] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, “Real-time estimation of human body posture from monocular thermal images,” in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, (Washington, DC, USA), p. 15, IEEE Computer Society, 1997.
- [21] C. Bregler, J. Malik, and K. Pullen, “Twist based acquisition and tracking of animal and human kinematics,” *International Journal of Computer Vision*, 2004.
- [22] G. Cheung, S. Baker, and T. Kanada, “Shape from silhouette for articulated objects and its use for human body kinematics estimation and motion capture,” *Computer Vision and Pattern Recognition*, 2003.
- [23] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, “Human body model acquisition and motion capture using voxel data,” *F.J. Perales and E.R. Hancock (Eds.): AMDO 2002, LNCS 2492*, 2002.
- [24] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, “Human body model acquisition and tracking using voxel data,” *International Journal of Computer Vision*, 2003.
- [25] S. J. Gortler, “Image-based visual hulls,” in *in Proceedings of ACM SIGGRAPH 2000*, pp. 369–374, 2000.
- [26] S. L. Darrell, G. Shakhnarovich, L. Lee, and T. Darrell, “Integrated face and gait recognition from multiple views,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 439–446, 2001.

- [27] R. Plankers and P. Fua, “Articulated soft objects for multiview shape and motion capture,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003.
- [28] E. Yu and J. Aggarwal, “Detection of fence climbing from monocular video,” in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR06)*, IEEE, 2006.
- [29] S. Park and J. Aggarwal, “A hierarchical bayesian network for event recognition of human actions and interactions,” in *Association For Computing Machinery Multimedia Systems Journal*, pp. 164–179, 2004.
- [30] A. B. Albu and T. Beugeling, “A three-dimensional spatiotemporal template for interactive human motion analysis.,” *Journal of Multimedia*, vol. 2, no. 4, pp. 45–54, 2007.
- [31] G. R. Bradski and J. Davis, “Motion segmentation and pose recognition with motion history gradients,” *Applications of Computer Vision, IEEE Workshop on*, vol. 0, p. 238, 2000.
- [32] J. W. Davis, “Hierarchical motion history images for recognizing human motion,” *Detection and Recognition of Events in Video, IEEE Workshop on*, vol. 0, p. 39, 2001.
- [33] A. Yilmaz and M. Shah, “Actions as objects: A novel action representation,” in *In: Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

- [34] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, “Actions as space-time shapes,” in *In ICCV*, pp. 1395–1402, 2005.
- [35] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, “Tracking groups of people,” *Computer Vision and Image Understanding*, vol. 80, 2000.
- [36] N. Zarka, Z. Alhalah, and R. Deeb, “Real-time human motion detection and tracking,” *Information and Communication Technologies: From Theory to Applications*, 2008.
- [37] J. Nascimento, M. Figueiredo, and J. Marques, “Independent increment processes for human motion recognition,” *Computer Vision and Image Understanding*, vol. 109, 2008.
- [38] R. Polana and R. Nelson, “Low level recognition of human motion (or how to get your man without finding his body parts,” *Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 77–82, 1994.