

2007

Exploring the use of human metrology for biometric recognition

Nikhil Mallikarjun Reddy Burri
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Burri, Nikhil Mallikarjun Reddy, "Exploring the use of human metrology for biometric recognition" (2007). *Graduate Theses, Dissertations, and Problem Reports*. 1847.
<https://researchrepository.wvu.edu/etd/1847>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Exploring the Use of Human Metrology for Biometric Recognition

by

Nikhil Mallikarjun Reddy Burri

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Electrical Engineering

Arun A. Ross, Ph.D., Chair
Donald Adjero, Ph.D.
Xin Li, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2007

Keywords: Biometrics, Anthropometry, Metrology, Hidden Markov Measure Field Models
(HMMF), Active Contours

Copyright 2007 Nikhil Mallikarjun Reddy Burri

Abstract

Exploring the Use of Human Metrology for Biometric Recognition

by

Nikhil Mallikarjun Reddy Burri
Master of Science in Electrical Engineering

West Virginia University

Arun A. Ross, Ph.D., Chair

This thesis explores the possibility of incorporating human body measurements in a biometric framework. While metrological features have been used for identifying persons in the late 19th century, there is limited work in automating this process for surveillance applications. We first establish the relevance of using metrological features in biometric systems by studying two anthropometric data-sets (NASA and NHANES). We then propose a technique to automatically extract a subset of these measurements from a video sequence. A robust segmentation technique (HMMF) to detect moving pixels corresponding to human objects is used in the first stage. Next, we use Active Contours to obtain a precise contour of the human body. Finally, we design a technique to extract the measurements of human body, viz., height, width of the head and the torso, from the segmented image. We show that the measurements extracted in this manner bear close resemblance to manual measurements in terms of their pixel count. To validate the procedure outlined here, we extract these measurements from different videos containing human objects and check for consistency across multiple stand-off distances between the subject and the camera. Data pertaining to 9 different individuals (3 video sequences each) was used in this research.

I dedicate my thesis to my parents, my love, family and friends

Acknowledgments

I must express my gratitude to all the individuals who helped make my education a success. Without their support I would not have come so far.

I appreciate the opportunity provided to me by Dr. Ross to work with him as a graduate student. It has been an honor to have him as my research advisor. His work ethics and attitude have been a total inspiration. I have learned a lot from his determination, dedication and ability to prioritize things. I would have not reached this phase of my life without his guidance and constant motivation. I would also like to thank Dr. Li and Dr. Adjeroh, my committee members, for their suggestions and guidance. I am also very grateful to Dr. Xin Li, my committee member, for his guidance, useful suggestions and classes that I have learned so much from. I appreciate the financial support provided by Center for Identification Technology Research (CITeR) for my thesis.

I would like to take this opportunity to thank my friends and lab-mates for the great times I have experienced at the Biometric Systems Lab. I would like to express my gratitude to Samir Shah and Rohan Nadgir for helping me in getting started with my research. I will always remember Shigefumi Yamada, Christopher Boyce, Simona Crihalmeanu, Phani Ivatury, Rajiv Mukherjee, Raghav Jillela, Nicholas Rymer and Aglika Gyaourova for their co-operation, their feedback on my research and also for the fun we had together.

I would also like to thank my dearest friends Sumanth Dommaraju, Dinesh Penigalapati, Shiva Somishetty, Kalpana Thota, Abhinav Gaddam, Venkat Sathi, Chaitanya Ravi, Madhav Madireddy and last but not the least Praveen Danda. Words are not enough to express my thanks for Manisha Sam Sunder and Rohan Nadgir my lab mates for their constructive criticism.

I will be forever indebted to my parents, my love Lekha, cousins and friends for their love and never-ending support. I will always remember the sacrifices they have made on my behalf.

Contents

Acknowledgments	iv
List of Figures	vi
List of Tables	viii
1 Introduction	1
1.1 Introduction	1
1.2 Literature Review	3
1.3 Anthropometric Features	6
1.3.1 NASA Anthropometric Data set	6
1.3.2 NHANES Anthropometric Data set	6
1.3.3 Experiments on Data-sets	7
1.4 Data Collection	15
2 Human Contour Extraction	17
2.1 Hidden Markov Models For Motion Segmentation	17
2.1.1 Classical Markov Random Field Models	17
2.1.2 Hidden Markov Measure Field Models	21
2.2 Tracking	27
2.3 Active Contours	32
2.3.1 Description of the Model	32
2.4 Measurements	36
3 Summary and Future work	47
3.1 Future Work	47
References	55

List of Figures

1.1	Partitioning the database based on height as an example for indexing.	2
1.2	Examples of human body measurements in a video.	3
1.3	Snap shot of the NHANES data-set [1].	7
1.4	Box plots of the NHANES data for different male and female groups.	8
1.5	Box plots of male NHANES data, variability with respect to different measurements.	9
1.6	Box plots of female NHANES data, variability with respect to different measurements.	10
1.7	Different measurements of the human body [2].	13
1.8	Flow chart depicts the salient components of the proposed system.	14
1.9	Illustrating a typical output (a) using HMMF, Active Contours and (b) Block matching techniques.	16
2.1	Definition of a neighborhood system for an arbitrary pixel X.	19
2.2	Clique Definition.	19
2.3	Initial K classes.	25
2.4	Initial probability distribution, $p(r)$, on a pixel lattice.	25
2.5	The two frames considered for estimating motion using HMMF.	26
2.6	Final probability distribution on the same lattice.	27
2.7	Output of the HMMF technique with the contours of the moving objects (human) on the original set of frames.	28
2.8	Shadows of the moving object (human) present in the video frames.	29
2.9	Output of the HMMF technique when there was significant amount of illumination change.	30
2.10	Moving objects tracked across frames along with their intensity histograms.	31
2.11	Evolving Surface C.	32
2.12	The contour of the human body obtained from using “Active Contours” overlapped on a video sequence.	37
2.13	Contour of the entire human body obtained by combining all the three techniques on a video sequence.	38
2.14	The final evolution curve (a) head, (b) torso, (c) legs	39
2.15	Overlapped ellipses on the video frames.	42
2.16	Ellipses overlapped on the original frames for extracting measurements at various distances.	45
2.17	Contours of the ellipses without any background at various distances.	46
3.1	3D start model	48
3.2	A part of the 3D model is zoomed	48

3.3	Agglomeration of the human body components	49
3.4	Range by number of edges and by distance (circle) (taken from [3])	50
3.5	Decay Functions (a) No decay function, (b) Cone decay function, (c) Bell decay function, (d) Cusp decay function, (e) random decay function, (f) Sine wave decay function (taken from [3])	51
3.6	Example for 3D point registration: (a) before stitching ; (b) after stitching	53

List of Tables

1.1	Mean and standard deviation of measurements across various ethnicity groups [1].	11
1.2	Different human body measurements from the NASA data-set.	12
2.1	Comparison of actual (A) vs estimated (E) values in terms of pixels.	43
2.2	Comparison of actual (A) vs estimated (E) values in terms of pixels.	43
2.3	Comparison of actual (A) vs estimated (E) values in terms of pixels.	43
2.4	Comparison of actual (A) vs estimated (E) values in terms of pixels.	44

Chapter 1

Introduction

1.1 Introduction

Biometric systems automatically recognize individuals based on their physical and/or behavioral characteristics like fingerprint, face, hand-geometry, iris, retina, palmprint, voice, gait, signature, and keystroke dynamics [4]. For a physical or behavioral characteristic to be declared as a biometric, it has to satisfy the following properties [4]: uniqueness, universality, permanence, collectability, performance, acceptability and difficult to imitate. In addition to the traditional biometric identifiers (such as fingerprint, face, hand-geometry), there can be other traits which may not satisfy the properties mentioned above, but may still be used in a biometric system. The ones falling under this category are generally referred to as “Soft Biometrics”. Soft biometrics, as defined by Jain et al. [5], are “characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals”. The ones falling under this category include height, weight, gender, eye color, etc. Metrology is one such soft biometric, and is the science of measuring the geometry of human body parts. Measurements such as the height and width of face, hands, arms, legs, etc. of an individual may not be permanent across time, but may be useful in certain applications. For example, one can use the information provided by metrology as an additional feature in conjunction with the traditional biometrics. Wayman [6] proposed the use of soft biometric attributes to partition a biometric database. Partitioning the database can greatly improve the search efficiency of the biometric system; however partitioning errors can degrade recognition performance.

One other class of applications for metrological features would be surveillance in hostile envi-

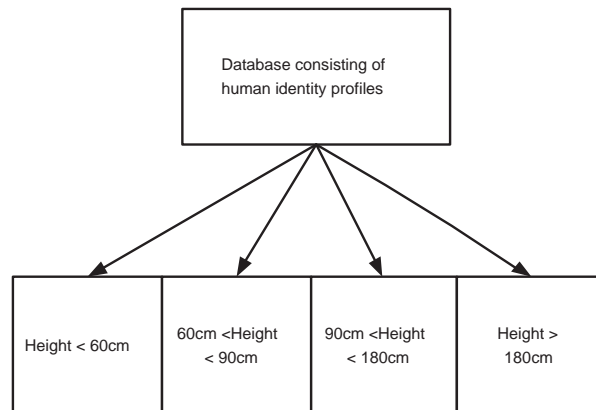


Figure 1.1: Partitioning the database based on height as an example for indexing.

ronments (e.g., night time environments, from a distance) involving non-cooperative individuals (e.g., masked), or battlefield environments (e.g., soldiers in full gear). In these scenarios, the human ‘shape’ can play a pivotal role in characterizing subjects. But video metrology poses its own set of challenges. Usually the surveillance videos have low resolution and high noise levels which make it difficult to extract the human measurements. Changes in illumination and pronounced shadows also make segmenting subjects of interest and extracting key measurements difficult. The other key issue in real time implementation of video metrological systems is that of computational complexity. Most processors, particularly the ones used in surveillance cameras, have limited computational power, and hence, the algorithms used for video metrology should have low complexity. Another challenge in video metrology is identifying the reference points for measurements and tracking them through the video. These reference points, for example, can be the tip of the head, shoulder joints, hip joints, etc. which can be handy in extracting the key measurements. Automating the selection of such reference points from low resolution surveillance videos presents a significant challenge. Segmentation of the subjects of interest, which is usually needed to perform metrology, is also challenging especially in low resolution surveillance videos.

The goals of this thesis are:

1. To establish the utility of anthropometric features in biometric indexing or database partitioning.
2. To design a technique that automatically extracts some metrological features from a video stream.

3. To develop a model for representing extracted features.
4. To evaluate the proposed feature extraction and feature matching technique.

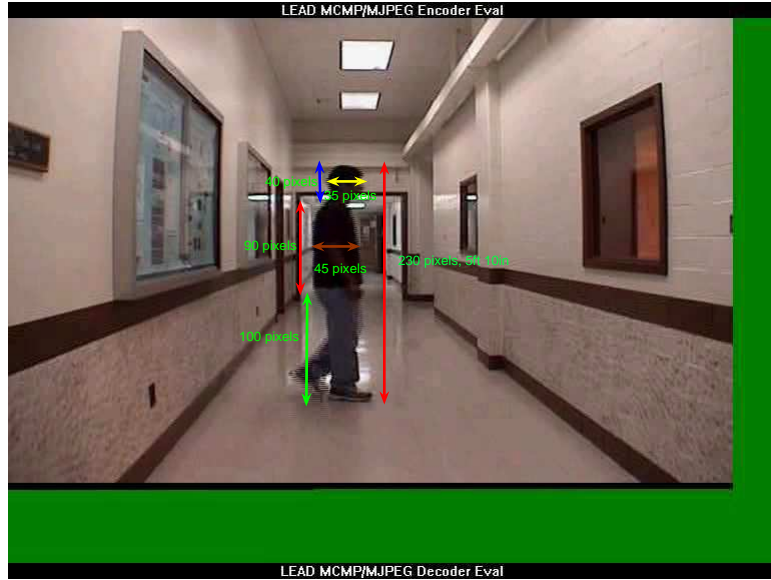


Figure 1.2: Examples of human body measurements in a video.

To accomplish goal (1), in this thesis a simple statistical analysis on the NASA [2], and NHANES [1] anthropometric data was conducted, which revealed the significance of using the measurements. The details of the NASA and the NHANES anthropometric data-sets and the experiments conducted on them are presented later in the chapter. From an automation perspective, the solution to the problem of extracting measurements is twofold. Firstly, an exact and complete contour of the human body has to be obtained with a fixed camera (this is related to object segmentation in a video sequence). Secondly, measurements of the body have to be extracted from the contour to accomplish goal (2).

1.2 Literature Review

Humans can be classified based on the shape information. Hand geometry and 3D face systems have been studied and in some cases deployed for use as biometrics. The technique of using human body measurements as biometric attributes is not new and dates back to the early

1880's. The French savant, Alphonse Bertillon was the first person (according to the literature [4]) to use different body measurements for identification. Significant developments have been made in the field of person identification since then and much more in recent years.

Several approaches have been used in the computer vision literature for automatic person identification using gait from video. In this thesis the problem addressed is closely related to the gait recognition technique [7], [8], [9], [10], [11], [12]. To our knowledge much work has not been done in the field of extracting human measurements from videos. Grother et al. [13] demonstrates the effectiveness and efficiency of the shape-based biometrics from the newly available CAESAR database [14]. The CAESAR database contains high resolution, three-dimensional, shape information of the human body for a large population . The paper talks about how an automated system would perform human identification by using a small amount of non-dynamic anthropometric information. They use twelve different body measurements that are pose-independent to demonstrate that they can be used for person identification.

A two-dimensional video sequences is used to demonstrate the use of metrology as a biometric. The problem is difficult because we are not only supposed to detect and track the humans but also are required to extract body measurements based on the entire video sequence. A brief description of the techniques related to the problem are presented. Firstly, the techniques that are related to moving object detection and tracking problem are presented. Optical Flow techniques have been used in the literature for motion estimation (Horn and Schunck [15], Lucas and Kanade [16], Uras et al. [17], Nagel et al. [18], Anandan et al. [19], Singh et al. [20], Heeger et al. [21], Waxman et al. [22] etc). These techniques are computationally complex and all of them cannot be employed directly to the problem as they output an approximate region corresponding to motion as opposed to computing a precise contour of the moving object. For segmenting human body regions, Level Sets have also been used [23].

Once the segmented region is obtained, it has to be tracked across different frames. A lot of work has been done in the field of object tracking. The techniques employ either the intensity information (color or gray scale) or the shape information to track objects (humans). In the case of using shape information, descriptors such as area, perimeter, non-circularity, eccentricity, etc. are used [24], [25], [26], [27], [28]. In the case of using intensity, a filter is applied to the selected region of interest and features are extracted from the ensuing image. Based on these features, the object of interest is tracked. Video object-tracking based on extended active shape models have

been used; however most of them use the information of markers placed on the human body [29].

Some of the literature work in the field of extracting human body measurements are described below. Crimini et al. [30] have suggested a height measurement technique in which the persons height is measured from a single image. This technique works on the assumption that the person is standing still. The drawback of this technique is that it is not a reliable assumption in a normal surveillance scenario where the subject is walking. The results of estimating the height are obtained from projective geometry and computer vision. The accuracy of measuring is demonstrated where the ground truth is known. Their paper also deals with uncertainty of heights with humans in motion. Madden and Picardi [31] have suggested a session-based biometric in which the height of an individual is estimated throughout a sequence of images. The technique employed by them determines the location of the top of the head in the image and its projected location on the ground plane to estimate the height of the individual. To estimate the ground point directly below the head, they consider the position of the heels, where the position of the heels is manually marked. These multiple estimates of heights are statistically analyzed to determine the exact height. The cameras used for their study are fixed cameras which allow for calibration. These measurements were conducted on five different video sequences and three different individuals. The paper reports the average height and the standard deviation of the height across different frames of the five video sequences of an individual. Godil and Ressler [32] describe a framework for similarity based retrieval and clustering from a 3D human database. The experiments were conducted on the CAESAR [33] anthropometric database. To effectively represent the 3D points they have developed two different representation schemes based on the human body shape. One of them uses the information of the human body measurements (mostly distance between single large bones), and the other one is based on Fourier descriptors of features that are obtained from three different silhouettes rendered by the human body from the front, side and top. These representations are used for similarity retrieval. The paper concludes saying that the body and head descriptors represent the CAESAR data with reasonable amount of accuracy. They also show the results of using human body shape descriptors for clustering the 200 subjects considered in their study from the CAESAR database.

1.3 Anthropometric Features

1.3.1 NASA Anthropometric Data set

A brief description of the NASA anthropometric data-set is presented here. The Anthropometric Source Book [2] was basically designed to provide NASA, its contractors, various other government agencies, and the civilian sector a comprehensive description of the anthropometric data. Specifically, the data meets the needs of engineers engaged in the design of equipment and clothing for the NASA Space Shuttle/Spacelab. The book contains tables of anthropometric data, that defines the physical size of various body components, mass distribution properties, and dynamic capabilities of U.S. and selected foreign populations. The material includes information pertaining to both male and female of various age groups, socio-educational backgrounds, races, and ethnic background. The anthropometric survey is of 61 military and civilian populations of both sexes from the U.S., Europe and Asia. The data consists of 295 accurately measured variables (not all measurements are available for all groups). The tables in the book also contain the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentile of the given population.

1.3.2 NHANES Anthropometric Data set

A brief description of the NHANES [1] anthropometric data-set is presented here and the primary reason why the data was collected. The information regarding nutrition is one of the important criteria for health, and to solve many of the public health concerns requires survey data. The major aim in conducting the survey by the NHANES is to provide information useful for studying the relationship among diet, nutritional status, and health. To assess the nutritional status a series of stature, weight and other anthropometric dimensions are required. The data collected during the survey is basically used for assessing growth, body fat distribution, child growth charts (i.e., generally to answer the questions related to health concern). There were around 9000 individuals examined during the survey. And each individual is grouped based on different races, age and gender. There were around 27 anthropometric measurements but some of them were ignored as they were not related to the physical dimensions of the human body (for example, the body weight or the mass index).

1.3.3 Experiments on Data-sets

The details of the statistical analysis based on the NHANES anthropometric data-set is as follows. The mean, standard error of the mean, and selected percentiles, by race-ethnicity and age are presented in the tables. The classification of the data in the first stage is based on the age for male and female sub-groups. From these sub-groups, the data is once again classified based on race/ethnicity (non-hispanic white, non-hispanic black, mexican american and all other races which do not fall in the first three categories). The data is also classified on the basis of age (i.e. 20-29 years, 30-39 years etc.). The data-set had about 22 measurements, which we narrowed down to 11 measurements. The eleven measurements are as follows: height, sitting height, upper-arm length, upper-leg length, mid-upper arm circumference, thigh circumference, waist circumference, wrist-breadth, biacromial breadth, biliac breadth and elbow breadth. Even though some of these are three dimensional measurements and seem to be unrelated, they can be deduced from the two dimensional measurements with some degree of uncertainty. A snap shot of the data is shown in Figure 1.3.

Table 11: Height in inches for males 20 years and over-number of examined persons, mean, standard error of the mean, and selected percentiles, by race-ethnicity and age - United States, 1966-1994

Race-ethnicity and age	Number of examined persons	Mean	Standard error of the mean	Selected percentile										
				5th	10th	15th	25th	50th	75th	85th	90th	95th		
All race/ethnicity groups¹														
20 years and over	7343	69.1	0.06	64.4	65.5	66.1	67.2	68.1	71.5	72.1	73.5	74.5		
20-29 years	1831	69.3	0.11	64.4	65.6	66.2	67.4	68.2	71.5	72.1	73.1	74.5		
30-39 years	1461	69.5	0.11	64.9	66.0	66.6	67.6	68.5	71.5	72.6	73.6	74.4		
40-49 years	1204	69.4	0.12	64.7	65.6	66.2	67.5	68.4	71.4	72.2	73.0	74.1		
50-59 years	655	69.2	0.15	64.7	65.7	66.4	67.4	68.3	71.5	71.9	72.5	73.5		
60-69 years	1176	69.5	0.12	63.9	65.1	65.6	66.6	67.6	70.5	71.5	72.0	72.9		
70-79 years	875	67.7	0.13	63.5	64.3	64.6	65.5	67.7	69.5	70.5	71.1	72.2		
80 years and over	699	66.7	0.16	62.0	63.1	63.9	64.9	65.7	68.6	69.6	70.4	71.1		
Non-Hispanic white														
20 years and over	3286	69.5	0.07	65.1	66.1	66.6	67.6	68.4	71.4	72.3	73.1	74.1		
20-29 years	783	69.9	0.13	65.5	66.6	67.3	68.5	69.4	71.7	72.9	73.3	74.1		
30-39 years	454	70.0	0.15	66.0	66.7	67.2	68.1	69.0	71.7	73.1	73.9	74.5		
40-49 years	626	69.8	0.13	65.4	66.3	67.0	68.1	69.0	71.6	72.4	73.1	74.1		
50-59 years	418	69.6	0.12	65.6	66.4	67.0	67.9	68.6	71.1	72.1	72.6	73.8		
60-69 years	510	69.9	0.11	65.0	65.9	66.5	67.1	68.0	70.6	71.5	72.2	73.1		
70-79 years	524	67.8	0.12	63.7	64.4	64.9	65.9	67.6	69.6	70.6	71.1	72.2		
80 years and over	556	66.9	0.12	62.4	63.3	64.1	65.0	66.8	68.6	69.6	70.5	71.1		
Non-Hispanic black														
20 years and over	2313	69.3	0.06	64.6	65.6	66.4	67.5	68.4	71.2	72.3	73.1	74.1		
20-29 years	469	69.7	0.14	65.2	66.1	66.8	67.8	68.5	71.6	72.8	73.6	74.4		
30-39 years	497	69.6	0.12	65.3	66.4	67.1	68.1	69.0	71.4	72.4	73.3	74.4		
40-49 years	370	69.6	0.16	65.0	65.5	66.4	67.5	68.5	71.5	72.9	73.5	74.2		
50-59 years	217	69.5	0.19	64.8	66.1	66.5	67.2	68.1	70.7	72.0	72.4	73.0		
60-69 years	295	68.3	0.17	63.8	64.9	65.4	66.4	68.2	70.2	71.0	72.0	72.7		
70-79 years	187	67.9	0.21		64.5	64.6	66.0	68.0	70.1	70.7	71.5			
80 years and over	52	66.5	0.46					65.1	67.1	68.6				
Mexican American														
20 years and over	2258	69.8	0.07	62.6	63.5	64.2	65.1	66.7	68.5	69.5	70.2	71.3		
20-29 years	675	69.8	0.12	62.8	63.6	64.3	65.1	66.9	68.5	69.5	70.1	71.1		
30-39 years	474	67.2	0.16	62.8	63.7	64.3	65.6	67.1	68.9	70.0	70.6	71.9		
40-49 years	261	66.6	0.15	63.1	63.7	64.4	65.1	66.6	68.4	69.5	70.2	71.6		
50-59 years	178	66.8	0.21		63.7	64.1	65.0	66.7	68.1	69.0	69.6	70.1		
60-69 years	335	66.3	0.15	62.1	63.3	63.9	64.6	66.3	68.1	69.0	69.4	70.1		
70-79 years	148	65.2	0.22		62.4	63.0	63.9	65.0	66.6					
80 years and over	63	64.0	0.36				62.1	64.2	65.3	67.9	69.2			

¹ Figures do not use more standard of reliability or precision.
² Includes all other race/ethnicity groups not listed separately.

Figure 1.3: Snap shot of the NHANES data-set [1].

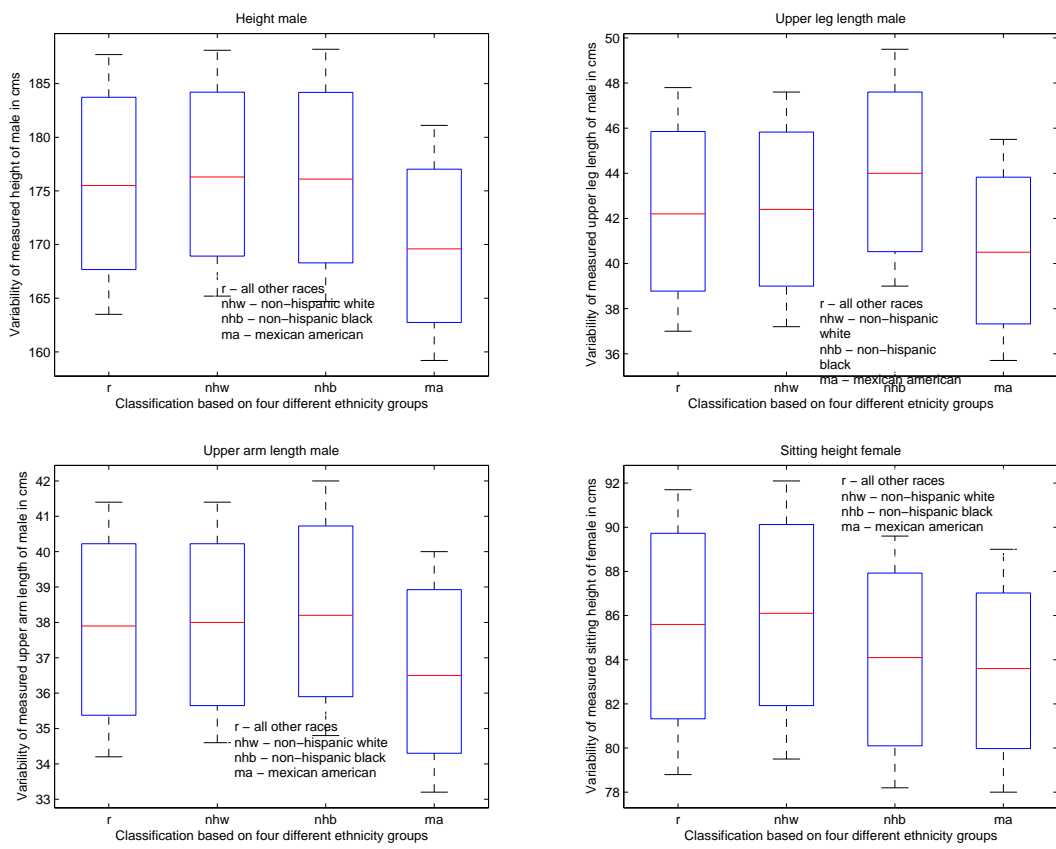


Figure 1.4: Box plots of the NHANES data for different male and female groups.

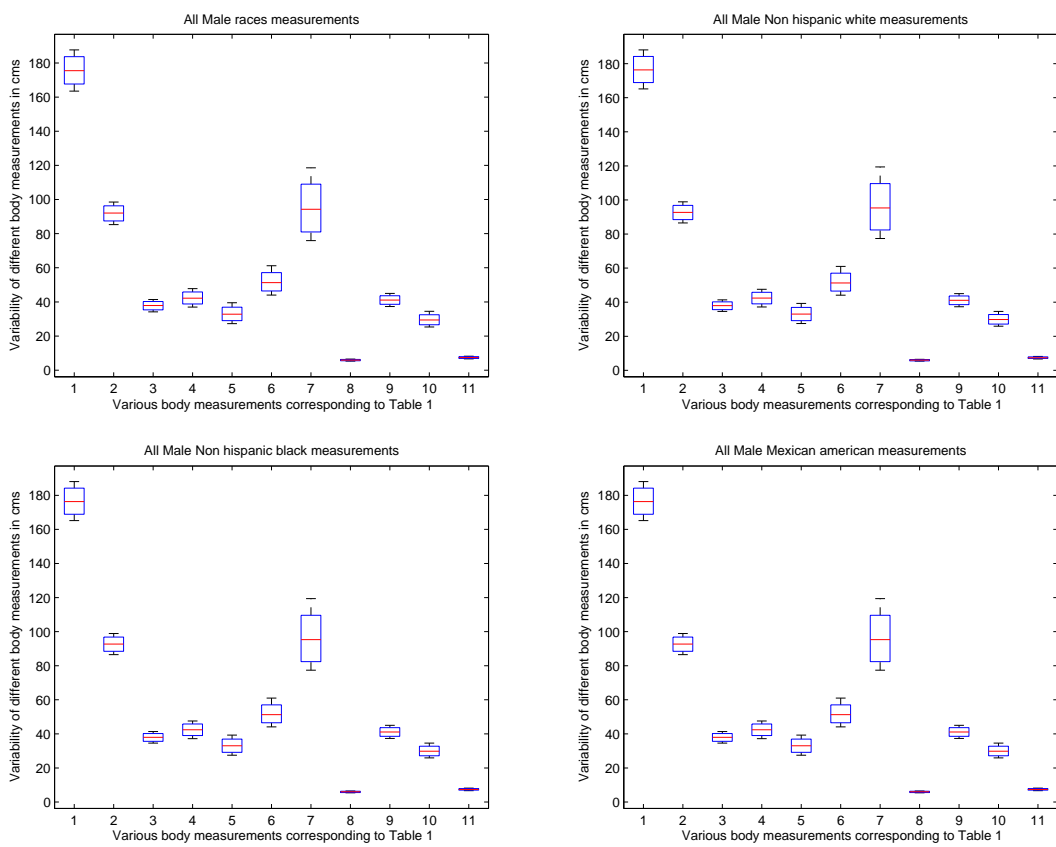


Figure 1.5: Box plots of male NHANES data, variability with respect to different measurements.

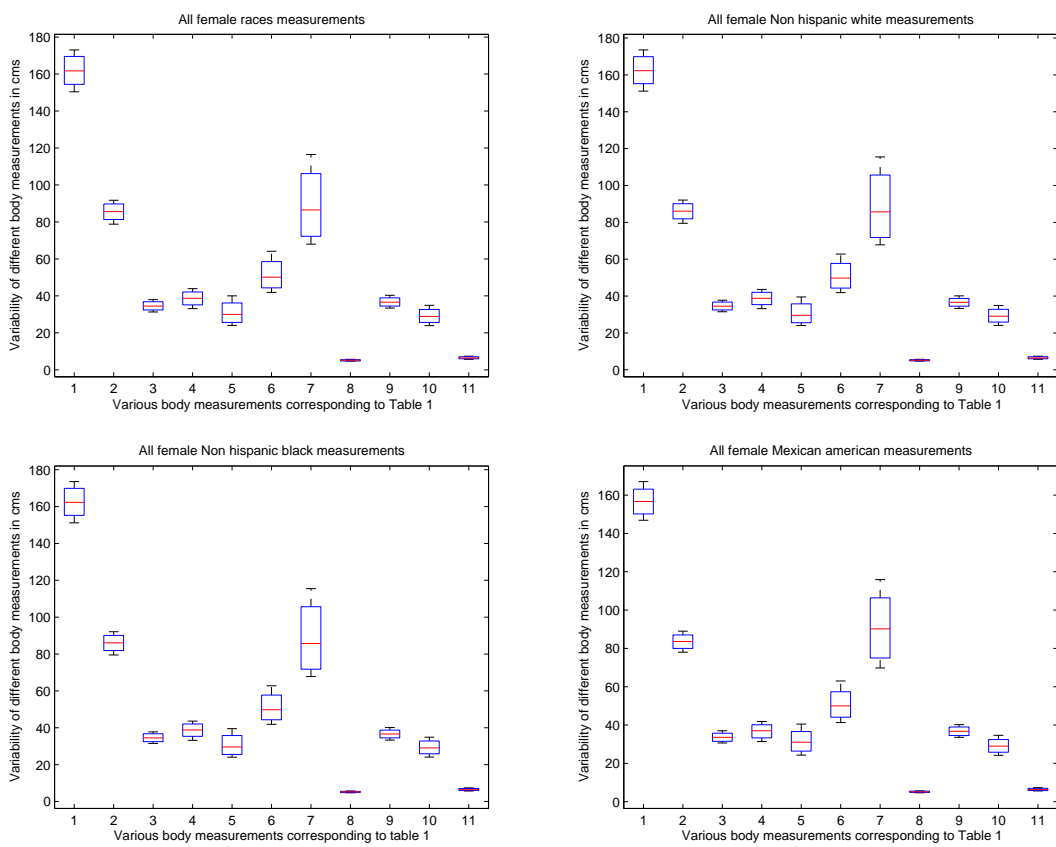


Figure 1.6: Box plots of female NHANES data, variability with respect to different measurements.

	Non-Hispanic White	Non-Hispanic Black	Mexican American	All Other Races
Height	176.5, 10.31	176.1, 9.19	169.7, 9.02	175.6, 14.25
Sitting Height	92.6, 5.56	89.6, 4.07	89.6, 4.66	91.9, 7.84
Upper arm	38.0, 2.75	38.3, 3.12	36.5, 3.25	37.8, 4.32
Upper leg	42.4, 9.88	44.0, 6.20	40.6, 7.42	42.3, 6.88
Mid Upper arm	33.1, 6.18	33.6, 5.37	32.5, 5.60	33.0, 7.85
Thigh Circumference	51.7, 7.67	53.5, 7.07	50.8, 7.86	51.7, 10.30
Waist Circumference	96.3, 20.34	91.8, 19.96	93.5, 19.91	95.2, 25.82
Wrist breadth	5.9, 0.55	5.9, 0.44	5.8, 0.46	5.9, 0.86
Biacromial Breadth	41.1, 5.52	41.8, 4.46	41.0, 3.25	41.1, 4.32
Biliac breadth	30.0, 5.50	28.2, 4.89	29.1, 4.63	29.6, 6.03
Elbow breadth	7.4, 1.10	7.4, 0.89	7.1, 0.92	7.4, 0.86

Table 1.1: Mean and standard deviation of measurements across various ethnicity groups [1].

Based on the percentile information, we plotted the Box plots. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. This gives us a clear indication about the variance of the data across different persons in the same group as well as different groups. From the plots, we can observe that, for example, one can distinguish Mexican Americans from the rest of the category based on the height of the person. Using the information related to sitting height of females from Figure 1.4, one can distinguish the Non-Hispanic Black and Mexican Americans from the rest of the categories. Also from Table 1.1 we observe that the variance for height, waist circumference is very large, making them an important distinguishable feature across different groups as well as for individuals within the same group. To some extent all other measurements like sitting height, upper arm, mid upper arm circumference, upper leg, biliac breadth (except wrist breadth) can be used for distinguishing groups.

Even though some of the features are three dimensional measurements, they can be derived from the two dimensional measurements with some degree of uncertainty. Consider the waist circumference; if we can obtain the approximate width of the torso from the frontal and the side view of the human with respect to a fixed camera we can estimate the circumference of the waist within a certain confidence interval. The three dimensional measurements suggest that their corresponding two dimensional measurements are also variant across different groups (though this is not very intuitive). Later in the thesis, we extract four measurements (height, width of the head and torso) and all these measurements correspond to the side view with respect to the

ID Code	Description
23	Acromial Height
39	Radiale Length
103	Biacromial Breadth
107	Biauricular Breadth
245	Chin prominence
389	Glabella to top of head
391	Glabella to wall
107	Biarcular breadth
142	Bitragion breadth
150	Bitragion Menton
289	Ectocanthus to Otobasion
291	Ectocanthus to wall
386	Functional Leg length
629	Nasal root height
698	Radial stylium length
758	Sitting height
798	Span Akimbo
805	Standing height (Stature)
931	Waist circumference

Table 1.2: Different human body measurements from the NASA data-set.

viewer (camera).

The second set of experiments were conducted on the NASA Human anthropometric database. The database consists of 10,000 people in 90 categories from 20 different countries. There were around 260 measurements all together that were collected, and based on the ease of automatic detection, redundancy we pruned them initially to 19 measurements and further reduced to 14 measurements. A sample list of measurements are listed in tabular form in Table 1.2. Figure 1.7 shows the various measurements as established manually.

We studied the relative significance of the individual measurements based on their discriminative ability, correlation with others and error resilience.

When looking at the NHANES and NASA data-sets, the measurements corresponding to different body parts are same ensuring us the fact that these set of measurements are variable across different individuals and groups (when one group is compared with another group).

This inspired us to work on extracting some of the measurements from a video sequence. The reason for using video sequences, instead of still images, is that information (measurements) can be extracted and accumulated across different frames. The videos considered in this work

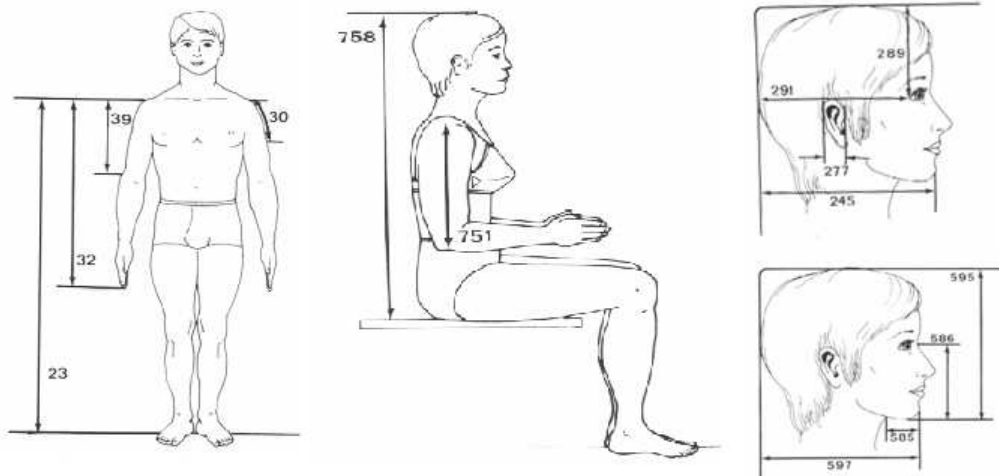


Figure 1.7: Different measurements of the human body [2].

consist of only a single person walking, with no other moving objects, obtained using a fixed camera. From an automation perspective, the solution to the problem is twofold. Firstly, an exact and complete contour of the human body has to be obtained with a fixed camera (this is related to object segmentation in a video sequence). Secondly, measurements of the body have to be extracted from the contour to accomplish goal 2.

Motion segmentation in a video sequence is the basic operation for retrieving information about moving objects. It is important to precisely detect the boundaries and segment the video into semantically homogeneous regions for video processing. To address the problem of segmentation, we have employed Hidden Markov Measure Field Models and Active Contours.

Markov Random Field (MRF) models have been used in the literature for solving problems related to image and video analysis; these include applications in different fields like image restoration, texture modeling, object segmentation, object matching, etc [34]. The success of these models can be attributed to the fact that they give rise to good, flexible, stochastic image models [35]. One type of problem that has been approached using MRF models is motion segmentation: this consists of partitioning a particular frame in a video sequence into regions corresponding to motion and those that do not. Let us define a set of frames, $\{F_1, F_2, F_3, \dots, F_k\}$

pertaining to a video sequence. The i^{th} frame is tessellated into a set of m non-overlapping regions $\{R_1^i, \dots, R_M^i\}$ such that the variation of some property like velocity between frames is either constant or follows a simple model Φ_k (e.g., a linear function) within each region. This is a difficult problem to solve because one has to find both the model parameters as well as the corresponding regions across the frames (i.e., the region where the model holds) at the same time [36]. In this thesis, we use the same class of probabilistic models used in [35] which characterizes the solution for complex motion segmentation in terms of minimizing a differentiable energy function. The chapters in this thesis are organized as follows: In Chapter 2, we present a brief description about the classical Markov Random Field models, introduce the Hidden Markov Measure Field Model technique (HMMF) and illustrate the results on a video sequence. We also present a technique to track individual parts viz. head, torso, and legs over an entire video sequence. From the results obtained using HMMF we observe that a more precise contour is required to extract human body measurements. To address this, we employ an active contours based technique. We then present a technique to extract the measurements of the different body components. In Chapter 3 we present the 3D models that can be generated from the two dimensional measurements and also the future work.

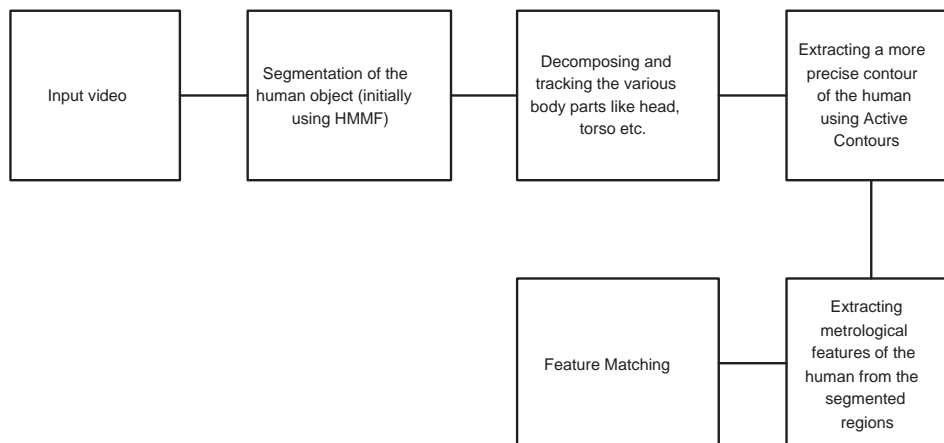


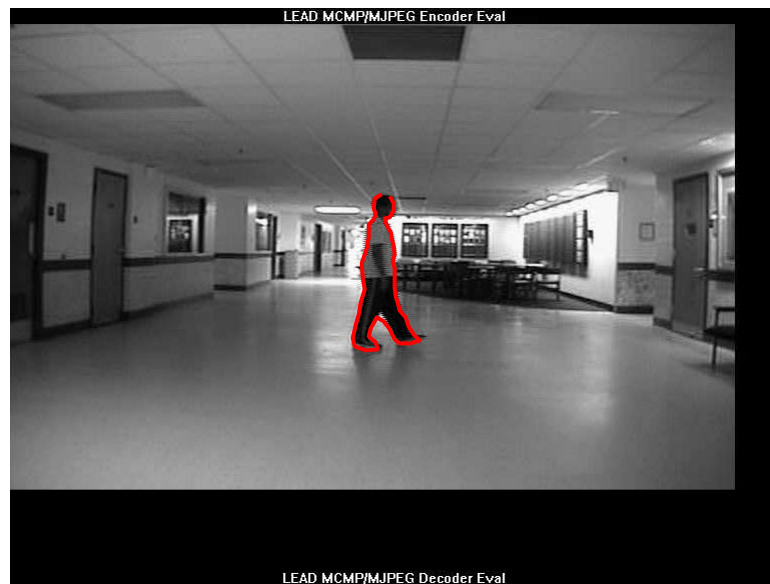
Figure 1.8: Flow chart depicts the salient components of the proposed system.

The primary reason why we chose to go with the HMMF model [35] rather than the traditional block matching algorithm is the need to obtain an exact and complete contour of the human form. Also, to our knowledge the traditional schemes do not output a precise contour but rather specify a broad region (e.g., rectangle). A simple background subtraction proceeded by

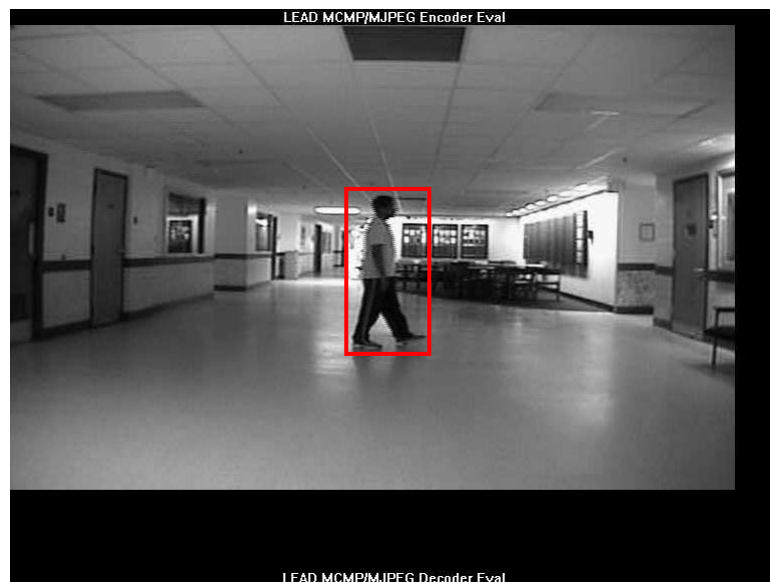
thresholding operation would also be sufficient for extracting the human form in cases where the background is stationary. But this would output only one region, whereas the HMMF model outputs regions with different velocities whose information can be later used for processing the input data.

1.4 Data Collection

The videos were collected using a Sony camera with a resolution of 768×568 in the color RGB space. The user is asked to walk across a hall way as shown in Figure 1.2 across three stand off distances from the camera. The camera is situated approximately at the same level of the human. In addition the user is requested to walk sideways. This simple scenario is not the case in practical situations like in airports. More rigorous data collection schemes like the camera situated at an altitude of say 20 meters and the individual walking in any direction have to be acquired. This type of data would be more useful for testing the proposed scheme. But due to time constraints, this type of data collection was not performed. Also the three stand off distances were chosen randomly (12ft, 24ft, 36ft) and there is no particular reason in choosing these specific ones. We would also like to acquire the data where the person is standing relatively far enough from the camera. For example the camera being situated on top of a building close to the parking lot and acquiring information corresponding to individuals. In the data collection procedure the individual is asked to walk relatively slowly to acquire more information pertaining to him/her, but in the proposed data collection scheme the individual can walk at his usual pace.



(a) Precise contour of the objects (human)



(b) Region as a rectangle corresponding to the moving objects

Figure 1.9: Illustrating a typical output (a) using HMMF, Active Contours and (b) Block matching techniques.

Chapter 2

Human Contour Extraction

2.1 Hidden Markov Models For Motion Segmentation

2.1.1 Classical Markov Random Field Models

Markov Random Field image models represent knowledge in terms of “local” probability distributions. Specifically, the kinds of probability distributions generated by MRFs have a local neighborhood structure. Given the conditional probability distribution (in our case: given two frames what is the probability of observing moving pixels in these frames) we need to obtain a joint probability distribution (probability of the two frames and the moving pixels). It is the joint probability distribution that gives the complete motion representation and not the conditional distribution. We model the *a priori* information using Gibbs distribution. The reason for using Gibb’s distribution is explained in the Hammersley-Clifford theorem (Besag [37]; Geman and Geman [38]) which states, that any conditional distribution has a joint distribution which is Gibbs (Kinderman and Snell [39]; Dubes and Jain [40]) if the following conditions hold: Positivity, Locality and Homogeneity. The term Positivity refers to $P(f) > 0$. The term homogeneity refers to the probability of observing a pixel is independent of the positions in the pixel lattice. The locality refers to the Markov assumption and the Markovianity property is defined as $P(f_i|f_{S-i} = P(f_i|f_{N_i}))$, where S refers to the entire pixel lattice. The term ‘N’ refers to a neighborhood system and is discussed later. ‘i’ denotes the current pixel in the neighborhood system. And site ‘S’ refers to neighboring pixels defined by cliques.

To describe the motion segmentation technique, we introduce the following notation. Let L

represent the pixel lattice where video frames are observed. The model assumes that there are M regions, $\{R_1, \dots, R_M\}$, such that L is the union of all the M regions (i.e., $L = \bigcup_{k=1}^M R_k$), and there are no overlapping regions (i.e., $R_i \cap R_j = \phi, i \neq j$). Consider two video frames, I_1 and I_2 , in this pixel lattice. The observation at pixel $r \in L$ can be written as:

$$I_1(r) = I_2(r + \phi(r, \theta_{f(r)})) + n(r). \quad (2.1)$$

where ϕ is a function of the pixel location, the parameter θ_k defines the transformation of the corresponding pixel between the two frames (i.e., the motion vector), $f(r)$ is a labeling function that indicates the region corresponding to r (i.e., $f(r)$ maps pixel r to a region $k \in \{1, 2, \dots, M\}$) and $n(r)$ is a white noise field with known distribution P_n . In the videos considered in this work, the subjects typically moved in the horizontal direction. Thus, we assume θ_k to be 1-dimensional (i.e., $\theta_k = x_k$) although a more complex transformation may be used. The goal is to estimate $f(r)$ (which can be thought of as a Markov Random Field) and $\theta_{f(r)}$. Since we use a piecewise constant model, the equation (2.1) can be modified as

$$I_1(r) = I_2(r + \theta_{f(r)}) + n(r), \quad (2.2)$$

where $n(r)$ is a white noise field with known distribution P_n (e.g., $n(r)$, $r \in L$ are 0-mean, independent, identically distributed Gaussian random variables with standard deviation σ). In this model, the a priori model f is assumed to be a sample from a MRF, obtained with a Gibbsian distribution:

$$P_f(f) = \frac{1}{Z_f} \exp[-\sum_C V_C(f)], \quad (2.3)$$

where Z_f is a normalizing constant and the sum in the exponent ranges over the cliques of a given neighborhood system on L , and $V_C(f)$ is the “potential function” corresponding to the clique C . These potential functions, together with the neighborhood system that is selected, control the appearance of the sample model f and, hence, the properties of the estimated motion region.

We use the neighborhood system defined in Figure 2.1(a) and the clique defined in Figure 2.2(b) due to their simplicity. However experimenting with other neighborhood systems and cliques should be possible..

The motion segmentation problem involves not only determining the motion but, also the region where the motion is present (i.e., to estimate the region and also the parameter vector $\theta = (\theta_1, \dots, \theta_M)$, given the observations I_1 and I_2).

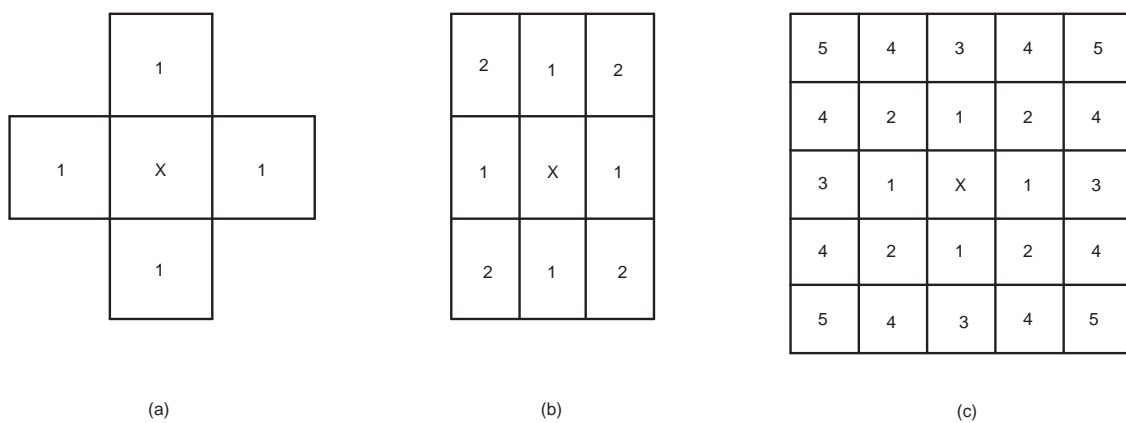


Figure 2.1: Definition of a neighborhood system for an arbitrary pixel X.

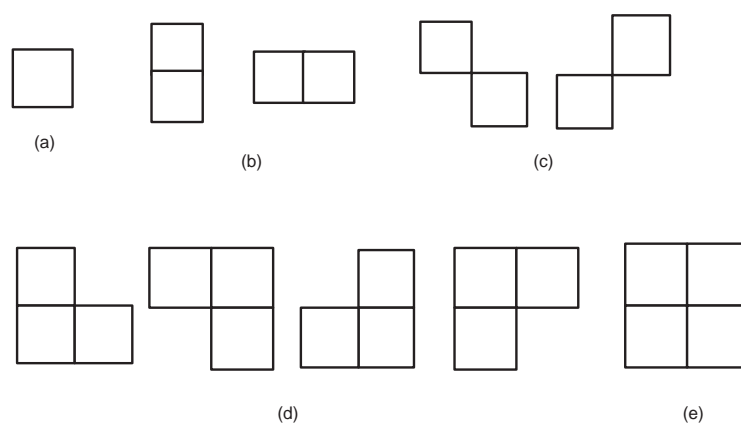


Figure 2.2: Clique Definition.

The following points indicate about solving the problem in a broader sense. A detailed description of the mathematics is presented next.

1. All the pixels have the same probability associated with them in order to determine whether they are moving or not. Only moving pixels have higher probability.
2. The parametric vector θ is deterministic; so the probability of observing the *Motion* given the prior model f and the parameter vector $P(\text{Motion}|f, \theta)$ depends only on modeling the noise parameter.
3. From $P(f)$ and $P(\text{Motion}|f, \theta)$ we can compute $P(f, \theta|\text{Motion})$. The *a priori* model f^* and parameter vector θ^* which maximize this distribution will be the solution to the problem.

Let I_2 be the second frame (a displacement vector added to the first frame) and I_1 be the first frame. Then,

$$I_1 = I_2 + n,$$

where n is zero mean independent noise, $n \sim P_N$. For Gaussian noise, $N(0, \sigma^2)$, the probability that the noise has a value $a-b$ is:

$$P(I_1(r) = a|I_2(r) = b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(a-b)^2}{2\sigma^2}\right). \quad (2.4)$$

Since the noise is independently distributed,

$$P(I_1|I_2) = \prod_{r \in L} P_n(I_1(r)|I_2(r)).$$

For the parametric model

$$I_2 = I_2(r + \theta_{f(r)}).$$

The parametric model depends on the *a priori* model, f , and the parameter vector, θ , so the likelihood term associated with the frames is $P(\text{Motion}|f, \theta)$. We compute the posterior $P(f, \theta|\text{Motion})$ by Bayes's rule as:

$$P(f, \theta|\text{Motion}) = \frac{P(f, \theta)P(\text{Motion}|f, \theta)}{P(\text{Motion})}, \quad (2.5)$$

where, $P(\text{Motion})$ is a normalizing constant, $P(f, \theta) = P(f)P(\theta)$.

We require the *a priori* model, f^* , and parameter vector, θ^* , which maximize the posterior distribution $P(f, \theta | \text{Motion})$. The minimization of step 2.5 is generally called as Segmentation/Model Estimation (SM) algorithm, and is usually a two step procedure (one particular example is the Expectation Maximization (EM) algorithm). In the EM algorithm, given the current estimate of parameter vector θ the best segmentation is found in step S, and given the current estimate for the segmentation, the optimum or the best values for parameter θ is found in step M. One can always resort to these techniques but the major problem is that it is not always possible to find the exact optimal segmentation in step S, and so one is forced to resort to approximations, the most common ones being stochastic, Markov Chain Monte Carlo [41], [38], which are all computationally very expensive. Other techniques which are fast are known in the literature [42] but they are highly sensitive to noise [35].

f^* and θ^* are called the MAP (maximum *a posteriori*) estimators. Finding the global maximum of $P(f, \theta | \text{Motion})$ can be difficult and slow. The hidden Markov measure field (HMMF) model efficiently computes the MAP estimator.

2.1.2 Hidden Markov Measure Field Models

A Markov random vector field ' p ' is first generated with distribution $P(p) = \frac{1}{K} \exp[-\sum_C W_C(p)]$ (a Gibbsian prior), where K is a normalizing constant, C is the set of cliques of a given neighborhood system in the lattice, and W_C is the potential function; $p(r)$ is a M -dimensional vector such that $0 \leq p_k(r) \leq 1$, $k = 1, 2, \dots, M$, and $\sum_{k=1}^M p_k(r) = 1$. The simple quadratic potential,

$$W_{rs}(p(r), p(s)) = \lambda |p(r) - p(s)|^2 = \lambda \sum_{k=1}^M (p_k(r) - p_k(s))^2,$$

is used as the potential function. Here, λ is a positive parameter controlling the smoothness of the field and pixels $\langle r, s \rangle$ are in the same clique. The field f is then generated in such a way that each $f(r)$ is an independent sample from the distribution $p(r)$. Thus,

$$P(f|p) = \prod_{r \in L} p(r).$$

In this formulation, motion can be detected by observing the variables $p(r)$ and $\theta_{f(r)}$. The conditional distribution $P(p(r), \theta_{f(r)} | \text{Motion}(r))$ can be obtained via Bayes rule as:

$$P(p(r), \theta_{f(r)} | \text{Motion}(r)) = \frac{1}{P(\text{Motion})} P(\text{Motion}(r) | p(r), \theta_{f(r)}) P_p(p) P_\theta(\theta),$$

where $P(\text{Motion})$ is a normalizing constant. We can write the joint conditional probability

$$P(\text{Motion}(r), f(r)|p(r), \theta_{f(r)}) = P(\text{Motion}(r)|f(r), p(r), \theta_{f(r)})P(f(r)|p(r), \theta_{f(r)}).$$

So, we can obtain $P(\text{Motion}(r)|p(r), \theta_{f(r)})$ by marginalizing over all possible regions. Thus,

$$P(\text{Motion}(r)|p(r), \theta_{f(r)}) = \sum_{k=1}^M P(\text{Motion}(r)|(f(r) = k), p(r), \theta_{f(r)})P((f(r) = k)|p(r), \theta).$$

This can be rewritten as,

$$P(\text{Motion}(r)|p(r), \theta_{f(r)}) = \sum_{k=1}^M v_k(r)p_k(r) = v(r).p(r), \quad (2.6)$$

where

$$v_k(r) = \sqrt{\frac{\gamma}{\pi}} \exp(-\gamma|I_1(r) - I_2(r + \theta_{f(r)})|^2)$$

is the additive Gaussian noise in region R_k and is assumed to be known. The term $v_k(r)$ is obtained by replacing ‘a’ with $I_1(r)$ and ‘b’ with $I_2(r + \theta_{f(r)})$ in Equation 2.4.

The posterior distribution can be written as

$$P(p, \theta | \text{Motion}) = \frac{1}{Z} \exp[-U(p, \theta)],$$

where $U(p, \theta)$ is the posterior energy:

$$U(p, \theta) = - \sum_{r \in L} \log(p(r).v(r)) + \Sigma_C W_C(p) - \log(P(\theta)).$$

$$\frac{\partial U(p, \theta)}{\partial \theta_k} = - \sum_{r \in L} \frac{2\gamma v_k(r)p_k(r)(I_1(r) - I_2(r + \theta_k))(-I_2(r + \theta_k))}{\sum_{k=1}^M v_k(r)p_k(r)} - \frac{1}{p_k(\theta)} \quad (2.7)$$

$$\frac{\partial U(p, \theta)}{\partial p_k(r)} = - \sum_{r \in L} \frac{v_k(r)}{\sum_{k=1}^M v_k(r)p_k(r)} + \sum_C \lambda \sum_{k=1}^M 2(p_k(r) - p_k(s)) \quad (2.8)$$

The MAP estimators p^* and θ^* for p, θ respectively are those which maximize posterior distribution (or minimize the posterior energy). For the HMMF model, the posterior energy is a differentiable function as can be seen from Equations 2.7, 2.8 which can be minimized using a gradient descent method. To find the MAP estimators

1. Optimize w.r.t to θ using a modified gradient descent ($\frac{d\theta}{dt} \approx -\nabla_{\theta} U(p, \theta)$)

2. Optimize w.r.t to p using a modified gradient descent ($\frac{dp}{dt} \approx -\nabla_p U(p, \theta)$)
3. Modify p so that the following constraints are satisfied. The constraints are
 - (a) Sum across all the probability fields should be 1 ($\sum_{r \in L} p(r) = 1$)
 - (b) Each probability field should lie in the range 0 and 1 ($0 \leq p(r) \leq 1$)

Below, we describe the modified gradient descent rule proposed by [35]. Consider the unknowns, x , to be the position of a particle of unit mass. Let the acceleration of the particle be given by

$$\ddot{x} = -\nabla U(x) - 2\alpha\dot{x} \quad (2.9)$$

- First term : the particle is being forced down hill (direction of decreasing U)
- Second term : restrict the speed of decent (friction or damping)

From kinematics we know that

$$x^{t+h} = x^t + \dot{x}^t h + \frac{1}{2} \ddot{x}^t h^2 \quad (2.10)$$

where x, \dot{x}, \ddot{x} are the position, velocity and acceleration of the particle. Substituting the given acceleration we have

$$x^{t+h} = x^t + \dot{x}^t h + \frac{1}{2} (-\nabla U(x^{(t)}) - 2\alpha\dot{x}^{(t)}) h^2 \quad (2.11)$$

Rearranging all the terms the final evolution term when discretized is given by

$$x^{t+h} = \frac{2}{\alpha h + 1} x^{(t)} + \frac{\alpha h - 1}{\alpha h + 1} x^{(t-h)} - \frac{h^2}{\alpha h + 1} (\nabla U(x^{(t)})) \quad (2.12)$$

This idea is applied to θ and p for minimizing the energy function (replace x with θ and p).

$$\theta^{t+h} = \frac{2}{\alpha h + 1} \theta^{(t)} + \frac{\alpha h - 1}{\alpha h + 1} \theta^{(t-h)} - \frac{h^2}{\alpha h + 1} (\nabla U(\theta^{(t)})) \quad (2.13)$$

$$p^{t+h} = \frac{2}{\alpha h + 1} p^{(t)} + \frac{\alpha h - 1}{\alpha h + 1} p^{(t-h)} - \frac{h^2}{\alpha h + 1} (\nabla U(p^{(t)})) \quad (2.14)$$

We observe that while updating the ‘ p ’ term the values may not be in the interval $[0, 1]$. To account for this we normalize the probability fields making sure that they lie in the range $[0, 1]$.

$$p_k(r) = \frac{p_k(r)}{\sum_{k=1}^M p_k(r)} \quad (2.15)$$

The results of this technique is discussed in the next paragraph and also these results illustrate the performance of the method. For ease we have used only Gaussian noise models but any other model (e.g. Rayleigh noise) can be incorporated by simply modifying the definition of the likelihood term in Equation (2.4).

“A robust behavior, both with respect to noise and with respect to initialization, is obtained with a reasonable computational cost. The enhanced performance is due to the nonlinear data term $-\sum_{r \in L} \log(p(r).v(r))$ which is used as $-\sum_{r \in L} \log(v(r))$. This term along with the quadratic potential function $\lambda \sum_{k=1}^M (p_k(r) - p_k(s))^2$ permits the energy function to have a good balance between two opposing tendencies. The data term pushes each distribution $p(r)$ towards a low entropy configuration and the potential term acts as a diffusion and hence tends to produce high entropy configurations [35]. The balance between these two, allows the solution to evolve from an initial state to a final low entropy configuration at an appropriate rate so that the model parameters θ can escape from local minima at the beginning when the segmentation is induced and be optimally adjusted at the end” [35].

In our implementation we divided each frame into 8 regions, and the k^{th} region was assigned the model parameter θ_k . The values of θ_k were randomly chosen initially and, as noted earlier, the HMMF technique does not depend on the initial θ values. Figure 2.4 shows the initial set of ‘p’ values. These set of ‘p’ values are generated randomly in the range [0, 1]. Two frames are being considered for estimating the motion between them. Figure 2.5 shows these initial two frames. The final probability distribution is shown in Figure 2.6. Once the final probability distribution is obtained this output is binarized using a threshold value. The threshold value for all the experiments we conducted is set to 0.1 (i.e. all the values of ‘p’ greater than 0.1 are set to 1 and all the values of ‘p’ less than 0.1 are set to 0). Once we binarize the final ‘p’ distribution we use the canny edge detector to obtain the contour of the moving objects. The output of the contour overlapped on the original set of frames is shown in Figure 2.7.

In the videos we collected, there were shadows around moving objects due to natural lighting. Figures 2.8 shows where there were considerable shadows of the moving object (human). The change in illumination is not properly handled in case of the HMMF technique i.e. when the shadows are present in one frame and not present in the other one it treats the shadows as moving objects, as illustrated in Figure 2.9. As can be seen from the output, even the shadow of the moving object (human) is considered as a region of motion.

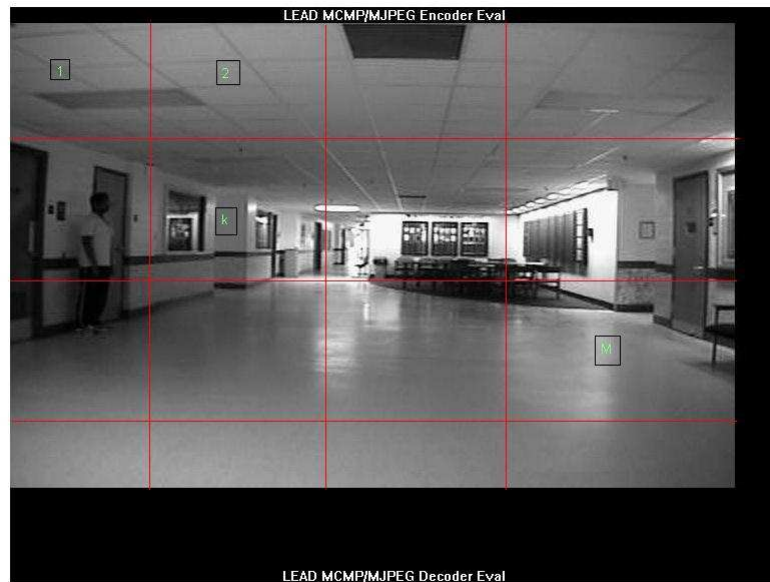


Figure 2.3: Initial K classes.

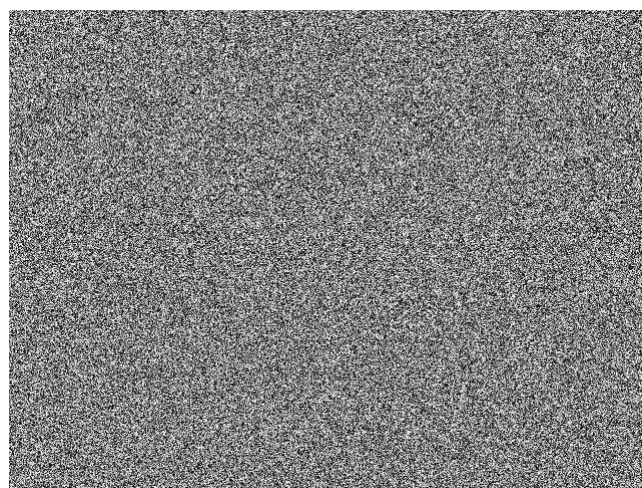


Figure 2.4: Initial probability distribution, $p(r)$, on a pixel lattice.



(a) Frame 1



(b) Frame 2

Figure 2.5: The two frames considered for estimating motion using HMMF.



Figure 2.6: Final probability distribution on the same lattice.

2.2 Tracking

Once the regions corresponding to the moving objects are obtained in the video sequence, these objects are tracked across different frames. In our case, instead of tracking the entire human we decompose the object into different components, viz., head, torso and legs and track these individual parts. Let B_1 , B_2 , B_3 be the three regions, each one corresponding to head, torso, legs respectively. We want to determine blocks B'_1 , B'_2 , B'_3 in second frame. Based on the segmented output of the HMMF procedure, we manually mark the three regions B_1 , B_2 , B_3 in one of the frames (say test frame), usually it is the first frame of the video sequence. Once these regions are marked manually, the histograms for these regions are obtained. We employ a histogram matching technique to track these regions in the corresponding frames. Since we have a prior knowledge of the direction of motion of the human in the video sequence (subject moving horizontally) we crop the vertical portions that do not correspond to motion from the output of the HMMF. In the retained portion, we calculate the histograms of intensities of the regions B_1 , B_2 , B_3 (the size of which is defined in the manual marking stage). Chi-square distance is used as a metric to compare the similarity between these two histograms.

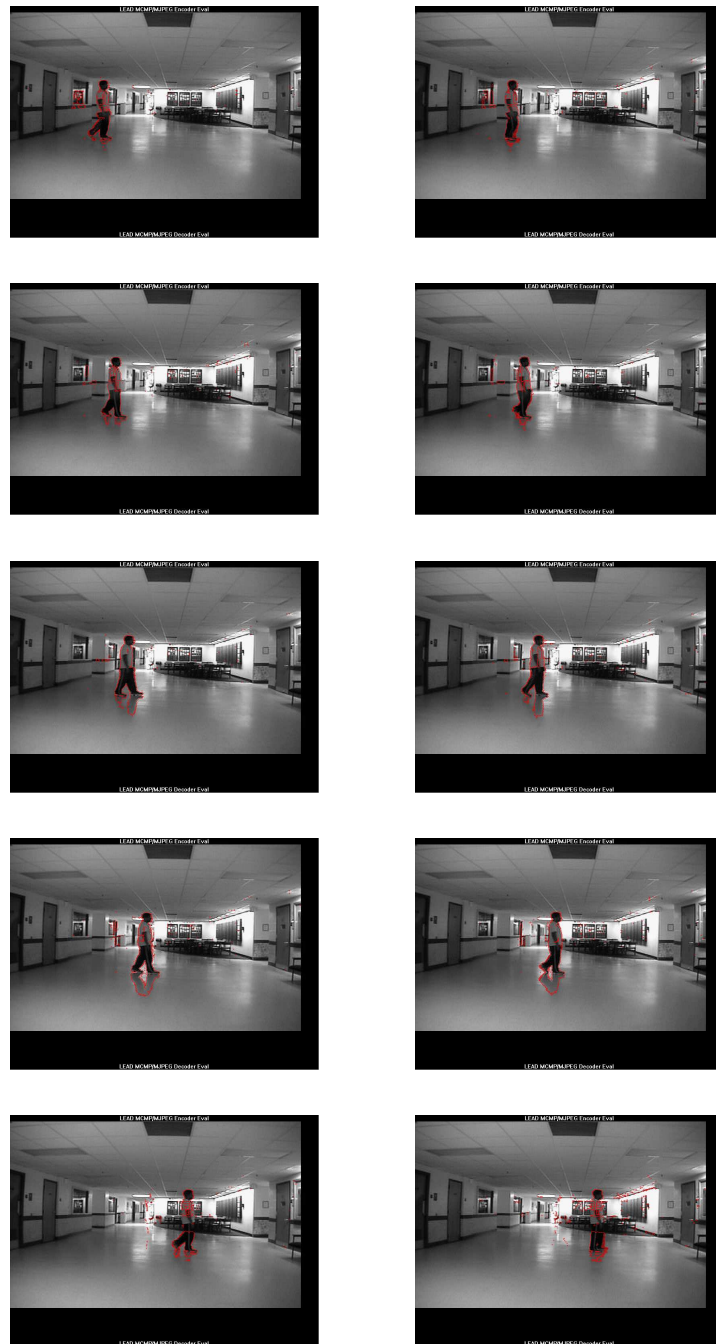


Figure 2.7: Output of the HMMF technique with the contours of the moving objects (human) on the original set of frames.



(a) Frame 1



(b) Frame 2

Figure 2.8: Shadows of the moving object (human) present in the video frames.



Figure 2.9: Output of the HMMF technique when there was significant amount of illumination change.

$$Chi - square = \chi^2 = \sum_{i=1}^M \frac{h_1(i) - h_2(i)}{h_1(i) + h_2(i)}^2 \quad (2.16)$$

M corresponds to the number of bins and in our case it is the range of gray scale intensity values (0-255). h_1 and h_2 are the histograms of blocks across two different frames. The videos considered for the experiments contain color information (RGB color space). We make use of the information in all the three channels. The above procedure of calculating the histograms of intensities of the three regions is performed in each of these channels separately (i.e. Red channel, Green channel, Blue channel). We then calculate the χ^2 values between test frame and the consequent frames of the video sequence for each of the blocks B_1, B_2, B_3 . For each block, there are three different χ^2 values (across R, G, B channels). All these values are then averaged out. The lowest χ^2 value is then retained for each of the three regions and the block that corresponds to the lowest value is output as the region of correspondence for each of the regions in the consecutive frames ($\chi^2(B_i, B'_i)$ is minimum). For robustness we assume that the torso portion is tracked properly, and depending on the relative position of the torso, we rely on the tracked portion of the head and legs.

The results of the technique discussed above are as shown in figure 2.10.

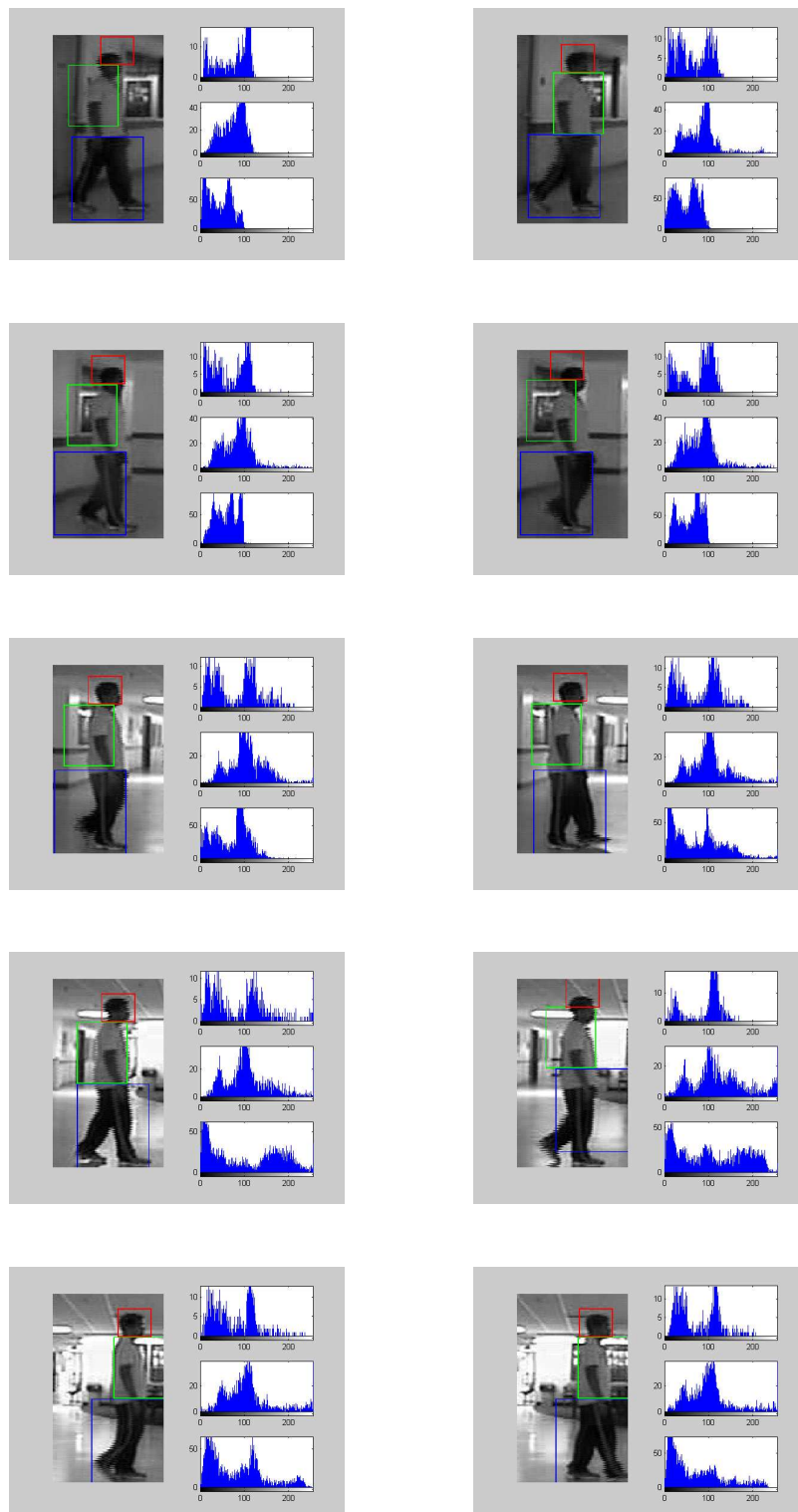


Figure 2.10: Moving objects tracked across frames along with their intensity histograms.

2.3 Active Contours

As can be seen from the output of the HMMF model, a precise contour of the human has not yet been obtained. For this reason, we further process the video by using the space “Active Contours” [43] technique. Specifically, we use the technique proposed by Chan and Vese [44]. The idea in using the active contour models or snakes is to evolve a curve or surface (surface in this thesis as we evolve in 3D space), in such a way that it is subject to image or video constraints, in order to detect objects in the image or moving objects in a video sequence. For instance, starting with a surface around the object (to be detected) in the video, the surface evolves towards the objects interior normal and stops (not always) on the boundary of it.

2.3.1 Description of the Model

Let us define the evolving surface C in Ω , as the boundary ω of Ω (i.e. $\omega \subset \Omega$, and $C = \partial\omega$). $inside(C)$ denotes the region ω , and $outside(C)$ denotes the region $\Omega \setminus \bar{\omega}$.

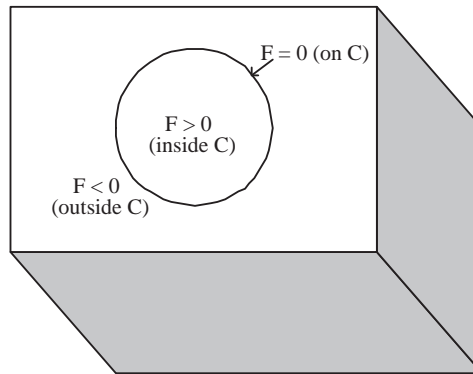


Figure 2.11: Evolving Surface C .

Assume that the video (set of frames) u_0 is formed by two regions, both of which are approximately piecewise-constant intensities, of distinct values u_0^i and u_0^o . In addition let us assume that the object (object in our case refers to portions corresponding to moving objects) to be detected is represented by the region with the value u_0^i , and the boundary is denoted by C_0 . Consider the following term

$$F_1(C) + F_2(C) = \int_{inside(C)} |u_0(x, y, t) - c_1|^2 dx dy dt + \int_{outside(C)} |u_0(x, y, t) - c_2|^2 dx dy dt$$

where C is any variable surface, and the constants c_1, c_2 , depending on C , are the averages of u_0 inside C and outside C , respectively. The boundary of the object is the minimizer to the above equation. In [44], the authors in addition to minimizing the above equation add some regularizing terms, like the length of the curve etc.. Therefore the above energy functional $F(c_1, c_2, C)$ is modified as

$$\begin{aligned} F(c_1, c_2, C) &= \mu.Length(C) + \\ &\lambda_1 \int_{inside(C)} |u_0(x, y, t) - c_1|^2 dx dy dt + \\ &\lambda_2 \int_{outside(C)} |u_0(x, y, t) - c_2|^2 dx dy dt \end{aligned}$$

where $\mu \geq 0, \lambda_1, \lambda_2 > 0$ are fixed parameters.

The Mumford-Shah functional for segmentation is [45]

$$\begin{aligned} F^{MS}(u, C) &= \mu.Length(C) + \\ &\lambda \int_{\Omega} |u_0(x, y, t) - u(x, y, t)|^2 dx dy dt + \\ &\int_{\Omega \setminus C} |\nabla(u(x, y, t))|^2 dx dy dt \end{aligned}$$

where μ and λ are positive parameters. The solution video u is obtained by minimizing this functional. The minimum partition problem as it is called is a reduced form by simply restricting the values of c_i to averages as u_0 on each side of the surface.

The Chan and Vese model is a particular case of the minimal partition problem, in which the best approximation is obtained u of u_0 as a function of only two values, namely

$$u = \begin{cases} average(u_0) & inside \ C \\ average(u_0) & outside \ C \end{cases} \quad (2.17)$$

and with one edge C , represented by the snake or the active contour, and this case of the minimal partition problem can be formulated and solved using the level set method [46].

More details of the method related to level sets can be obtained from [46]. The next set of equations is the formulation for solving the minimum partition problem. In the level set method [46], $C \subset \Omega$ is represented by the zero level set of a Lipschitz function such that,

$$\left\{ \begin{array}{l} C = \partial\omega = (x, y, t) \in \Omega : \phi(x, y, t) = 0 \\ inside(C) = \omega = (x, y, t) \in \Omega : \phi(x, y, t) > 0 \\ outside(C) = \omega \setminus \bar{\omega} = (x, y, t) \in \Omega : \phi(x, y, t) < 0 \end{array} \right.$$

For the level set formulation of variational active contour model, the unknown variable C is replaced by the embedding function ϕ (e.g. signed distance function).

Using the Heaviside function H , and the one-dimensional Dirac measure δ_0 , which are defined, respectively, as

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0, \\ 0, & \text{if } z < 0, \end{cases}$$

$$\delta_0(z) = \frac{d}{dz}H(z), \quad (2.18)$$

the terms in the energy are expressed in the following way:

$$\begin{aligned} Length\{\phi = 0\} &= \int_{\Omega} |\nabla H(\phi(x, y, t))| dx dy dt \\ &= \int_{\Omega} \delta_0(\phi(x, y)) |\nabla \phi(x, y)| dx dy dt, \\ Area\{\phi \geq 0\} &= \int_{\Omega} H(\phi(x, y)) dx dy dt \end{aligned}$$

and

$$\begin{aligned} \int_{\phi > 0} |u_0(x, y, t) - c_1|^2 dx dy dt &= \int_{\Omega} |u_0(x, y, t) - c_1|^2 H(\phi(x, y, t)) dx dy dt, \\ \int_{\phi < 0} |u_0(x, y, t) - c_2|^2 dx dy dt &= \int_{\Omega} |u_0(x, y, t) - c_2|^2 (1 - H(\phi(x, y, t))) dx dy dt, \end{aligned}$$

Then, the energy $F(c_1, c_2, \phi)$ can be written as

$$\begin{aligned} F(c_1, c_2, \phi) &= \mu \int_{\Omega} \delta(\phi(x, y, t)) |\nabla \phi(x, y, t)| dx dy dt + \\ &\quad \lambda_1 \int_{\Omega} |u_0(x, y, t) - c_1|^2 H(\phi(x, y, t)) dx dy dt + \\ &\quad \lambda_2 \int_{\Omega} |u_0(x, y, t) - c_2|^2 (1 - H(\phi(x, y, t))) dx dy dt \end{aligned} \quad (2.19)$$

u as defined in equation 2.17 is solution to the model as a particular case of the Mumford-Shah minimal partition problem, and can be written using the level set formulation as

$$u(x, y, t) = c_1 H(\phi(x, y, t)) + c_2 (1 - H(\phi(x, y, t))), \quad (x, y, t) \in \bar{\Omega}$$

Here the authors are minimizing the energy functional with respect to the two constant c_1 and c_2 keeping ϕ fixed, in which case the values of c_1 and c_2 are given by,

$$\begin{cases} c_1(\phi) = \text{average}(u_0) & \phi \geq 0 \\ c_2(\phi) = \text{average}(u_0) & \phi < 0 \end{cases} \quad (2.20)$$

The next set of equations indicate how the values of c_1 and c_2 as indicated in Equation 2.20 are defined by. The minimum $\min_c \sum_{i=1}^n (c - x_i)^2$ of what can be found by differentiating with respect to 'c'. This becomes

$$\frac{d}{dc} \sum_{i=1}^n (c - x_i)^2 = \sum_{i=1}^n 2(c - x_i) = 0, = 2(nc - \sum_{i=1}^n x_i) = 0$$

So, $c = \frac{1}{n} \sum_{i=1}^n nx_i$ which is nothing but the term given by Equation 2.20.

Once we obtain the values of c_i 's for the piece-wise constant case, the minimum of the function F with respect to ϕ can be obtained by keeping the values of c_i 's constant, the associated Euler-Lagrange equation for ϕ is obtained.

$$\frac{\partial \phi}{\partial \text{iter}'} = \delta_\epsilon(\phi) [\mu \text{div}(\frac{\nabla \phi}{|\nabla \phi|}) - \lambda_1 (u_0 - c_1)^2 + \lambda_2 (u_0 - c_2)^2] \quad (2.21)$$

The term $-\text{div}(\frac{\nabla \phi}{|\nabla \phi|})$ is called the curvature term and is usually represented by κ . The curvature term κ is discretized based on the up-wind finite differences.

$$\kappa = - \frac{\psi_{xx}\psi_y^2 - 2\psi_x\psi_y\psi_{xy} + \psi_{yy}\psi_x^2 + \psi_{xx}\psi_z^2 - 2\psi_x\psi_z\psi_{xz} + \psi_{zz}\psi_x^2 + \psi_{zz}\psi_y^2 - 2\psi_z\psi_y\psi_{zy} + \psi_{yy}\psi_z^2}{(\psi_x^2 + \psi_y^2 + \psi_z^2)^{\frac{3}{2}}} \quad (2.22)$$

where ψ_x is the gradient of the image in the x direction; ψ_y is the gradient in the y direction; ψ_z is the gradient in the z direction; ψ_{xx} is the 2nd order gradient in the x direction; ψ_{yy} is the 2nd order gradient in the y direction; ψ_{zz} is the 2nd order gradient in the z direction; ψ_{xy} is the 2nd order gradient, first in the x direction and then in the y direction; ψ_{yz} is the 2nd order gradient, first in the y direction and then in the z direction; ψ_{xz} is the 2nd order gradient, first in the x direction and then in the z direction. The first term ' κ ' provides the smoothing constraints on

the evolving surface by reducing the total curvature of the evolving surface. The second term acts like a balloon force and it pushes the surface towards the object boundary. c_1 and c_2 are the means of the values defined by the embedding function ϕ . u_0 is the original image (set of frames in our case) $I(x, y, f)$. λ_1 and λ_2 are constants. The stopping criteria in our case is the number of iterations, i.e., the evolution of the curve, will terminate after a certain number of predefined iterations.

The terms λ_1 and λ_2 are both set to 1. The step length ‘iter’ is set to 0.01. The term μ is set to 2. The values were fixed for all the experiments that were conducted. The results of the algorithm discussed above are as shown in Figure 2.13.

The initial ψ function is a cylinder whose center coincides with the center of the block of the individual components. The final evolution surface of the head, torso, legs is shown in Figure 2.14.

2.4 Measurements

As mentioned earlier, the next step is to extract the measurements of the human body from the segmented results. For this, we fit ellipses to the segmented regions: one for the head and one for the torso. The reason for choosing the ellipses is that we can get both the measurements (i.e., the width and the height of the body components) simultaneously. For this we use the technique proposed by Radim [47]. As said in the previous section a contour is obtained for each of the body components. We get the coordinates (x, y) of the contour in the image plane, and fit an ellipse to them. A brief description of the ellipse fitting technique is presented next.

An ellipse is a special case of a general conic which can be described by an implicit second order polynomial

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0$$

with an ellipse-specific constraint which is $b^2 - 4ac < 0$ where a, b, c, d, e, f are coefficients of the ellipse and (x, y) are coordinates of points lying on it. The polynomial $F(x, y)$ is called the *algebraic distance* of the point (x, y) to the given conic. By introducing vectors

$$\mathbf{a} = [a, b, c, d, e, f]^T$$

$$\mathbf{x} = [x^2, xy, y^2, x, y, 1]$$

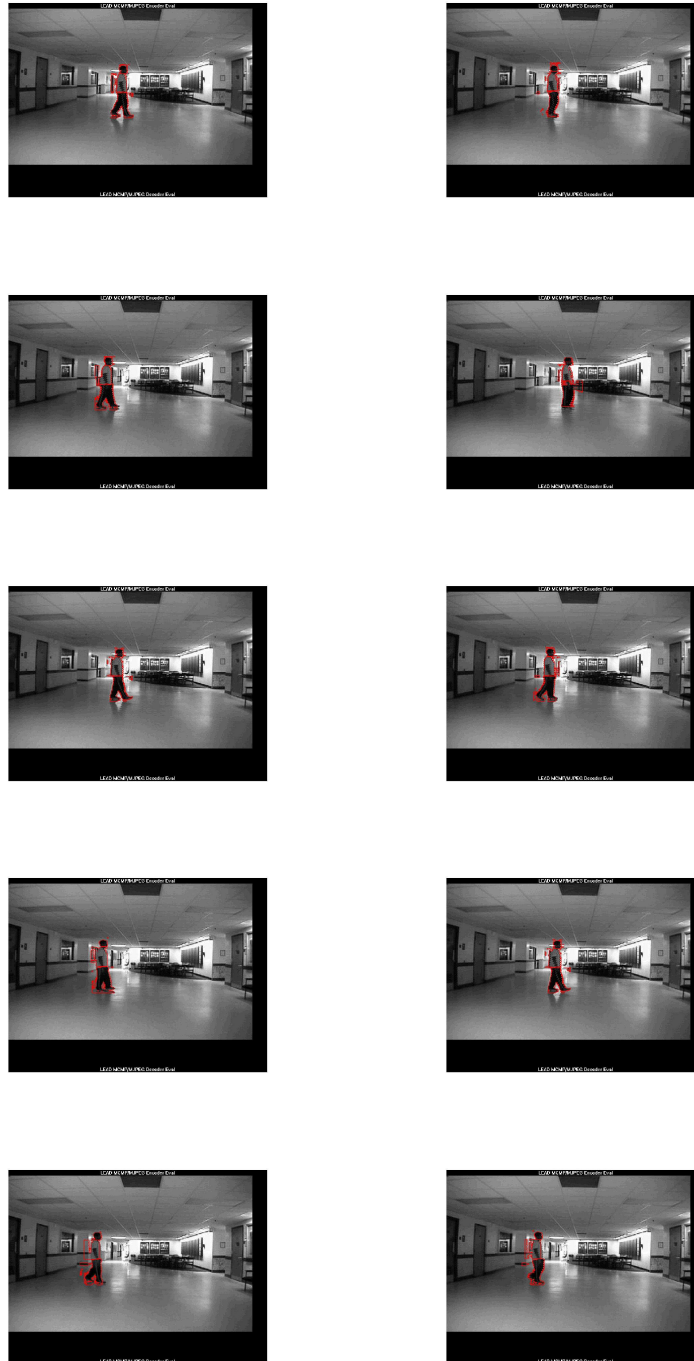


Figure 2.12: The contour of the human body obtained from using “Active Contours” overlapped on a video sequence.

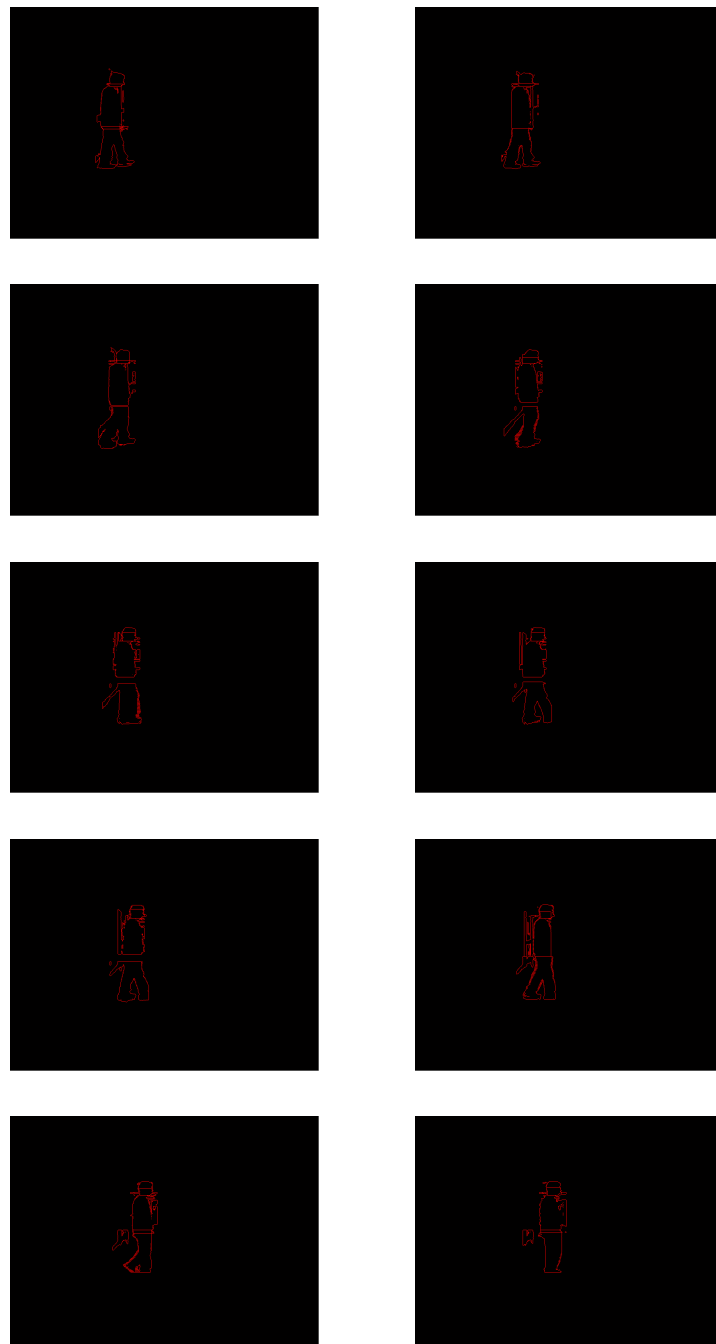
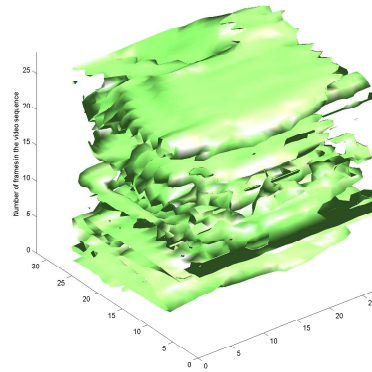
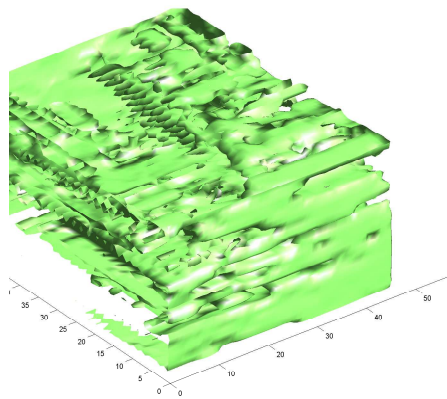


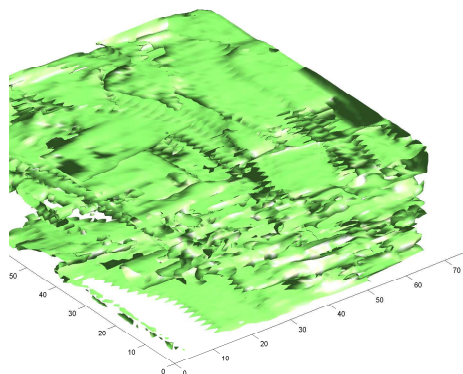
Figure 2.13: Contour of the entire human body obtained by combining all the three techniques on a video sequence.



(a) head



(b) torso



(c) legs

Figure 2.14: The final evolution curve (a) head, (b) torso, (c) legs

it can be rewritten to the vector form $F_a(\mathbf{x}) = \mathbf{x} \cdot \mathbf{a} = 0$. The fitting of a general conic to a set of points $(x_i, y_i), i = 1, \dots, N$ may be approached by minimizing the sum of the squared algebraic distances of the points to the conic which is represented by coefficients \mathbf{a} :

$$\min \sum_{i=1}^N F(x_i, y_i)^2 = \min \sum_{i=1}^N F_{\mathbf{a}}(\mathbf{x}_i)^2 = \min \sum_{i=1}^N (\mathbf{x}_i \cdot \mathbf{a})^2$$

The equation can be directly solved by the standard least squares approach, but the result of such fitting is a general conic and it need not be an ellipse. To ensure an ellipse-specificity of the solution, the appropriate constraint $b^2 - 4ac < 0$ has to be considered. In Fitzgibbon's paper [48] it was shown that such a system is hard to solve. Under a proper scaling, the inequality constraint can be changed into an equality constraint $4ac - b^2 = 1$ and the ellipse-specific fitting problem can be reformulated as

$$\min \|\mathbf{D}\mathbf{a}\|^2 \text{ subject to } \mathbf{a}^T \mathbf{C}\mathbf{a} = 1$$

where the *design matrix* \mathbf{D} of size $N \times 6$,

$$\mathbf{D} = \begin{pmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_N^2 & x_N y_N & y_N^2 & x_N & y_N & 1 \end{pmatrix}$$

represents the least square minimization and the *constraint matrix* \mathbf{C} of size 6×6 ,

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

expresses the constraint $4ac - b^2 = 1$. The minimization problem is solved by a quadratically constrained least square minimization. First, by applying the Lagrange multipliers we get the

following conditions for the optimal solution \mathbf{a}

$$\mathbf{S}\mathbf{a} = \lambda\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{C}\mathbf{a} = 1$$

where \mathbf{S} is the *scatter matrix* of size 6×6 ,

$$\begin{aligned} \mathbf{S} &= \mathbf{D}^T\mathbf{D} \\ &= \begin{pmatrix} S_{x^4} & S_{x^3y} & S_{x^2y^2} & S_{x^3} & S_{x^2y} & S_{x^2} \\ S_{x^3y} & S_{x^2y^2} & S_{xy^3} & S_{x^2y^2} & S_{xy^2} & S_{xy} \\ S_{x^2y^2} & S_{xy^3} & S_{y^4} & S_{xy^2} & S_{y^3} & S_{y^2} \\ S_{x^3} & S_{x^2y} & S_{xy^2} & S_{x^2} & S_{xy} & S_x \\ S_{x^2y} & S_{xy^2} & S_{y^3} & S_{xy} & S_{y^2} & S_y \\ S_{x^2} & S_{xy} & S_{y^2} & S_x & S_y & 1 \end{pmatrix} \end{aligned}$$

in which the operator S , denotes the sum

$$S_{x^a y^b} = \sum_{i=1}^N x_i^a y_i^b$$

The equation $\mathbf{a}^T\mathbf{C}\mathbf{a} = 1$ is solved by using generalized eigenvectors. There exist up-to six real solutions $(\lambda_j, \mathbf{a}_j)$, but because

$$\|\mathbf{D}\mathbf{a}\|^2 = \mathbf{a}^T\mathbf{D}^T\mathbf{D}\mathbf{a} = \mathbf{a}^T\mathbf{S}\mathbf{a} = \lambda\mathbf{a}^T\mathbf{C}\mathbf{a} = \lambda$$

we are looking for the eigenvector \mathbf{a}_k corresponding to the minimal positive eigenvalue λ_k . After a proper scaling ensuring $\mathbf{a}_k^T\mathbf{C}\mathbf{a}_k = 1$, we get a solution of the minimization problem which represents the best-fit ellipse for the given set of points.

Since we decompose the human into three regions and there is no notion of width and height for the legs (as we define the entire legs region as a whole) we all together have five measurements. The four measurements are width and the height of the head, torso. Based on this algorithm we measure the width and the height of the body parts. The results for one of the video sequences are shown as bar graphs

The overlapped ellipse on the original frames is as shown in Figure 2.15:



Figure 2.15: Overlapped ellipses on the video frames.

	Head Width					
	A(12ft)	E(12ft)	A(24ft)	E(24ft)	A(36ft)	E(36ft)
1	51	54	22	23.06	17	16.45
2	47	51.2	20	19.13	16	15.47
3	48	53.28	18	19.91	15	15.79
4	40	42.17	20	20.14	10	10.23
5	40	40.92	21	20.41	11	11.67
6	43	46.74	20	18.19	12	11.92
7	40	42.34	20	21.19	12	14.72
8	35	37.96	19	20.63	11	11.92
9	38	40.12	19	21.78	12	13.97

Table 2.1: Comparison of actual (A) vs estimated (E) values in terms of pixels.

	Head Height					
	A(12ft)	E(12ft)	A(24ft)	E(24ft)	A(36ft)	E(36ft)
1	43	42.4	21	20.65	13	12.79
2	40	43.44	20	21.69	16	16.73
3	39	46.51	16	15.12	11	11.71
4	40	38.16	20	19.07	14	14.32
5	37	35.14	20	19.79	14	14.93
6	40	42.17	20	21.13	11	11.41
7	40	43.59	20	21.68	10	11.91
8	40	44.92	18	20.17	11	11.17
9	43	48.22	19	22.05	11	13.01

Table 2.2: Comparison of actual (A) vs estimated (E) values in terms of pixels.

	Torso Width					
	A(12ft)	E(12ft)	A(24ft)	E(24ft)	A(36ft)	E(36ft)
1	50	57.76	21	22.36	20	21.19
2	75	84.83	46	50.26	29	30.12
3	50	80.22	40	43.18	25	27.03
4	60	57.15	29	32.09	18	17.01
5	49	46.43	22	24.12	17	18.31
6	54	58.96	21	19.16	15	16.09
7	50	51.17	20	22.43	16	16.94
8	40	44.39	20	21.19	16	16.79
9	43	47.10	21	24.28	17	20.01

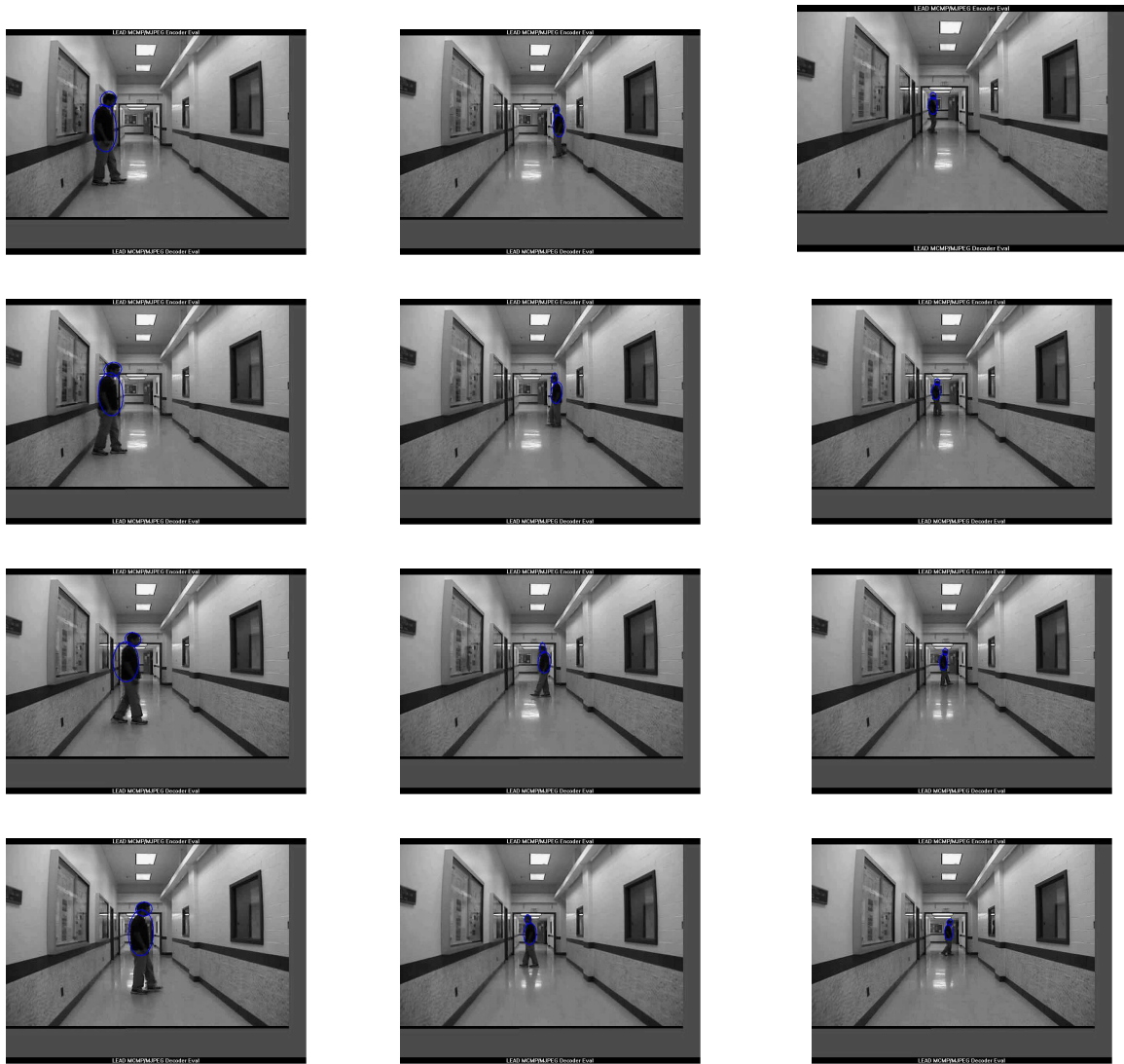
Table 2.3: Comparison of actual (A) vs estimated (E) values in terms of pixels.

	Torso Height					
	A(12ft)	E(12ft)	A(24ft)	E(24ft)	A(36ft)	E(36ft)
1	84	76.09	47	49.92	31	30.25
2	105	115.09	60	66.94	40	28.25
3	110	123.85	60	56.19	35	43.69
4	110	121.19	60	64.85	40	37.17
5	105	110.96	50	54.07	33	42.69
6	103	110.19	50	52.93	32	34.09
7	95	100.05	45	49.59	40	36.39
8	100	103.27	43	47.69	36	39.19
9	102	109.27	45	50.12	38	42.19

Table 2.4: Comparison of actual (A) vs estimated (E) values in terms of pixels.

The dataset consists of 9 people and 3 video sequences of each person. The body measurements are taken across different distances. The following table gives the details of the measurements from the experiments and from manual (actual) measurements.

The measurements (height and width of the torso, head) in terms of pixels is within 10% of the manual measurements. Even though 10% error is bit high we can potentially reduce this error by finding techniques to extract measurements from the contour instead of using a simple ellipse fitting procedure. For example we can use joints corresponding to the human but the joints are not well defined in the human contour form. And this is one of the challenges in extracting measurements. The measured values as can be seen from the table are consistent across all frames and across various distances too. As the segmentation technique works well on the side view, the measurements can be obtained only with respect to side view. The results of the measurements at three distances (12ft, 24ft, 36ft) are as shown in Figure 2.17 :



(a) Measurements at 12ft

(b) Measurements at 24ft

(c) Measurements at 36ft

Figure 2.16: Ellipses overlapped on the original frames for extracting measurements at various distances.

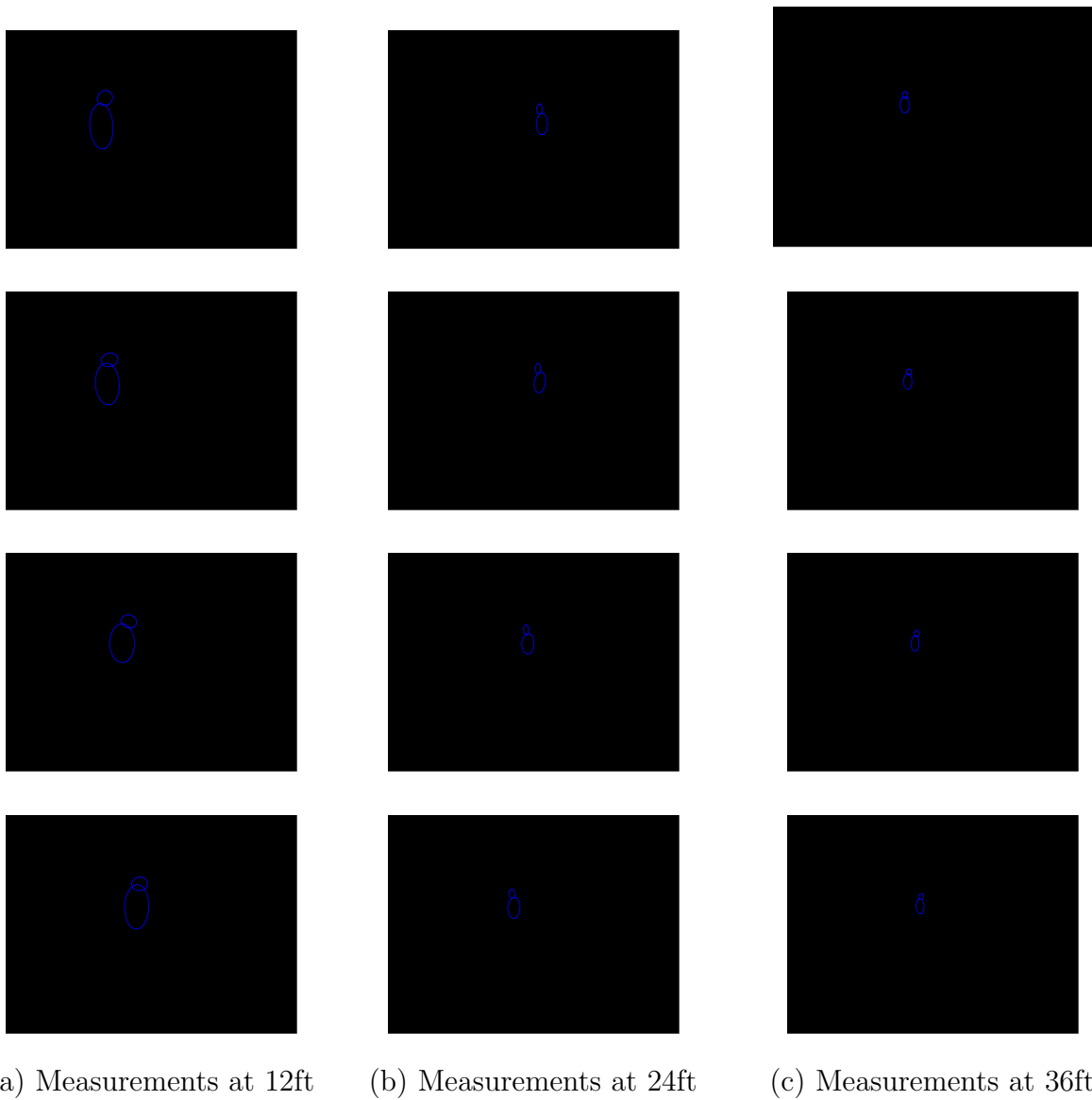


Figure 2.17: Contours of the ellipses without any background at various distances.

Chapter 3

Summary and Future work

3.1 Future Work

As we can see, we could extract only four measurements but in the future we can extract more body measurements either by obtaining a more clear and precise contour of the human or by developing a better tracking algorithm, in which we consider some key points in one of the input frames and then track these key points based on some local or global information. If we can track these key points with some degree of redundancy, then we can extract the human body measurements. The purpose of the 3d model is to map the metrological features of a subject as measured from a video sequence onto one of the generic 3D models, and then individualize the model for each subject. With respect to this, we have started modeling a human. The template was obtained from cyberware [49] web site. The model looks as shown in Figure 3.1.

Each 3D model is viewed as an agglomeration of multiple components viz., head, torso, legs, arms, etc. as shown in Figure 3.3, similar to the case when we deal with obtaining the human measurements. Based on the anthropometric data from the NASA [2] data-set, we define the shape, structure, and size of the individual components for each generic model. We then parameterize the individual components.

The representation of the components is based on polygon mesh modeling which is used to represent each component. This enables the geometry of a mesh to be manipulated in a general and powerful way. Euler operations can be used to modify vertices of the mesh. However these could be cumbersome and awkward to use. Therefore, we define a set of primitive mesh manipulation operators. The different types of operators can be listed as parallel, stretch, grow,

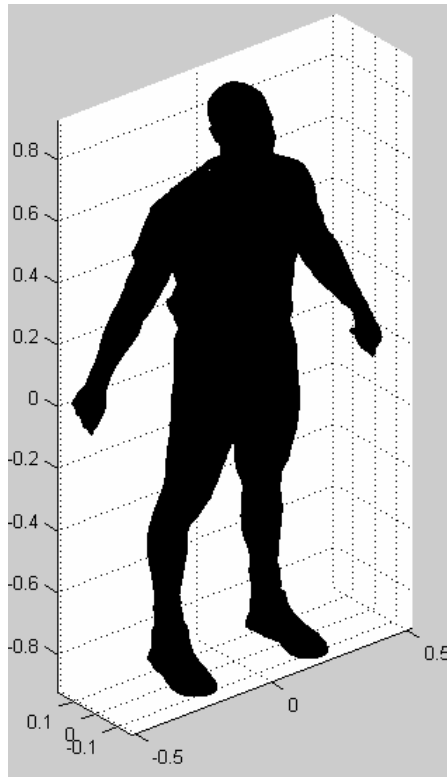


Figure 3.1: 3D start model

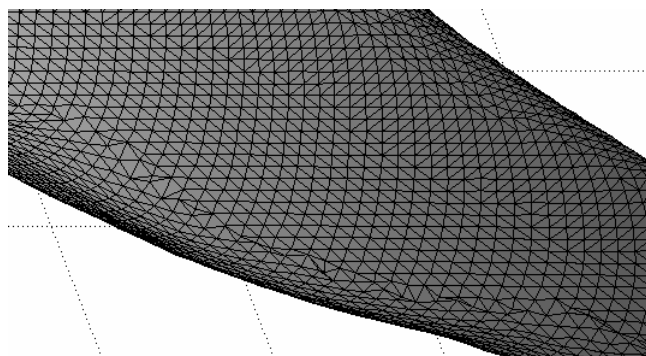


Figure 3.2: A part of the 3D model is zoomed

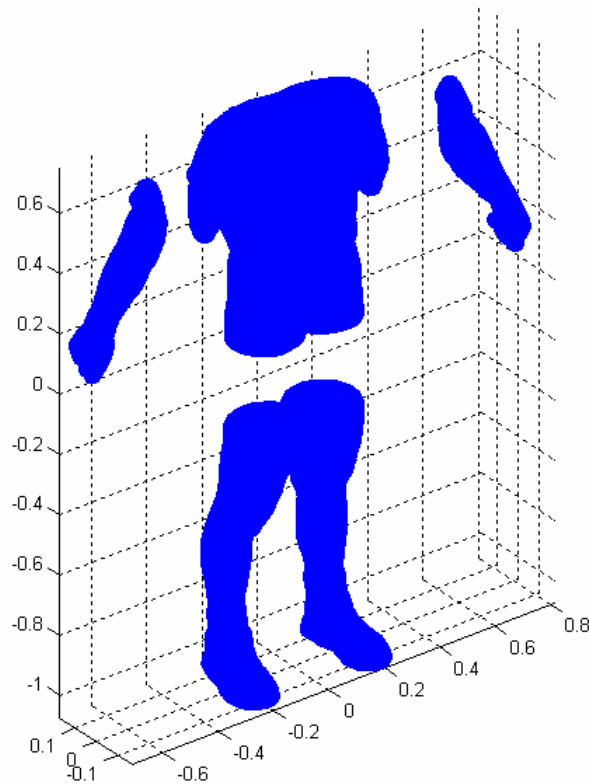


Figure 3.3: Agglomeration of the human body components

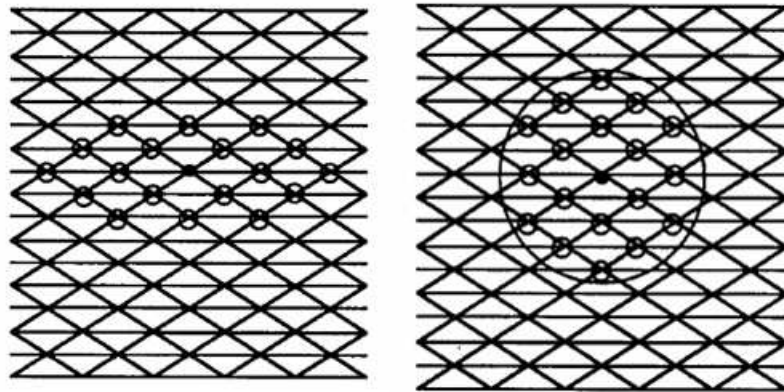


Figure 3.4: Range by number of edges and by distance (circle) (taken from [3])

randomize, and smooth operators, and for each of these operators we define parameters such as range of influence based on the number of edges or based on the range by distance (say circle) (Figure 3.4), decay functions over range of influence like, Cone decay, Bell decay, Cusp decay etc. (Figure 3.5), and by binding vertices. A decay function is one that is used to determine the amount by which each vertex moves. Together, the range of influence and decay functions offer a powerful set of mesh manipulation operators. However, they have the limitation that only vertices lying in a connected sub-mesh can be moved. Often it is necessary to move some vertices without affecting intervening ones. This type of operation is called the binding vertices operation.

The generic 3D model obtained from [49] of a male has about 148,000 points and 290,000 edges. Each component is modified independently based on the type of mesh manipulation. We use 3D transformation matrix to stitch the modified components into one single 3D entity once the components are modified. The 3D point set registration work is obtained from Paul J. Besl and Neil D. McKay [50].

Let $P = p_i$ be the data point set in one component that has to be aligned with the point set $X = x_i$ on another component. The unit quaternion $q_R = [q_0 q_1 q_2 q_3]^t$ where $q_0 \geq 0$ and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The 3×3 rotation matrix formed by the Unit quaternion is

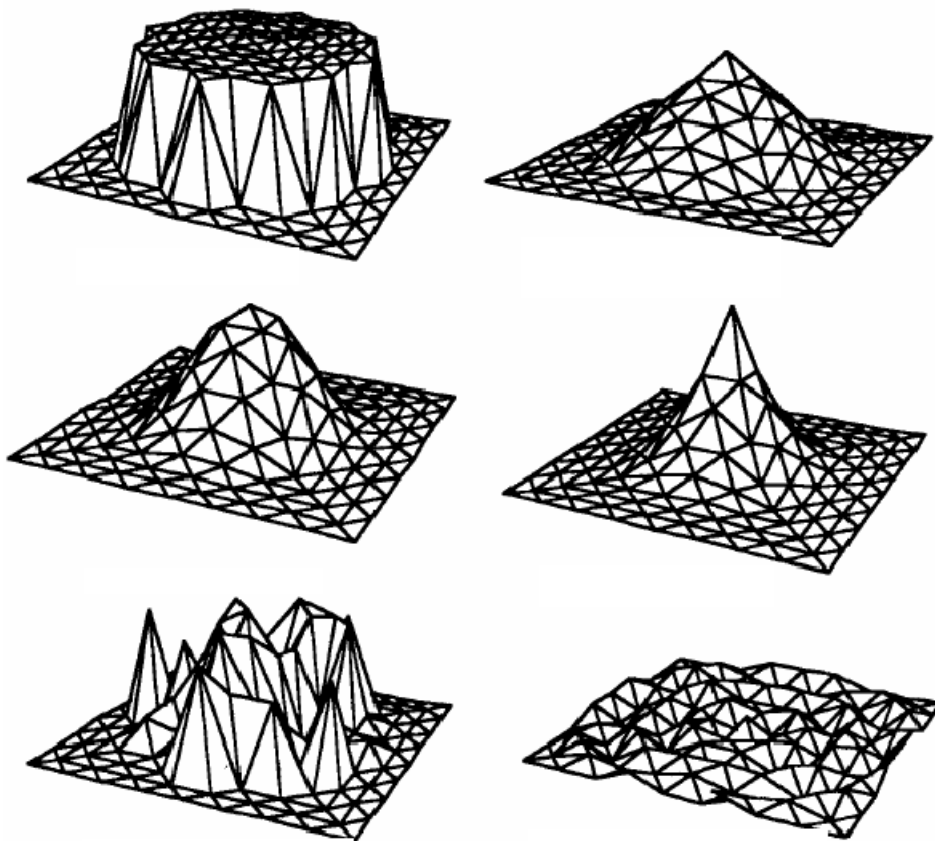


Figure 3.5: Decay Functions (a) No decay function, (b) Cone decay function, (c) Bell decay function, (d) Cusp decay function, (e) random decay function, (f) Sine wave decay function (taken from [3])

$$R = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{pmatrix} \quad (3.1)$$

and the translation vector is represented by $q_T = [q_4q_5q_6]^t$. The mean square objective function that has to be minimized to obtain the best registration of the 3D point set is given by $f(q) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|x_i - R(q_R)p_i - q_T\|^2$. μ_p and μ_x are the ‘‘center of mass’’ of point sets $P = p_i$ and $X = x_i$, respectively. $\mu_P = \frac{1}{N_p} \sum_{i=1}^{N_p} p_i$, $\mu_x = \frac{1}{N_x} \sum_{i=1}^{N_x} x_i$ where N represents the total number of points in each set. The unit eigen vector $q_R = [q_0q_1q_2q_3]^t$ corresponding to the maximum eigenvalue of the following matrix

$$Q(\sum_{px}) = \begin{pmatrix} tr(\sum_{px}) & \Delta^T \\ \Delta & \sum_{px} + \sum_{px}^T - tr(\sum_{px} I_3) \end{pmatrix} \quad (3.2)$$

is selected as the optimal rotation. The term ‘tr’ refers to Trace of a matrix and I_3 refers to an Identity matrix the subset denoting the order of the matrix. The optimal translation vector is given by $q_T = \mu_x - R(q_R)\mu_p$. The results of the above technique are used to stitch the decomposed 3D model. For example the arm part is modified in a way such that the new body part (arm) is 1.1 times the original arm length and you are trying to stitch this modified body part to the torso, the results of which are shown in Figure 3.6.

In this thesis, techniques to extract only a few set of human body measurements, namely the height, width of the head and the torso, are discussed. Further more, the videos collected have just one person walking in the scene and the subject is constrained to walk in a particular direction. But in a typical surveillance system this is not the situation. The camera in our case is stable (i.e., there is no motion of the camera). However, in a surveillance system the camera might experience some motion. The technique designed in the thesis is not fully automated. For example, during the tracking stage one has to manually mark the regions corresponding to the head, torso and legs. Thus a fully automated technique is desired. The parameters used during the implementation may not hold for all types of videos. In the future, one can use the same techniques for detecting humans in the IR spectrum in night time environments. In cases where there are multiple persons, we need to resort to new tracking techniques, since the histogram based matching scheme is not a robust technique in such scenarios. The HMMF model can still be used for motion detection, but the background has to be stable with the simple formulation

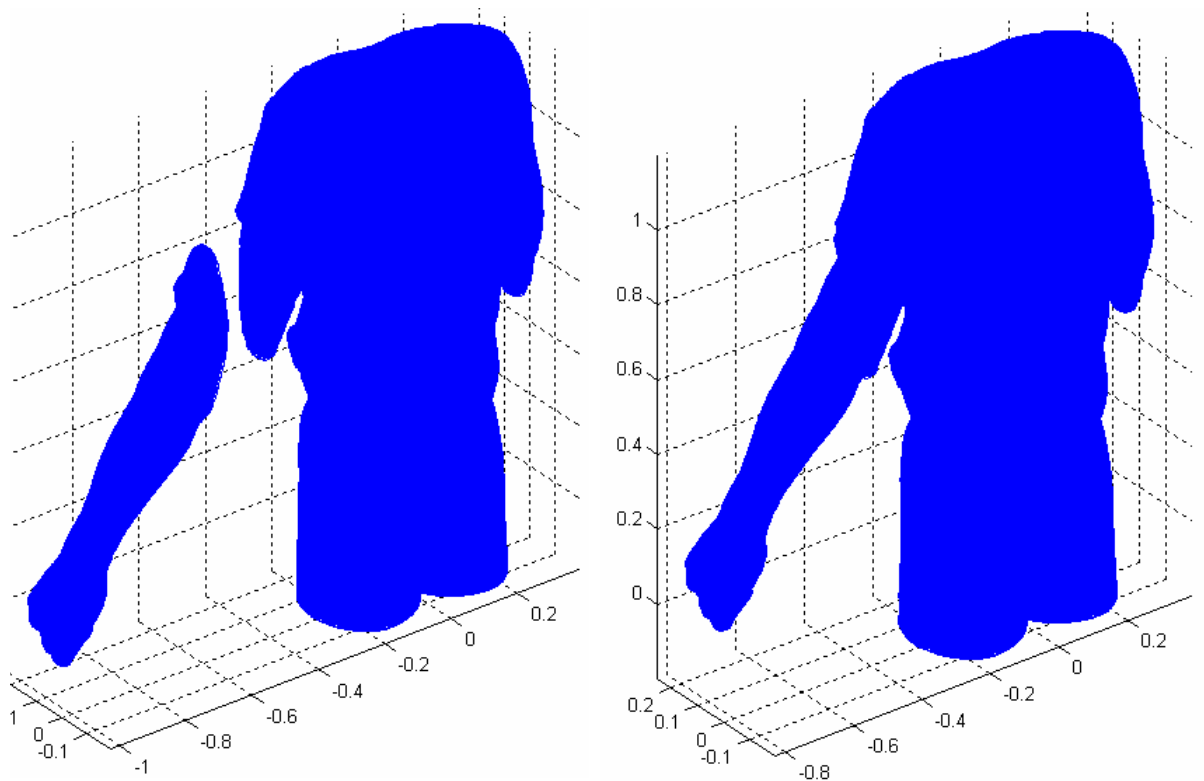


Figure 3.6: Example for 3D point registration: (a) before stitching ; (b) after stitching

that we considered. Otherwise, more complex ϕ functions can be used in the case where both the camera and objects are moving.

References

- [1] NHANES, “Anthropometric reference data,” 1988-1994.
- [2] NASA, *Anthropometric Source Book, A handbook of anthropometric data*. NASA, 1978.
- [3] J. B. Allan, B. Wyvill, and I. H. Witten, “A methodology for direct manipulation of polygon meshes,” in *Proceedings of Computer Graphics International*, November 1988, pp. 451–469.
- [4] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer, 2005.
- [5] A. K. Jain, S. C. Dass, and K. Nandakumar, “Soft biometric traits for personal recognition systems,” in *Proceedings of International Conference on Biometric Authentication*, Hong Kong, July 2004, pp. 731–738.
- [6] J. Wayman, “Large-scale civilian biometric systems - issues and feasibility,” in *Proceedings of Card Tech / Secur Tech ID*, 1997.
- [7] M. S. Nixon, T. N. Tan, and R. Chellappa, *Human Identification on Gait*. Springer, 2006.
- [8] K. Akita, “Image sequence analysis of real-world human motion,” *Computer Vision and Image Understanding*, vol. 17, no. 1, pp. 73–83, 1984.
- [9] C. Cedras and M. Shah, “Motion based recognition, a survey,” *Image and Vision Computing*, vol. 13, no. 2, pp. 129–155, 1995.
- [10] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata, “Nonrigid motion analysis: Articulated and elastic motion,” *Computer Vision and Image Understanding*, vol. 70, no. 2, pp. 142–156, 1998.
- [11] D. M. Gavrilu, “The visual analysis of human movement: A survey,” *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.
- [12] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture,” *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.
- [13] A. Godil, P. Grother, and S. Ressler, “Human identification from body shape,” in *Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM03)*.
- [14] C. American and E. S. A. R. C. web site, “<http://www.sae.org/technicalcommittees/caesumm.htm>.”
- [15] B. Horn and B. Schunck, “Determining optical flow,” in *Artificial Intelligence 17*, 1981, pp. 185–204.

- [16] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of 7th International Conference on Artificial Intelligence*, Vancouver, British Columbia, August 1981, pp. 674–679.
- [17] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biological Cybernetics*, vol. 60, pp. 79–97, 1988.
- [18] H. Nagel, "On the estimation of optical flow: Relations between different approaches and some new results," in *Artificial Intelligence 33*, 1987, pp. 299–324.
- [19] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal on Computer Vision*, vol. 2, pp. 283–310, 1989.
- [20] A. Singh, "An estimation-theoretic framework for image-flow computation," in *Proceedings of 3rd International Conference on Computer Vision (ICCV)*, Osaka, 1990, pp. 168–177.
- [21] D. Heeger, "Optical flow using spatiotemporal lters," *Intenational Journal on Computer Vision*, vol. 1, pp. 279–302, 1988.
- [22] A. Waxman, J. Wu, and F. Bergholm, "Convected activation profiles and receptive fields for real time measurement of short range visual motion," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Ann Arbor, 1988, pp. 717–723.
- [23] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1531–1536, 2004.
- [24] L. J. Latecki, R. Lakamper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 424–429.
- [25] K. Rohr, "Towards model-based recognition of human movements in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 59, no. 1, pp. 94–115, 1994.
- [26] R. Polana and R. Nelson, "Detecting activities," in *Conference on Computer Vision and Pattern Recognition*, June 1993.
- [27] J. O'Rourke and N. Badler, "Model-based image analysis of human motion using constraint propogation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 522–536, 1980.
- [28] R. F. Rashid, "Towards a system for the interpretation of moving light displays," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 574–581, 1980.
- [29] A. Koschan, S. Kang, J. Paik, B. Abidi, and M. Abidi, "Video object tracking based on extended active shape models with color information," in *Proceedings of 1st European Conference Color in Graphics, Imaging, Vision*, April 2002, pp. 126–131.
- [30] A. Crimnsi, A. Zisserman, L. V. Gool, S. Bramble, and D. Compton, "A new approach to obtain height measurements from video," in *Proceedings of SPIE Investigation and Forensic Science Technologies*, vol. 3576, November 1999, pp. 227–238.

- [31] C. Madden and M. Piccardi, "Height measurement as a session-based biometric for people matching across disjoint camera views," in *Image and Vision Computing New Zealand*, Dunedin, New Zealand, November 2005.
- [32] A. Godil and S. Ressler, "Retrieval and clustering from a 3d human database based on body and head shape," in *SAE Digital Human Modeling conference*, Lyon, France, August 2006.
- [33] K. M. Robinette, "CAESAR measures up," *Ergonomics in Design*, 2000.
- [34] S. Z. Li, *Markov random field modeling in image analysis*. Springer, 2001.
- [35] J. L. Marroquin, E. A. Santana, and S. Botello, "Hidden markov measure field models for image segmentation," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1380–1387, November 2003.
- [36] A. Sundaresan, A. RoyChowdhury, and R. Chellappa, "A hidden markov model based framework for recognition of humans from gait sequences," in *Proceedings IEEE International Conference on Image Processing*, 2003, pp. 143–150.
- [37] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, vol. 36, pp. 192–236, 1974.
- [38] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 6, pp. 564 – 584, 1984.
- [39] R. Kinderman and J. Snell, "Markov random fields and their applications," AMS Press, 1980.
- [40] R. Dubes and A. K. Jain, "Random field models in image analysis," *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131–164, 1989.
- [41] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computational machine," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.
- [42] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society: Series B*, vol. 48, no. 3, pp. 259–302, 1974.
- [43] M. Ristivojevic and J. Konrad, "Space-time image squence analysis: Object tunnels and occlusion volumes," *IEEE Transactions On Image Processing*, vol. 15, 2006.
- [44] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions On Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [45] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math*, vol. 42, pp. 577–685, 1989.
- [46] S. Osher and A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulation," *Journal of Comput. Physics*, vol. 79, pp. 12–49, 1988.

- [47] R. Halir and J. Flusser, “Numerically stable direct least squares fitting of ellipses,” in *International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*.
- [48] A. W. Fitzgibbon and R. B. Fischer, “A buyer’s guide to conic fitting,” in *Proceedings of the 6th British conference on Machine vision*, vol. 2, 1995, pp. 531–522.
- [49] Cyberware web site, “<http://www.cyberware.com>.”
- [50] P. J. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, February 1992.