

Estimating Human Limb Motion from Video Sequences with Anatomical Knowledge

by

Ali Ghazisaeidi, M.Sc.

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of
Master of Applied Science in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering (OCIECE)

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada, K1S 5B6

January 2011

© Copyright 2011, Ali Ghazisaeidi



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-81677-6
Our file *Notre référence*
ISBN: 978-0-494-81677-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

The undersigned recommend to
the Faculty of Graduate and Postdoctoral Affairs
acceptance of the thesis

Estimating Human Limb Motion from Video Sequences with Anatomical Knowledge

submitted by

Ali Ghazisaeidi, MSc.

in partial fulfillment of the requirements for

the degree of Master of Applied Science in Electrical and Computer Engineering

Chair, Howard Schwartz, Department of Systems and Computer Engineering

Thesis Supervisor Professor Pierre Payeur

Thesis Supervisor Professor Richard Dansereau

Carleton University
January 2011

Abstract

The rapid progress in computer vision has broadened its applications to medical science, especially computer-aided diagnosis as an advancing subfield. Musculoskeletal examination is one the most important exams in medical diagnosis. In this thesis, we introduce a novel method for automating the musculoskeletal examination of the human upper limb. We propose a markerless approach for the analysis of the upper limb structure. We calculate 3D positions of the joints and bones in the upper limb and estimate the range of motion of the elbow. Furthermore, with the proposed system we are able to estimate the length of the forearm and upper arm bones and reconstruct a 3D model of the upper limb structure. Finally, the 3D model of upper limb structure is reconstructed. Compared to other markerless approaches, our experimental results show favorable accuracy and reliability of the proposed system. We believe that the future medical applications will benefit from the time- and cost-efficiency of image processing techniques to provide much flexibility for customers. In this research direction, computer vision, and especially markerless approaches play a key role.

This thesis is dedicated to my family and friends for their support and encouragement.

Acknowledgments

I would like to thank my thesis supervisor Dr. Pierre Payeur, for his guidance through my research.

Table of Contents

Abstract	iii
Acknowledgments	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Motivations	2
1.2 Objectives	3
1.3 Thesis contributions and outline	3
2 Literature Review	5
2.1 Motion capture methods	6
2.1.1 Manual motion capture	6
2.1.2 Marker-based motion capture	6
2.1.3 Markerless motion capture	10
2.2 Human motion analysis	11
2.2.1 Human body models	11
2.2.2 Image descriptors	12
2.2.3 Conclusion	15
3 Upper Limb Anatomy	16
3.1 Elbow	16
3.2 Wrist	19

3.3	Forearm and upper arm	20
3.4	Shoulder	22
4	Methodology	24
4.1	Marker-based vs. markerless motion estimation for clinical applications	24
4.1.1	2D image analysis vs. volumetric model analysis	25
4.2	Arm and joint models	26
4.2.1	Arm model	26
4.2.2	Joint movement model	26
4.2.3	Joint shape model	28
4.3	Proposed approach based on 2D image analysis	30
4.4	System description	30
4.5	Video analysis algorithm	32
4.6	Pre-processing	34
4.6.1	Background removal	34
4.6.2	Skin detection filter	34
4.6.3	Noise effects	36
4.6.4	Medial representation	37
4.6.5	Contour detection	38
4.7	Two dimensional joints location estimation	39
4.8	Three dimensional processing	47
4.9	Locating bone ends	55
5	Experimental Work and Results	58
5.1	System description	58
5.2	Degrees of freedom	61
5.2.1	Shoulder yaw	61
5.2.2	Shoulder pitch	63
5.2.3	Shoulder roll	65
5.2.4	Elbow pitch	67
5.2.5	Wrist roll	70
5.3	Bone lengths	73
5.3.1	Humerus	73
5.3.2	Radius	75
5.3.3	Ulna	76

5.3.4	Average error	76
5.3.5	Discussion	78
6	Conclusions and Contributions	83
6.1	Summary	83
6.2	Limitations	84
6.3	Contributions	84
6.4	Future work	85
	List of References	87
	Appendix A Application specifications	92

List of Tables

3.1	The approximate ranges of motion for the adult wrist.	20
3.2	The approximate ranges of motion for the adult shoulder	23
4.1	Skeletonization methods comparison.	37
4.2	Time consumed for different skeletonization methods.[seconds]	38
5.1	The measured lengths of the major upper limb bones by a doctor [inches].	59
5.2	The studied degrees of freedom in each articulations.	61
5.3	Shoulder yaw estimated and real values [degrees].	63
5.4	Shoulder pitch estimated and real values [degrees].	65
5.5	Shoulder roll estimated and real values [degrees].	69
5.6	The elbow angle [degrees] where the shoulder pitch is equal to 0 degree.	69
5.7	The elbow angle [degrees] where the shoulder pitch is equal to 30 degrees.	70
5.8	The elbow angle [degrees] where the shoulder pitch is equal to 45 degrees.	70
5.9	The actual and estimated results for wrist angles [degrees].	72
5.10	The calculated bones lengths [inches].	78
A.1	Main parts of the application.	92

List of Figures

3.1	Upper limb anatomy.	17
3.2	(a) The bony components of the elbow joint (frontal view) , and (b) the bony components of the elbow joint (lateral view)	18
3.3	Wrist anatomy	19
3.4	(a) Forearm bones , and (b) upper arm bone.	21
3.5	Shoulder anatomy	22
4.1	Upper limb model.	27
4.2	The arm models for (a) elbow joint, and (b) shoulder joint.	27
4.3	The proposed upper limb model.	28
4.4	MRI of the wrist.	29
4.5	MRI of the elbow.	29
4.6	MRI of shoulder.	30
4.7	Acquisition system with 4 cameras and 2 synchronized computers.	32
4.8	System flowchart.	33
4.9	Effect of inappropriate lighting: (a) input frame , (b) background (c) output frame.	35
4.10	Background removal and skin detection: (a) input frame , (b) background removal and skin detection filter applied, and (c) output silhouette.	36
4.11	Results for the skeletonization of Figure 4.10:(a) distance transform skeleton, (b) thinning technique skeleton, and (c) Voronoi skeleton.	37
4.12	Results of the edge detection for methods: (a) Prewitt, (b) Canny and (c) Sobel.	39
4.13	Locating the elbow joint.	41
4.14	Upper limb - lateral view.	42
4.15	Convex hull estimated on the forearm and hand.	42
4.16	Locating the wrist joint.	43

4.17	Locating the shoulder joint.	44
4.18	Locating the areas of interest, step 1: locating on the medial line . . .	46
4.19	Locating the areas of interest, step 2: locating on the contour.	47
4.20	Derivative algorithm description.	48
4.21	Three dimensional process.	49
4.22	Ellipse fitting.	49
4.23	Original ellipse located on the wrist and the perspective ellipses seen from the camera views.	50
4.24	Elliptical cones intersection.	51
4.25	Joints cross section planes.	53
4.26	Joints cross section vectors.	53
4.27	Elliptical cones and their intersection.	54
4.28	The cross section of the elbow joint and the corresponding geometry model.	56
4.29	The cross section of the wrist joint and the corresponding geometry model.	56
4.30	The cross section of the shoulder joint and the corresponding geometry model.	57
5.1	Acquisition system.	59
5.2	Some movement sample frames: (a) frame 1, (b) frame 30, (c) frame 50, (d) frame 80, (e) frame 130, (f) frame 155, (g) frame 200, (h) frame 225	60
5.3	Shoulder yaw.	61
5.4	Estimated shoulder yaw angles [degrees].	62
5.5	Shoulder pitch.	63
5.6	Estimated shoulder pitch angles [degrees].	64
5.7	Shoulder roll.	65
5.8	Shoulder roll.	67
5.9	Estimated shoulder rolling angles [degrees].	68
5.10	Elbow pitch.	68
5.11	Estimated elbow pitch angles [degrees].	71
5.12	Wrist roll.	71
5.13	Wrist and elbow estimated ellipses.	72
5.14	Estimated wrist roll angles [degrees].	73

5.15	Estimated length of humerus bone [inches].	74
5.16	Absolute error for the estimated humerus bone lengths.	74
5.17	Estimated length of radius bone.	75
5.18	Absolute error for the estimated radius bone lengths.	76
5.19	Estimated length of ulna bone.	77
5.20	Estimation error for the ulna bone length.	77
5.21	Arm positions for which the system does not respond properly: (a) forearm occludes the elbow and upper arm , (b) the elbow and the wrist are not distinguishable from the lateral view.	80
5.22	Camera views (case I).	81
5.23	Upper limb 3D illustration of arm bones for case I, seen from the right side.	81
5.24	Camera views (case II).	82
5.25	Upper limb 3D illustration of arm bones for case II, seen from above.	82
A.1	Application main window.	93
A.2	Application sample code.	94

Chapter 1

Introduction

The cave paintings show the great fascination in capturing human motion even in the early ages of humankind. Later in ancient Greece, Aristotle wrote a book on the movement of animals called “De Motu Animalium”. Two thousand years later, Leonardo da Vinci described the mechanics of jumping, standing, walking up and down hill, and rising from a sitting position in his anatomic drawings. Four hundred years later, in the 19th century, two photographers named Marey and Muybridge dissected human and animal motion and created the famous, “The Horse in Motion”, photo series. Since then human motion analysis and synthesis has received increasing attention. Particularly during the last decade, with the advances in computer vision and graphics and decreases in the price of digital cameras, human motion capturing and analysis using computer vision have become more popular.

A major application of human motion capture is in computer character animation and virtual reality (VR). The position information is gathered from the sensors that are attached to the whole body of the actors and sent to the computers. This information is used for mapping the movements of real actors to the animated characters. Another application of motion analysis systems is in sports. The sports performance professionals use motion capture systems to perform accurate functional evaluations for clinical and research oriented purposes such as injury prevention and performance enhancement. Surveillance systems, man-machine interfaces, content-based image storage and retrieval, and video conferencing are other examples of motion capture and analysis systems.

In general, for designing an automated video capturing system, two approaches are considered. In the first approach, markers are attached to the patient’s joints and multiple cameras record their position and motion. The disadvantages of this

approach include the need for trained personnel to attach and adjust the markers on the body. As well for medical applications, attaching the markers on the body of a patient inherently limits the range of motion and might corrupt precious information. In the second approach, namely markerless, no markers are attached to the subject. In this scheme, the cameras send the video streams to the computer, where the videos are processed, and the motion is recognized and analyzed.

1.1 Motivations

This thesis work started up some years ago in an earlier work done within our group (Viva Group in uOttawa). Using the multi camera system previously developed by another group member [1], we now developed a fully automatic system for musculoskeletal examinations of human upper limbs. In the proposed system, there is no need for the user to interact with the system (e.g. no need to locate any joint in the video frames).

This work includes the use of model-based estimation techniques for detecting the joints in the upper limb. Examination of the musculoskeletal system is one of the most complex aspects of the general physical exam. Assessing active range of motion is an important part of the musculoskeletal examination procedure. In this exam, a practitioner instructs a patient to flex and extend the joints, such as wrist, elbow, and shoulder. The practitioner examines and records the passive and active range of motion of the joints. The competency of the practitioner is a subjective matter and varies with the experience of the person.

Automated measuring of the range of motion at the joints can be considered as an assistive solution, meant to increase accuracy and reduce fluctuations in measurements due to human factors. The beneficial properties of using an automated measuring includes standard and uniform measuring among all clinical institutes and the automated recording system, which simplifies accessing the data for further analysis. However, the traditional motion capture and analysis systems in clinical practice are very expensive, mostly because of the high cost electromagnetic devices used in these systems or long setting times associated with marker-based techniques. This gives a strong motivation for using markerless computer vision based systems for human motion analysis.

Using inexpensive consumer-level cameras in the clinical applications decreases

the medical examination expenses. In this research project, we develop a markerless image analysis system for examination of the musculoskeletal system.

1.2 Objectives

A typical musculoskeletal examination consists of several examinations, such as joint-muscle examination, percussion, and inspection. The inspection consists of the measurement of range of motion of joints (active and passive). To examine range of motion of the joints, the practitioner asks the patient to flex and extend his/her elbow or shrug his/her shoulders. The range of motion is measured using a goniometer. The objective of this thesis is designing and implementing a system for measuring the range of motion of the upper limb articulations using a multi camera video capturing system. One of the applications of such a system is in the arthritis problem. The inflammation of the joints affects the range of motion on the joints. Exercise for arthritis is necessary to reduce pain and improve muscle and bone strength. The exercise program should be tailored to each individual. Our proposed system can be used in the arthritis diagnosis by recording the changes in the range of motion of every joint. This information helps a physical therapist to give the patient a suitable exercise program.

The system is supposed be able to estimate the upper limb bone lengths. It should calculate the range of motion of the articulations and show a virtual 3D model of the bones at each timestamp. In addition, the system should perform automatically and individually without the need for human manipulation or supervision at any stage.

1.3 Thesis contributions and outline

To our best knowledge this is the first time that a markerless system is used for accurate upper limb analysis. The system has the advantage over the marker-based systems since only passive vision is used for monitoring. The hassle of putting on the suits with the markers on them is therefore avoided. Almost all of the markerless systems have been developed for applications such as gesture detection or motion tracking of the whole body. These applications do not need very precise analysis of the images. But in a medical application, the goal is to perform detailed structure analysis which means that the accuracy of results should be comparable with that of

marker-based techniques.

Our proposed system is able to find and track the upper limb movements including rolling the wrist with no need for detecting the fingers. In other words this method can be combined with any method that works on the hand, without affecting its results. This system is able to find the position of the bone ends and the joint angles without any user supervision or interaction. Overall five degrees of freedom (DoF) are studied. These joints contain three DoF in the shoulder, one DoF in the elbow along with one DoF at the wrist. The outline of this thesis is as follows: Chapter 2 is the literature review and contains two parts. In the first section, the state-of-the-art in motion capture is examined. The second section presents the techniques that have been applied for human motion analysis. Weaknesses and strengths of today's approaches are discussed and our own views upon the ways forward are given. In Chapter 3 the upper limb anatomy is summarized, in support to the proposed technique. Chapter 4 presents the markerless method proposed in this thesis to locate the upper limb bones. Chapter 5 describes the experimental work and obtained results. The final chapter contains the thesis conclusion, contributions, and possible future work.

Chapter 2

Literature Review

More than a century ago, Marey and Muybridge started their pioneering work on animal locomotion by shooting multiple photographs of a galloping horse over a short period of time. Since then human and animal motion analysis and synthesis have received increasing attention. Especially during the last decade, researchers in the field of the computer vision and graphics have been developing new mathematical models for human motion analysis. Segmenting the parts of the human body in an image, tracking the movement of joints, and generating the 3D model of body structure are useful for precise analysis of medical videos. Motion tracking is of interest for police, military, and security organizations. In the military, it often refers to the tracking of an object such as a missile or a plane, while in film animation, it is used for recording the motion of the actors and mapping the real actor to a three-dimensional animated character. It is comparable to the old rotoscoping technique where a movie was captured and then the hand drawn animated character was produced according to each frame. Today the computer video games industry and animation movie industry are generating 3D human avatars and human-computer interaction is very common. Motion capture and analysis find applications in sports and clinical practice. In this thesis, our focus is on medical applications of the motion capture and analysis, as it is the native field of motion analysis. This chapter presents a review of the advanced methods used in the field of motion capture and analysis. The first section is an overview of motion capture technologies. The second section provides reviews related to the human motion analysis methods.

2.1 Motion capture methods

In this section, we study motion capture methods. Based upon where the capture sources and sensors are placed, Perales [2] has classified motion capture systems into the three following groups: outside-in, inside-out, and inside-in. An outside-in system is a system that uses external sensors such as cameras to collect data from sources (e.g. markers) placed on the body. In inside-out systems, such as electromagnetic systems, the sensors placed on the body collect external source signals. Finally, inside-in systems use both sources and sensors placed on the body. For instance, in electromechanical suits, the sensors are potentiometers and the sources are actual joints inside the body.

In general, the vision related motion capture systems are divided into the three following categories: manual, marker-based, and markerless. During the last decade, the marker-based motion capture has been very common in the machine vision industry and substituted the traditional manual capturing methods. Until now, markerless systems were mostly used for gesture detection systems or motion tracking. In the following section, the vision related motion capture systems are described in detail.

2.1.1 Manual motion capture

In manual motion capture, a video of the object or patient is recorded and later a human operator (e.g. practitioner) analyses the video to find out how the movement has been performed and find movement problems. The disadvantage of manual motion capture is that a human observer must monitor and control all the processes and the results manually. Therefore, the results are subjective and can be different with various interpreters. Hence this technique is not reliable and tends to be inaccurate. In addition, the video is two dimensional data and lacks the depth information which makes the analysis of a three dimensional data more comprehensive.

2.1.2 Marker-based motion capture

Marker-based motion capture is currently the most popular method in human motion capture, especially for clinical applications. In a typical marker-based system, markers are attached to the body of the patient. Vision sensors record the positions of the markers over time and the (two-dimensional) coordinates of the markers' locations in several images are reported to the central system. By combining the coordinates of the

same marker from different sensors, three-dimensional information of the movement can be estimated. Therefore, the operator needs to identify the optimum points for placing the markers. In other words, anatomically trained personnel is needed to place the markers on the patient's body. The positions of these points depend on the part of the body the practitioner is working on. For anatomical purposes, it is important to place the markers at a joint position to be able to find the 3D position of the joint and the joint angle. A major problem with this method is that it is a time consuming task to place the markers in the anatomically correct locations. For achieving the highest accuracy in medical applications, Lucchetti *et al.* [3] use bone pins screwed into the bones as the gold standard for evaluation. In other medical applications, magnetic tracking sensors for capturing the position of the markers have been used [4,5]. Other than magnetic trackers, reflective markers or small LEDs are very common because of their light weight. For example, the Vicon Motion Capture [6] uses high-resolution cameras with LED flashes along with reflective markers. In their system, several LEDs are installed around the lens of each camera for a better illumination and the reflective markers attached to the specific body joints reflect the light emitted from the LEDs.

A multi-camera system is commonly used for capturing the videos. The cameras are mounted on the walls and ceiling, pointed at a common space from different angles. The actor or the patient moves within that space and those cameras capture the visible marker locations. This information is gathered by one computer or by several computer nodes attached to a network. Finally, a central computer performs the task of matching markers location between the views. Given that at least two cameras see a specific marker, the three-dimensional coordinates can be estimated. System calibration is a common and critical step in both markerless and marker-based systems, and the accuracy of the system depends on it. After this step, the subject with all the markers attached to him/her, is placed in the space surrounded by calibrated cameras. The placement of the markers should be done accurately, so that the light ray can reach as many markers as possible from all the angles during the subject movement. If a marker is located in a occluded area or not captured by any of the cameras, the data from that marker is lost and it affects the output of the system.

- **Estimation Methods:** In some marker-based capturing system applications, where the human body is under analysis, a general model of human skeleton is

used. In order to map this model to the body of the patient, markers are preferably positioned at the joints. The joints are generally the best places because the soft tissue at the joints is very thin and the contractions and relaxations of the muscles have minor effects on the position of the joints. In some approaches, many markers are spread on the muscles in order to model the relaxation and contraction of muscle. In [7] Holmberg and Lanshammar added an artificial texture on the lower limbs in order to find the location of knee joint. An accurate model contains many parameters that should be estimated. Therefore a large number of markers leads to more accurate model estimation. The estimation methods are divided into the two following main categories:

1. Prediction Methods: Prediction approaches have been applied for estimation of the variables in the model. In [8] and [9], the authors use regression equations based on standardized body geometry. They obtained regression coefficients from the hip joint of a number of adults. Schwartz and Rozumalski [10] proposed a predictive method for joint parameter estimation. They started by motion tracking of two adjacent segments sharing a single joint. Then they found the axis of rotation for each joint. Using an iterative and non-parametric method, they estimated the most likely effective joint center of all joints and the most likely direction of the axis of rotation.
2. Functional Methods: The techniques based upon the geometrical relation of joint centers and nearby anatomical landmarks are known as functional methods. In functional methods, the goal is to locate the center of rotation of the joint by numerical methods rather than relying on the prediction equations or predefined models. These methods are mostly useful in studying a ball and socket or hinge joint, like the knee or pelvis articulations. The idea behind them is estimating the center of the joint assuming the joint resembles a sphere. Therefore, by attaching the markers all around it, these markers will form a sphere and the coordinates of the center of the sphere can be calculated by numerical methods. In a research on locating the hip joint center, Halvorsen *et al.* [11] presumed the pelvis as stationary. Therefore, they attached the markers on the thigh so that they lie on a sphere with the hip joint as the center. In this situation the line linking two different positions of the marker at two different moments is perpendicular to the line linking the hip joint and their midpoint. In this way

they managed to find the hip joint with an acceptable accuracy. Marshall and Shaw [12] have applied the downhill simplex method by Nelder and Mead [13]. Their algorithm defines a cost function and tries to find the minimum of the function that corresponds to the position of the hip joint center. Leardini *et al.* [14] reported that the functional methods were more accurate than any prediction approach. However, Bell *et al.* [15] reported that the regression method significantly outperformed all the functional methods for hip joints with a small range of movement.

- **Errors:** Although marker-based systems are considered as very accurate systems and their use is very common in commercial applications such as virtual reality, there are three common sources of error in marker-based systems:
 1. **Marker Placement Errors:** In medical applications, the accurate placement of the markers is very important. The proper treatment of the patient can take several months; placing the markers in not exactly the same positions at each examination can undesirably affect the result of each treatment in comparison to the preceding ones.
 2. **Skin Movement:** As stated in [16], the error from the skin movement especially in elderly people can reach several centimeters. In order to avoid this type of error, modeling the skin movement has been proposed in [17].
 3. **Predictive Methods:** Since the prediction coefficients are defined according to a body model, which is based upon the average dimensions of the human body, errors are inevitable.

Currently, marker-based systems are dominant in practical motion analysis applications. Marker-based systems are considered to be adequate for applications in clinical environments. On the other hand, the cameras just track the markers and the information for analysis is only the location of the markers. Marker-less systems operate on whole images captured by the cameras, which require sophisticated algorithms for image processing.

There are also disadvantages to using marker-based systems. When all the errors mentioned above add up, the results can be inaccurate in some applications, such as locating the hip joint. The main limitation is that a large part of the information of the movement is thrown away and only the marker positions are

retained. However, the disposed information could be useful for achieving more accurate estimation of the model parameters. As an alternative for solving this problem, using a large number of markers spread over the skin could provide more data for a new model of soft tissue movement, which could help reduce soft tissue movement error.

2.1.3 Markerless motion capture

Markerless motion capture systems attempt to address some of the weak points identified with marker-based systems. Markerless systems rely on the images only and do not need any sensors and markers to be mounted on the human performer/patient. Therefore, using markerless systems eliminates the need for the patient to put on any suit with markers or textures, and this makes it possible for the patient to move freely and more naturally. In addition, complete images are sent to the central computer and this gives the possibility for various feature extractions.

The attempts in markerless capture systems can be divided into three groups:

In the first group, a 3D articulated body model is fitted to data by projecting this model to the several image views. Afterwards, these projections are compared and fitted on the images taken by the cameras. For example in [18], Delamarre and Faugeras compared detected silhouettes of the person to the projections of a 3D model of a human being. The difference between the projected images and the silhouettes were used to find the motion of the articulated 3D model.

In the second group, which has been implemented in more recent research, the 3D model is generated by combining the images from different views. Using the detected human silhouette from each view is a popular method for three-dimensional reconstruction of visual hull [19,20]. The method proposed in [21] uses data of typical human shape gathered from a large number of people. With this data set as a basis, an active shape model [22] scheme is adopted. The active shape model is then compared and fitted to the visual hull constructed from several video camera views. In another example of this type of system, the voxels of the object are constructed and classified as foreground and background [23]. Combining the foregrounds from multiple views is used to generate a visual hull of the object. The advantage of this work is that it does not need the performer to start from a specific initial posture. However, using only the visual hull as a basis for alignment of two three-dimensional objects is ambiguous. Therefore, some researchers try to extract more information from the texture and add

it to the voxel information. Cheung *et al.* proposed voxel coloring which is estimating the colors of the pixels and adding it to the voxels [19]. This approach relies heavily on the assumption of Lambertian surfaces, which states that the color of the point is independent from the viewing angle.

The third group of markerless techniques is model-based methods. In model-based techniques, the area of interest is not just the shape of the subject under study but also the parameterization of it. In the case of human motion analysis, the parameters are characteristics such as inertial properties of the segments or segment lengths. Variables of the model are joint positions and joint angles. Recently an interesting approach based on hands-on fit of a detailed model to image material from uncalibrated video sequences has been proposed [24]. In the paper, a three-dimensional model of a human being is represented. Then the model parameters such as size, position, and articulation are changed to make the image of the model fit to the real data images. This method gives good results when contraction and relaxation of muscles is considered. However, it performs poorly in the estimations of internal and external rotation. The latter methods are good representatives of the state-of-the-art in markerless human motion analysis.

2.2 Human motion analysis

After reviewing various motion capture technologies employed in industry and research, we now present the existing techniques used in the analysis of the human body motion. In [25], Aggarwal and Cai divided human motion analysis into three categories: human motion tracking, human activity recognition and, body structure analysis. Body structure analysis, which is the topic of this thesis, is divided into model-based and model-free approaches. In model-based approaches, modeling means the construction of the likelihood function, as well as defining the camera model, the image descriptors, and human body model.

2.2.1 Human body models

Human body models consist of the kinematic properties of the skeleton (static models), and the shape and appearance (dynamic models).

- **Kinematic Models:** Most of the models found in the literature describe the human body as a tree, consisting of segments that are linked by joints. Kinematic models of a human body are defined in 2D and 3D. In 2D pose estimation applications such as 2D human pose recovery from a single image, spatial constraints are applied for the human body model. Then, using the tree structured model, the most probable pose is estimated by maximum a posteriori estimation [26]. Both 2D and 3D models most often model segments as individual rigid parts. Unlike 2D models, 3D models can allow a maximum of three rotations per joint [27–29].
- **Shape Models:** The human shape is also modeled in 2D and 3D. Segments in 2D models are commonly modeled as rectangular or trapezoid-shaped. In 3D models, individual segments are modeled by volumetric or surface-based representations. Commonly used volumetric models are spheres [30] and cylinders [31]. Unlike kinematic models that model each segment as a separate rigid shape, surface-based models often use a single surface for the entire body. These models typically consist of a mesh of polygons that is deformed by changes to the underlying kinematic structure [28, 32]. In [33], a combination of kinematic model and skin surface model is used.

2.2.2 Image descriptors

Often filters are applied to images for extracting the useful information and further working on these outputs. In the human motion analysis, the data related to the clothing and lighting conditions should be removed from the image. Most used image descriptors include silhouettes, edges, 3D reconstructions, motion, and color. In the following, we describe each of these descriptors.

- **Silhouettes:** Silhouettes are insensitive to variations in surface such as color and texture, so using the silhouettes can help recover 3D poses [34]. However, performance is limited due to shadows and noisy background segmentation.
- **Boundary representation:** In the boundary representation methods, the object is represented by boundary related correspondents. In the most simplistic decomposition, the points located on the boundary of the object are used for the object representation. The mesh of tiles on the boundary of the object

can also be applied in some other approaches [35]. Moreover, some researchers proposed the decomposition of the boundary surface using orthogonal functions [36]. The boundary surface is represented by the coefficients of the basis orthogonal functions.

- **3D Reconstruction:** As 2D edges and silhouettes lack depth information, in a multi-camera system, using the images taken from different views, a 3D silhouette can be extracted. Two common techniques are volume intersection [37] and voxel-based approach [38]. In the voxel representation schemes, the objects are represented by a voxel displacement vector and a voxel label field. In the voxel label field each voxel is labelled as background/foreground or the object name. The strength of this method is its ability to rapidly access the entire interior and exterior object information. This representation method needs a high amount of storage and memory for both saving and working on the objects. Recently a fast 3D scanning system has been introduced by the XYZ RGB company. The scanning system captures geometrical data with the fidelity required for accurate industrial applications. It is claimed that a 360 degree full color body scanning takes 0.01 seconds using this system.
- **Medial Representation:** In the medial representation method, the object is represented by the set of the points equidistant to the object boundaries called skeleton. Typically, the skeletons have different mathematical definitions. Therefore different variants of skeleton can be found, such as straight skeletons, morphological skeletons, and Voronoi skeletons [39]. The original idea of the skeletonization was introduced by Blum [40]. He described an algorithm named media axis transform (MAT), which is based on the grass fire analogy. There are three major categories of the medial representation techniques depending on the definition of skeleton: (i) Distance transform technique, (ii) Morphological technique (thinning), and (iii) Voronoi diagram technique.

In the following section, we describe the three major categories of the medial representation techniques:

1. *Distance Skeleton:* Every point of a distance skeleton has at least two points on the boundary of the image with equal distances. Any distance measure function, such as Euclidean or Manhattan distance transform [41], can be used for measuring the distances. In this thesis, we applied the

Manhattan distance because of its simplicity in implementation. For the distance skeleton extraction, first the boundary of the image is extracted. The distance map is generated, where each element is the distance of the corresponding pixel to the nearest boundary pixel. The ridges are detected and marked as the skeletal points.

2. *Thinning Technique*: The thinning technique is an iterative method that erodes the surface of the object in each phase. The beneficial properties of thinning include connected topology as well as retaining topology and shape. Moreover it produces the skeleton in the middle of the object.
3. *Voronoi Diagram*: The Voronoi diagram of a collection of points (generating points) divides the space into a set of cells. Each cell contains one central generating point and its adjacent points which are far from the centers of other cells. If the density of the generating points approaches infinity, then the Voronoi diagram converges to the medial axis. Voronoi diagrams skeletonization technique can keep both topological and geometrical properties of the object, but it is an expensive process, especially for large and complex objects.

- **Other Properties**: Taking into account that texture and color of body parts do not change in different body poses, these two values can be used as parameters for human body modeling. In [42], Gaussian color distributions have been used for segmentation and modeling. Similarly in [43] skin color histograms have been utilized for modeling. However, in practice, texture and color are not static and different clothing and lighting can make the model unreliable. To overcome these problems, a 3D model was proposed in [44] and human body parts are modeled with cylinders and a multi-dimensional color histogram was utilized for representing the feature distributions and clustering.
- **Combination of Descriptors**: Combinations of descriptors have been used in several attempts. For instance in [45] silhouette information is combined with edges. Similarly, Bériault *et al.* used color and 3D silhouettes [1].

Furthermore, model-free approaches are used when no human body model is available. In these techniques, a direct relation between body pose and the captured image is presumed. Model-free approaches are categorized into learning-based and example-based methods. In learning-based methods such as the method used in [46], some

data are used for training the system and a function which relates the image space to pose space is learned. Example-based approaches utilize a collection of patterns, together with their corresponding pose descriptions. For a given input image, a search is performed among the patterns and candidate poses are interpolated [47].

2.2.3 Conclusion

This chapter presented a review of state-of-art human motion capture methods. Markerless full 3D human kinematics extraction (e.g. [1] [48], or [45]) generate a comprehensive body model by utilizing shape-from-silhouette techniques. Still they remain far from real-world applications because of their lack of robustness. Moreover, after the 3D shape generation, three dimensional shape analysis is needed for extracting the information, which is a time consuming task.

In medical applications, more robust, accurate and fast solutions are needed. Hence, we decided not to use a full 3D human model and perform all the analysis on the two dimensional images which is a less computational consuming task. Finally, by relying on the anatomical shape of the human body, we represent a 3D model for the human body anatomy which makes the analysis and measurements less complex. By analyzing all of the camera views simultaneously and using the anatomical knowledge, the desired body parts are located in each camera view and by combining the location information in all of the camera views, 3D representation of the body parts is obtained.

In Chapter 3, the upper limb anatomy is summarized. In Chapter 4, the proposed method for using the upper limb anatomy knowledge is introduced.

Chapter 3

Upper Limb Anatomy

This chapter describes the anatomy of the human upper limb from the shoulder to the wrist. The structure of the upper limb contains three major bones that are linked together by ligaments in articulations, which are shown in Figure 3.1. Bones have a dual major role as facilitators of movement as well as protectors of the soft tissues of the body. To produce movement, many muscles pull various parts of the arm while the bones play the role of levers to preserve a certain posture. Like any other part of the body, flat connective tissue sheets called aponeuroses connect these muscles to the bones. In addition, bones and muscles are connected to each other by tendons which have cable-like structures. The following four sections provide the anatomy of three articulations and the structure of the forearm and the upper arm.

3.1 Elbow

Elbow articulation is made up of the three long bones, the cartilage coverings, ligaments, tendons, and muscles. Three bones involved in elbow joint are the lower end of humerus, the upper end of radius, and the upper end of ulna. Figure 3.2 shows the elbow joint. The humerus is the only bone of the upper arm and forms the top of the elbow joint. Ulna is located on the forearm on the same side as the little finger, while radius is the shorter forearm bone and located on the side of the thumb. These three bones construct three separate but interrelated articulations, which are named according to their position. These articulations are humero-ulnar (between humerus and ulna), humero-radial (between humerus and radius) and superior radio-ulnar (between ulna and radius). All three joints are surrounded by a single enclosure called

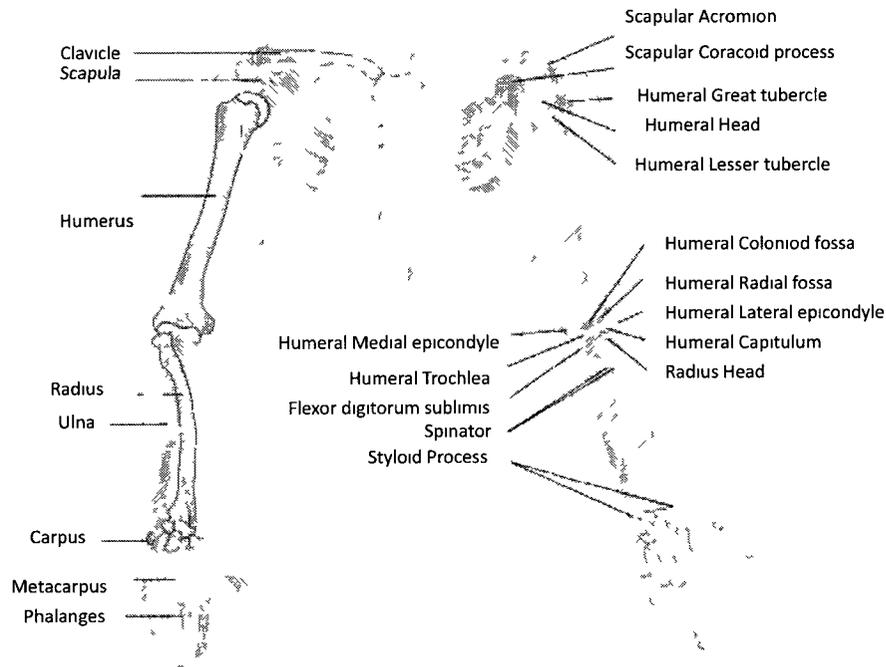
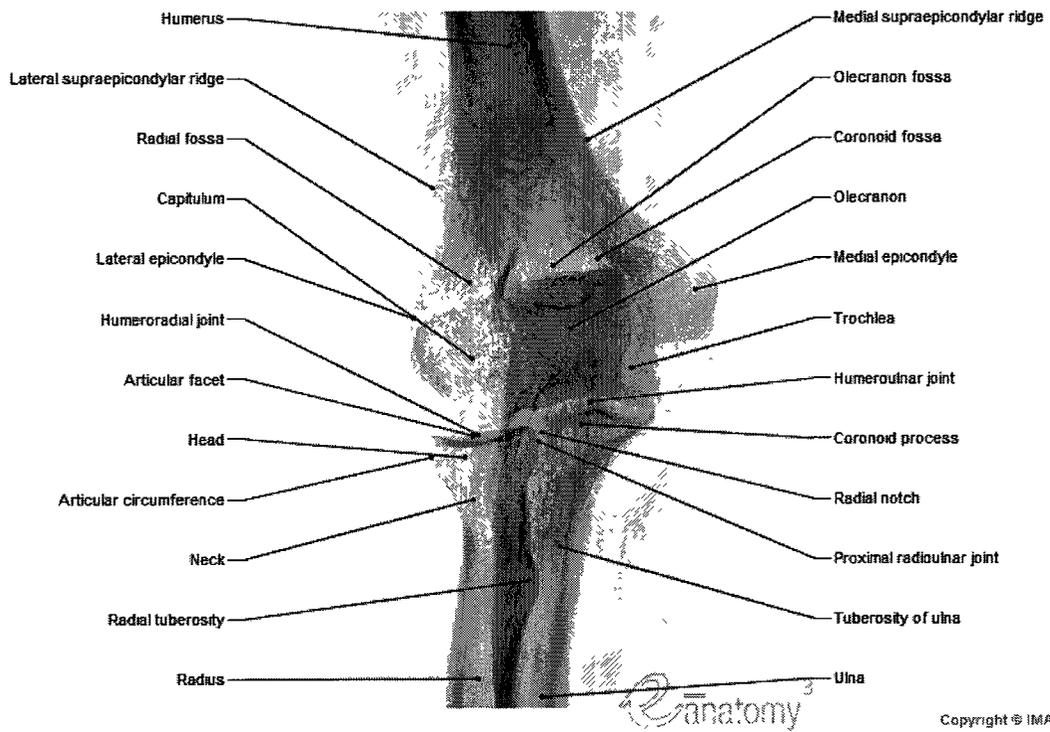


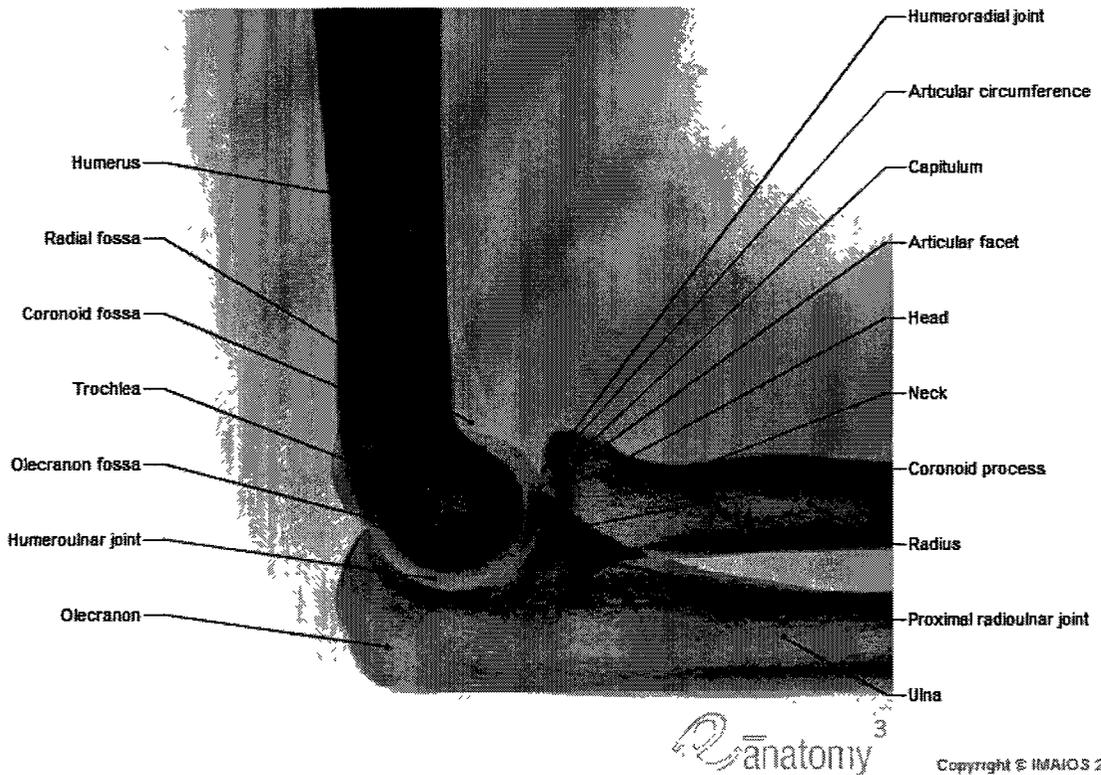
Figure 3.1: Upper limb anatomy. Adapted from [49]

the synovial joint capsule.

Articular cartilage is a substance found on the surface of all movable human bones, which causes the ease of the motion in the joints. At the elbow, the cartilage has two roles. First, it absorbs forces that are received in the joint. In addition, as the elbow joint moves, cartilage reduces friction between the bones. Flexion and extension are two movements that take place at the elbow, that is at the humero-ulnar and humero-radial joints. Flexion happens by touching the shoulder with the fingertips of the same side. Extension occurs when somebody straightens the upper limb. Normally, the range of motion of flexion and extension is between 0 to 135 degrees actively and 0 to 150 degrees passively [50]. Main muscles acting across the elbow include biceps brachii (responsible for supination and flexion), brachialis (causes flexion), triceps (causes extension), and pronator teres (causes pronation).



(a)

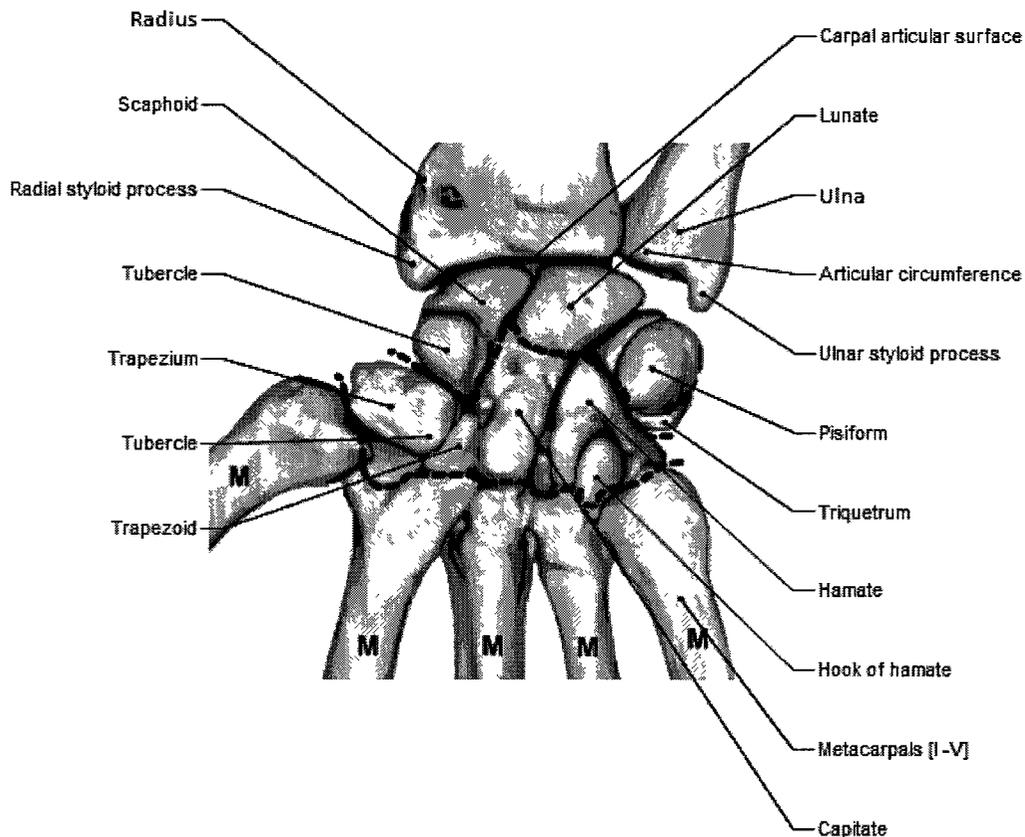


(b)

Figure 3.2: (a) The bony components of the elbow joint (frontal view) (reproduced with permission, [51]), and (b) the bony components of the elbow joint (lateral view) (reproduced with permission, [51]).

3.2 Wrist

This section is a review of wrist and hand biomechanics. The wrist joint contains different bones including the end of the ulna, the end of the radius, and the eight carpal bones. The main joints formed by these bones are the distal radio-ulnar joint, the radiocarpal joint, the ulnocarpal joint, the midcarpal joints, and the intercarpal joints. The distal radio-ulnar joint is the joint that is between the ends of the radius and ulna. It allows the wrist movements of pronation and supination to occur. Supination is rotation of the forearm so that the palm is turned up. Pronation is rotation of the forearm so that the palm is turned down. Figure 3.3 illustrates the wrist anatomy.



Illustrations . A. Micheau - MD

Copyright © IMAIOS 2010

Figure 3.3: Wrist anatomy (reproduced with permission, [51]).

Functionally the radio-carpal, ulno-carpal, midcarpal and intercarpal joints work together to allow the wrist to bend (flexion), straighten (extension), move inwards (ulnar deviation), and move outwards (radial deviation). The wrist is a condyloid joint,

that is to say, it allows flexion, extension, abduction, adduction and circumduction, the last being a combination of the previous four. Table 3.1 shows the approximate ranges of motion for the adult wrist.

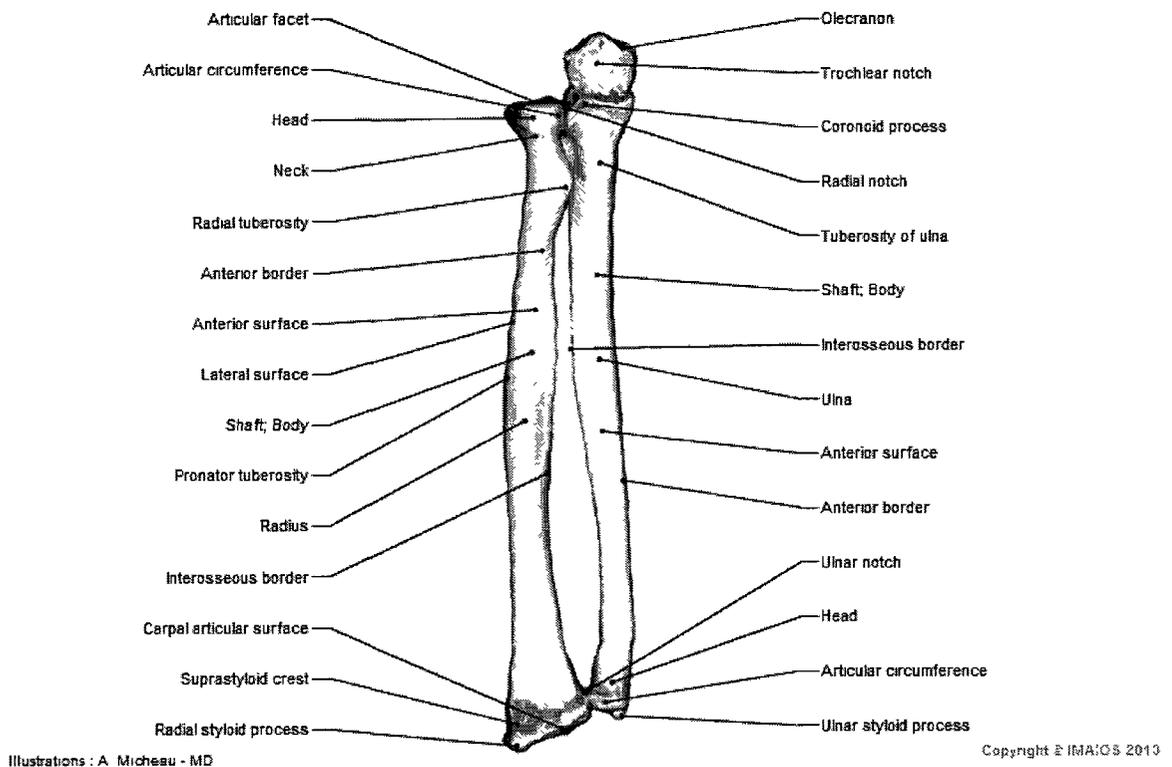
Table 3.1: The approximate ranges of motion for the adult wrist. [50].

Movement	Maximum range of motion (degrees)
Flexion	80
Extension	80
Ulnar deviation	30
Radial deviation	20

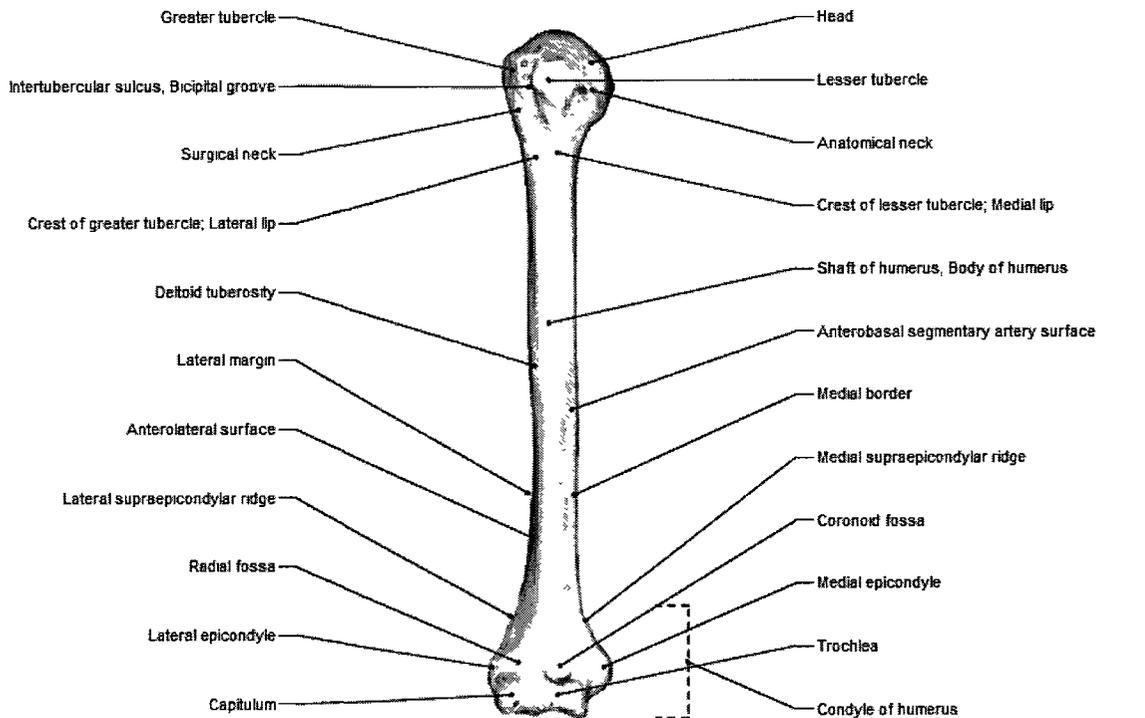
3.3 Forearm and upper arm

The forearm is made up of two bones: ulna and radius. The radius is composed of the head, neck, shaft, and distal end. The ulna consists of trochlear fossa, coronoid process (with its radial notch for articulation with the radial head), shaft and small distal head.

There are two types of forearm movements: pronation and supination. Pronation is the inward rotation of the forearm in a way that makes the head of radius swing around the ulnar towards the body. Supination is the rotation of the arm in opposite direction. Both movements occur through the forearm at the proximal and distal radio-ulnar joints. The maximum range of motion of pronation and supination is 80 degrees. The upper-arm contains only one bone named humerus which connects the scapula and the forearm. Figure 3.4(a) and Figure 3.4(b) show the anatomy of the forearm and upper arm, respectively.



(a)



(b)

Figure 3.4: (a) Forearm bones (reproduced with permission, [51]), and (b) upper arm bone (reproduced with permission, [51]).

3.4 Shoulder

The shoulder is the most mobile and the most flexible of the human joints. The omni-directional construction of the shoulder permits the arm to be rotated 360 degrees. The skeletal components of the shoulder include humerus, the long bone of the upper arm; the scapula or the shoulder blade, the clavicle, or the collarbone, and the sternum, or breastbone. Figure 3.5 shows the shoulder anatomy.

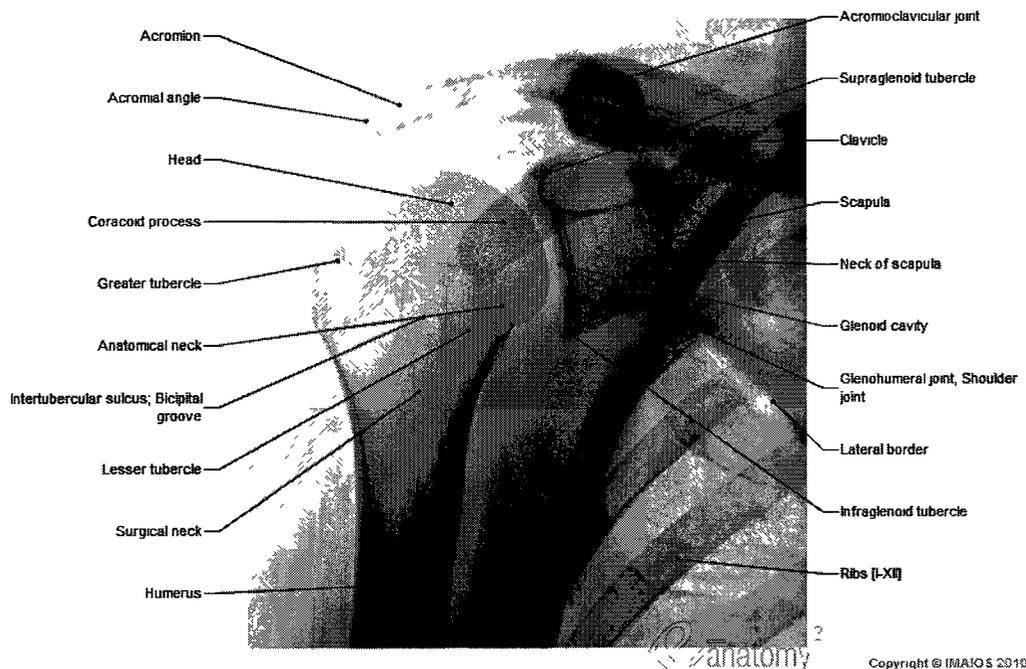


Figure 3.5: Shoulder anatomy (reproduced with permission, [51]).

These bones make three joints in the shoulder complex. The main joint is a ball and socket joint named glenohumeral joint. In this joint, the top of the ball-shaped end of humerus bone sits in the cavity on the end of the scapula. The other two joints in the shoulder complex are the sternoclavicular joint and the acromioclavicular joint. Acromioclavicular is the junction between the scapula at the highest point of the shoulder and the outer part of collarbone. The sternoclavicular joint connects the medial part of the collarbone to the breastbone. The movements of the glenohumeral joint include flexion, extension, inward or outward rotation, abduction and adduction. Flexion and extension occur by forward and backward lifting of the arm respectively. Abduction and adduction happen by movement of the arm away and towards the body. Movement at the glenohumeral articulation requires motion at the other joints

of the shoulder complex. This coordinated movement of the joints is referred to as the scapulo-humeral rhythm. For example, for every 15 degrees of shoulder abduction, 10 degrees occur at the glenohumeral joint and the rest occurs at the scapulothoracic joint.

Table 3.2 provides the approximate ranges of motion for the adult shoulder.

Table 3.2: The approximate ranges of motion for the adult shoulder [50].

Movement	Maximum range of motion (degrees)
Flexion	165 to 180
Extension	35 to 60
Abduction	180
External rotation	90
Internal rotation	70

The muscles of the shoulder complex provide stability and movement. During shoulder movements such as lifting, certain muscle groups help to move the shoulder, while other muscle groups help to stabilize the shoulder complex. Much of the stability in the shoulder complex is provided by this muscular coordination.

Chapter 4

Methodology

In this chapter, we explain the markerless method proposed in this thesis to locate the upper limb bones in the video frames taken by a multicamera system. The goal of this work is to find the bone ends and measuring their movement in 3D. In the following section, we explain our rationale in choosing this particular method. Then, we present the proposed method in detail.

4.1 Marker-based vs. markerless motion estimation for clinical applications

In Chapter 2, different human motion capture and analysis techniques were reviewed. Marker-based motion capture systems provide reliable results for various applications. However, in clinical applications, markers can limit the ability of the patient to move correctly and naturally. Even with patients who suffer from joint pains such as arthritis (inflammation of joints), bursitis, or muscle pain, attaching the markers to the joints may be very bothersome or even impossible. In addition, marker-based systems need trained personnel for installing the markers on the patient's body, which is also very time-consuming. Therefore, markerless systems are more suitable for clinical applications where there is a lack of operators trained to use image processing techniques.

Markerless techniques have been mostly used for human motion tracking or gesture recognition, which do not need high accuracy. Our goal in this thesis is to design a markerless technique for human body structure analysis in clinical applications, therefore we need to develop a higher accuracy system that is able to locate the exact

position of the articulations and bones.

4.1.1 2D image analysis vs. volumetric model analysis

Researchers in both markerless and marker-based motion capture systems usually focus on building volumetric representations of the object. If the 3D reconstructed volume has enough resolution, locating its various parts and performing motion analysis can be implemented by utilizing 3D algorithms which give more robust results in comparison with working on individual 2D images. However, there are disadvantages in this type of approach. The first drawback is that several cameras are needed to cover the object from every angle, such that the triangulation algorithm required for finding the 3D voxels occupied by the subject can produce an acceptable result. Obviously, this involves an expensive and advanced hardware with well calibrated cameras which should be spread all around the object. In addition, the 3D model construction is a time and memory consuming process and cannot be implemented as an online solution. Moreover, in many cases the resolution of the final silhouette is limited. Especially the resolution of videos captured by cameras is generally lower than that of still images taken by the same cameras. Therefore, a 3D approach appears less suitable for analyzing the structure and movement of an object, which in our case is the human upper limb.

One of the two strategies consists of converting 2D images to a 3D volume and working solely in 3D (volumetric reconstruction). With this approach, structural analysis and location of the desired points such as the wrist or the elbow require deploying 3D algorithms. Working on volumes brings the advantage of a 3D comprehensive vision of the arm, provided that the cameras are positioned properly. However, as already stated, volume representation and working in 3D are both time consuming. A second alternative consists of working in 2D. By working separately on each 2D view and studying the images we can find the joints in each image by applying some anatomical knowledge. The points do not necessarily correspond to the actual joints, but in refining the analysis, the real positions of the joints and bones in 3D can be identified. The advantage of the second approach is that it requires fewer cameras while achieving equivalent results as with a 3D approach, for the medical application considered in this work.

In our initial experimental results, we observed that after several attempts for 3D reconstruction of the arm, the resolution of the output was not very high, unless a

great amount of time was spent to deal with complex 3D models, which also requires massive computational infrastructure. Therefore, we opted for the second approach and concentrated our work on the analysis of a set of synchronized 2D images and then reconstructed 3D information from the processed 2D images.

4.2 Arm and joint models

In this section, the models proposed for the arm and the joints are described.

4.2.1 Arm model

Most of the human body models rely on the bone structure of the body, i.e. bones and articulations. This is motivated by the solid shape of the skeleton, which requires a lower number of parameters for modeling in comparison with muscles that have different forms during relaxation and contraction. Some researchers tried to model the muscle deformations for specific purposes such as computer graphics. In this thesis, the first objective is to detect the human arm skeleton, in order to measure its movement with a passive markerless vision system. Therefore muscle modeling, which is a computationally intensive task, is not needed. Instead, we model the forearm and the upper arm as two elliptical pipes, and try to find the parameters of the cross section ellipses at the joints. Figure 4.1 shows the assumed model.

As mentioned in Chapter 3, the main articulations of the upper arm are the wrist, the elbow, and the shoulder. One end of the upper limb is formed at the shoulder where the upper limb and torso are connected. The elbow is located in the middle of the upper limb. The wrist is the final joint that is located between the hand and the forearm. Note that analysing the movement of the hand and fingers is not the objective of this thesis.

4.2.2 Joint movement model

The joints can be modeled by their major movements. In this regard, Figure 4.2(a) shows the elbow joint and its corresponding engineering model. In Figure 4.2(b), the shoulder joint is modeled. Note that the same model as shoulder can be used for the wrist articulation but with limited range of motion.

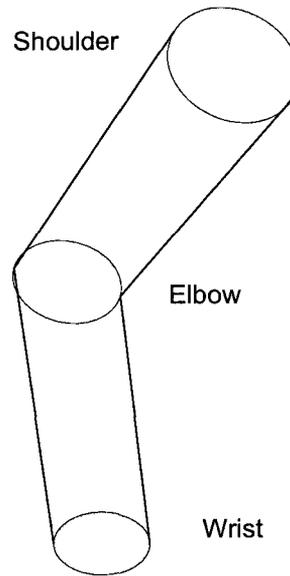


Figure 4.1: Upper limb model.

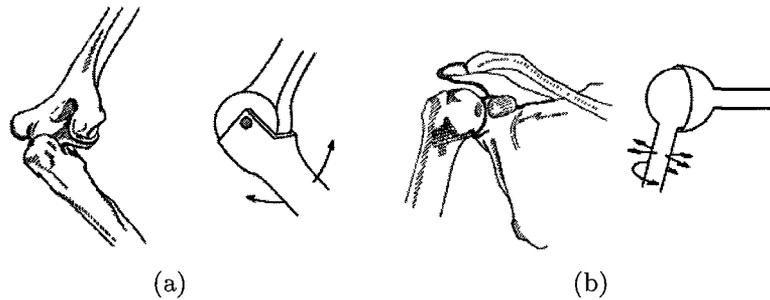


Figure 4.2: The arm models for (a) elbow joint, and (b) shoulder joint [52].

For modeling the arm, we used the structure shown in Figure 4.3. The number of degrees of freedom is the number of different ways that each joint can move.

Rotating the arm like a screwdriver when the arm is held straight is called rolling. Pitching is the movement of the joint in the up and down directions, while the movements to left and right are called yawing. In our model, the shoulder has three degrees of freedom: pitch, yaw, and roll. For the elbow, pitching means bending the elbow. We just consider rolling for the wrist joint although rolling in fact happens at the forearm.

We modeled the forearm and the upper arm as two pipes. The shoulder is a socket ball joint and has three degrees of freedom. The elbow acts as a door joint with one degree of freedom and the wrist is a socket ball joint with three degrees of freedom.

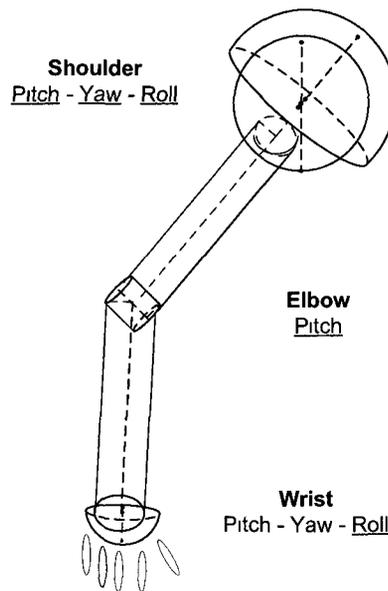


Figure 4.3: The proposed upper limb model.

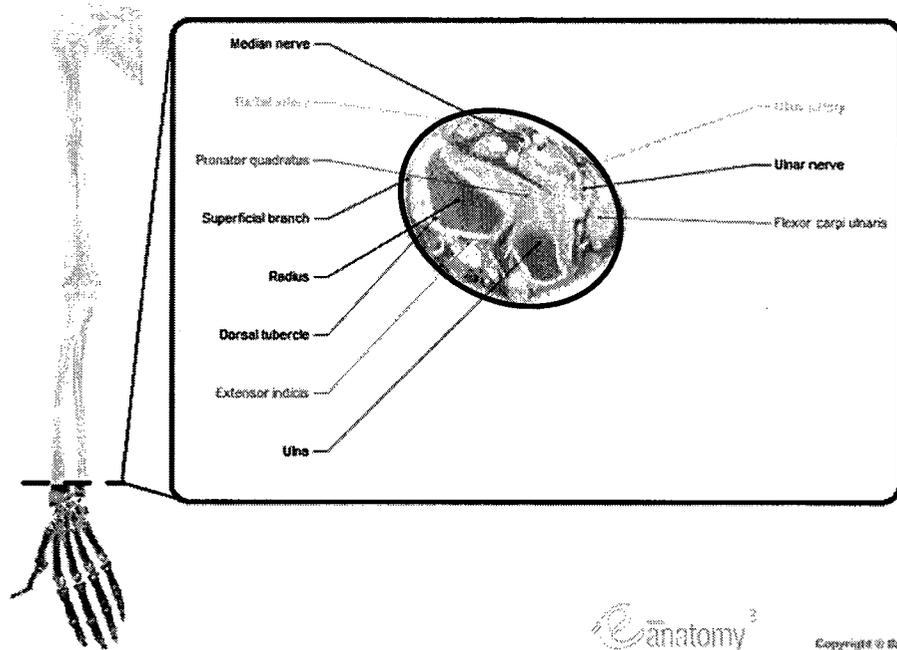
Since the hand is not included in our work, we consider one degree of freedom only for the wrist. Thus we modeled the arm with five degrees of freedom.

4.2.3 Joint shape model

Starting with one wrist cross section, as it is shown in Figure 4.4 for a Magnetic Resonance Image (MRI) of the wrist, the cross section of the wrist can be approximated by an ellipse.

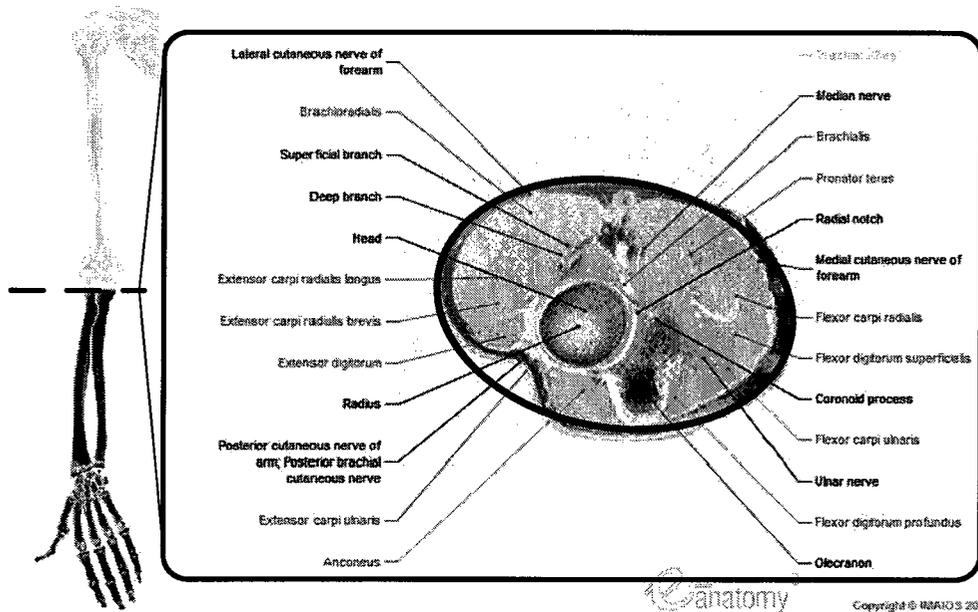
The elbow cross section is shown in Figure 4.5 for an MRI image of the elbow. The cross section of the elbow can be estimated as an ellipse.

An MRI of the shoulder is shown in Figure 4.6. Similar to the other joints the shoulder can be considered as a deformed circle or an ellipse.



Anatomy³ Copyright © IMAIOS 2019

Figure 4.4: MRI of the wrist. (Reproduced with permission, [51])



Anatomy³ Copyright © IMAIOS 2019

Figure 4.5: MRI of the elbow. (Reproduced with permission, [51])

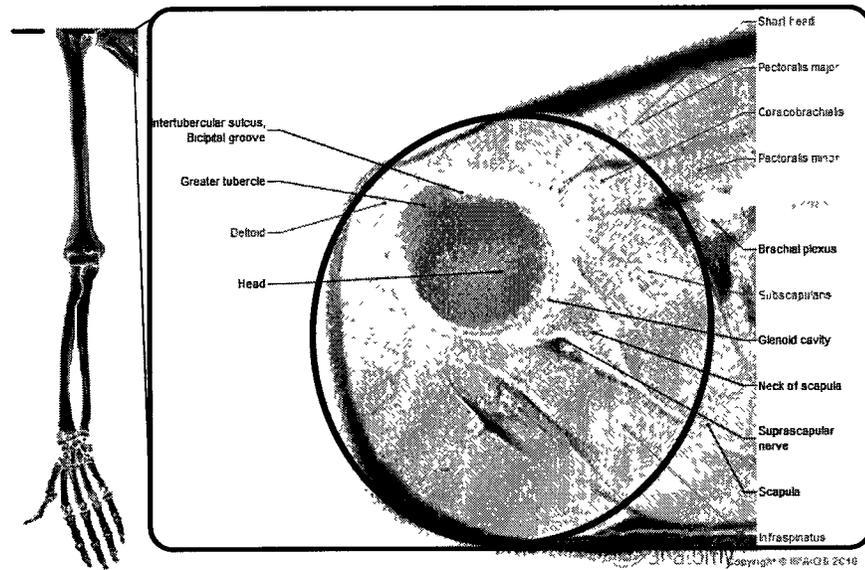


Figure 4.6: MRI of shoulder. (Reproduced with permission, [51])

4.3 Proposed approach based on 2D image analysis

In this work, 2D medial axis, or skeleton along with the contour of each image, is used to find a number of critical points in each of the camera views. Here, the critical points are the points in the articulation locations. In other words, first, we locate the joints in each view. Later, by combining this 2D information from all of the views, we model each joint in 3D relying on the inter calibration of the cameras between the views. As already mentioned in our work, we presume that the joints have elliptical shapes. By finding the parameters of these ellipses and considering that the bone ends have a fixed location at the articulations, the location of the bone ends can be estimated.

4.4 System description

The acquisition system used in this work is shown in Figure 4.7. It was composed of four Point Grey Research Flea2 IEEE 1394b Firewire cameras and two Pentium IV 3.4GHz. All cameras were mounted on a reconfigurable structure. In this system, each camera acquires one stream of video separately. The camera setup used to

monitor the arm movement occupies a 2.5m x 2.5m x 2.5m space. Each camera has a fixed focal length and the videos were recorded in 640x480 format with 15 frames per second. As shown in Figure 4.7, the PCs are connected to the four cameras. Both PCs are synchronized using a IEEE 1394a sync link.

The camera system has the following setup: Camera 0 is used as the reference camera and is mounted above the subject. Camera 2 is located on the right side of the subject aiming in the horizontal direction. Camera 1 is also located on the right side, above the camera 2 and pointing down left. Camera 3 is on the left and shooting towards down right. Cameras 1 and 3 are located about 200cm above the floor.

Before the experiment, the cameras are calibrated intrinsically and extrinsically. The calibration system was designed and implemented by Silvain Bériault [1] in the VIVA lab at the University of Ottawa. Camera calibration consists of the intrinsic and extrinsic parameter estimation. In the intrinsic calibration procedure, an 8x10 checkerboard pattern with 2cm x 2cm cell dimensions is utilized. For each camera, multiple views of the checkerboard are recorded. Using the OpenCV implementation of Zhang's method [53], the intrinsic camera parameters are estimated.

In extrinsic camera calibration, the calibration target consists of two LEDs mounted at the ends of a wood stick. By waving this tool in front of the cameras, a virtual cloud of 3D points is created. For every camera, the markers' pixel positions are recorded for each frame. When the markers are seen at least in two camera views, a match is recorded. After acquiring enough matches, using the 8-point RANSAC implementation provided by the OpenCV library, the fundamental matrix between pairs of cameras is computed. A retro projection accuracy of half a pixel is achieved using this calibration approach [1].

Following calibration, motion capture is performed as follows: While the subject moves his/her hands, the cameras record the movement. Each camera is connected to one of the two computers where the video streams are stored. After acquiring the videos, they are transferred to one computer for processing. Each frame of a video is grouped with the frames in other video streams that have the same timestamp. This unifies the start time and the length of the videos. In this system, the Visual C++ 2008 and OpenCV library package are used for image processing. Appendix A presents the application specifications.

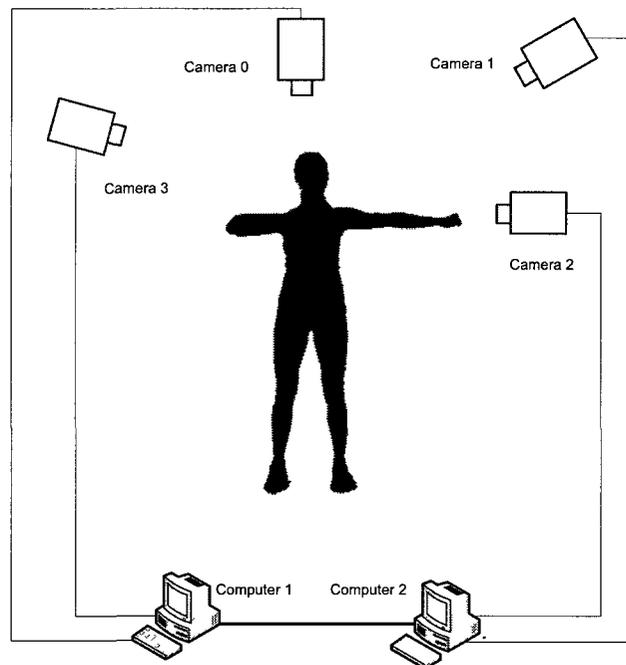


Figure 4.7: Acquisition system with 4 cameras and 2 synchronized computers.

4.5 Video analysis algorithm

Figure 4.8 shows the system flowchart and the two dimensional section which handles video capturing and estimating the location of the three joints and bones ends in the frames recorded by all cameras with the same timestamp. The inside flowcharts of the blocks for locating the elbow, wrist and shoulder in Figure 4.8, are shown in detail in Figures 4.18, 4.19 and 4.20 that will be explained in section 4.7.

For each video, frames are extracted and processed. The first frame of the first video is extracted. After preprocessing and background removal, we extract the medial line and contour of the image. In the next step, the elbow area is located on the skeleton and contour of the image. Two points are found on the contour of the image which are used for finding the wrist and shoulder areas. The method for locating the joints will be discussed in details in Section 4.9. For each joint, the two points located on the contour are saved to be used in the 3D process. Since we study three joints in the upper limb, 6 points on the contour are saved in this step. The above-mentioned process is repeated for the video frames taken by the other cameras, with the same timestamp. The result is sent to the 3D processing step. In this step the three dimensional position of the joints is estimated. This process is repeated for

all the timestamps.

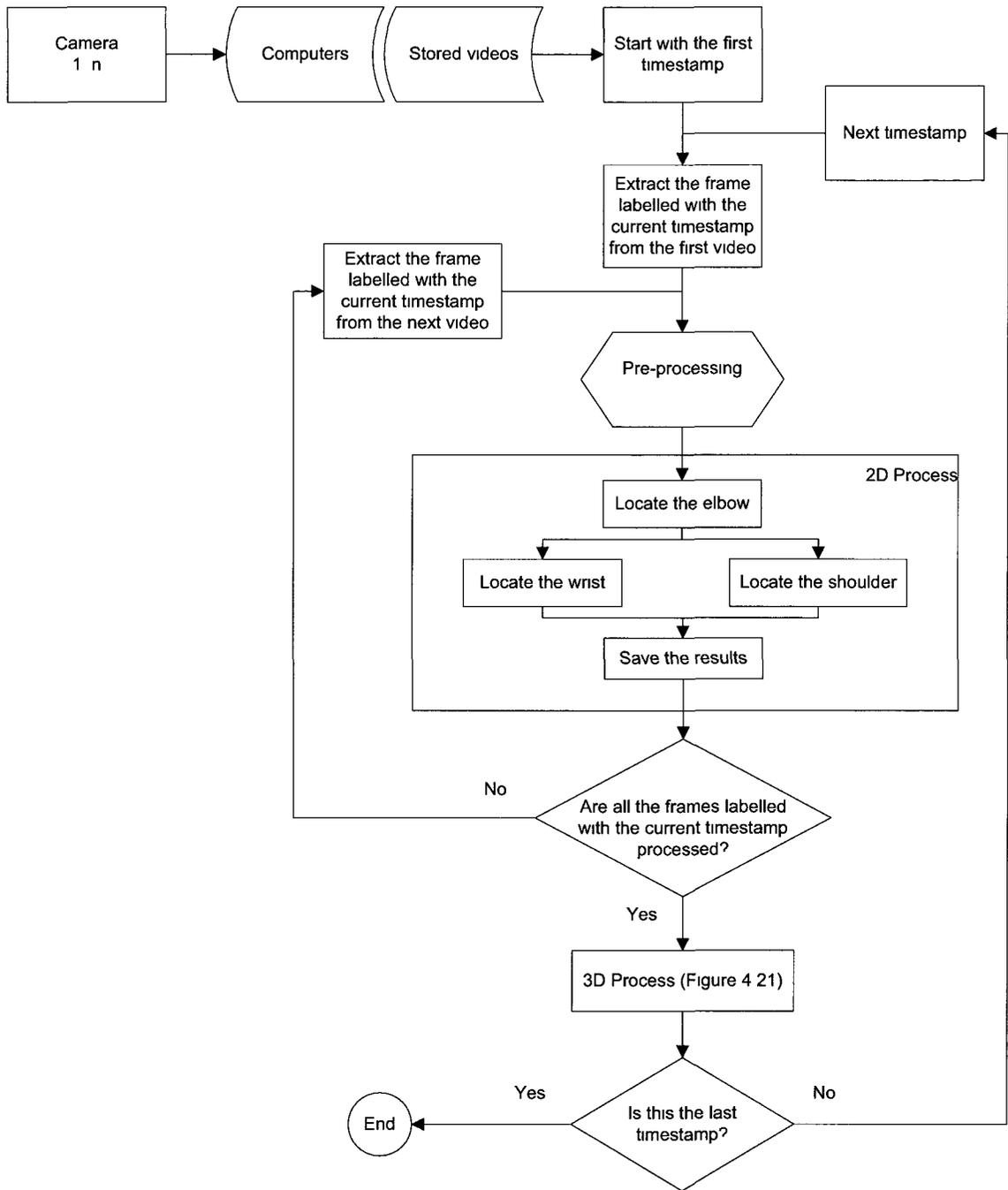


Figure 4.8: System flowchart.

4.6 Pre-processing

In the following section, the pre-processing and image representation steps are described.

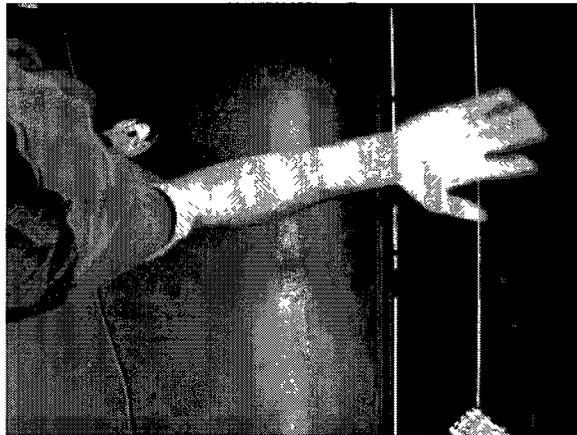
4.6.1 Background removal

In this work, we considered a statistical background removal method based on Mixture of Gaussians. We used the method proposed by KadewTraKuPong and Bowden [54] and is a built-in command in OpenCV. In our approach, background is learned during the first 10 seconds of the video which ideally does not contain the human candidate. Then background is considered as static. For each frame of the video, the difference between the frame and background is calculated for detecting the foreground in the image. Background should remain static for the rest of the acquisition. However, it was noticed that changes in the background, lighting, and shadow of the candidate affect the results significantly. In Figure 4.9 the effect of inappropriate lighting is shown. The reflection of light on the floor has affected the result which can be seen in the middle of the output image.

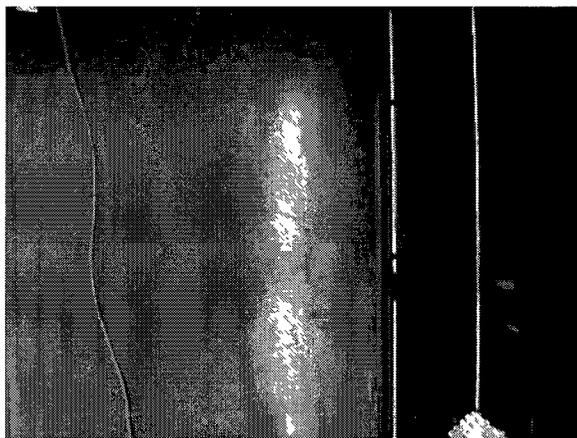
Moreover, it is important that the background color be chosen different from the skin color of the candidate. In general, the color of the clothes visible in the field of view of the cameras can affect the background removal accuracy and lead to false detection of background and foreground. As shown in Figure 4.9(c), part of the arm is detected as background because of the color of the shirt which was the same as the color of the floor. Also, wearing loose sleeve shirts or thick suits affect the precision of the system, especially for medical applications.

4.6.2 Skin detection filter

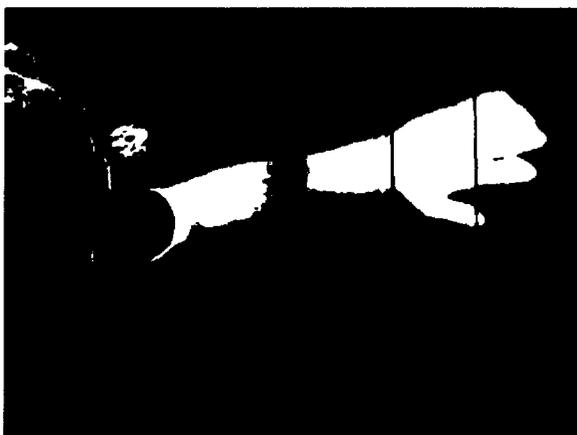
In order to make the system more precise, we used bare arm for recording the sample videos. In this way, not only the suits do not affect the output, but also by using a skin detection filter, the problems with the imperfect background model and shadow effects are mostly solved. We applied the HSV space skin filter proposed by [55]. In



(a)



(b)



(c)

Figure 4.9: Effect of inappropriate lighting: (a) input frame , (b) background (c) output frame.

this scheme, the pixels are classified as skin if the following conditions are all satisfied:

$$\begin{cases} S \leq 21, V \geq 2.5S \\ 158 \leq H + V \leq 400, H + V > 13S \\ H > 0.2V, H > 4S \end{cases} \quad (4.1)$$

where H, S and, V are hue, saturation, and value [56].

The ideal skin filter is independent of the skin color. It can be adjusted for patients suffering from diseases that change the skin color or birthmarks. Figure 4.10 shows the sample results of background removal and skin detection steps. In Figure 4.10(c), the generated silhouette is used for studying of various parts of the arm.

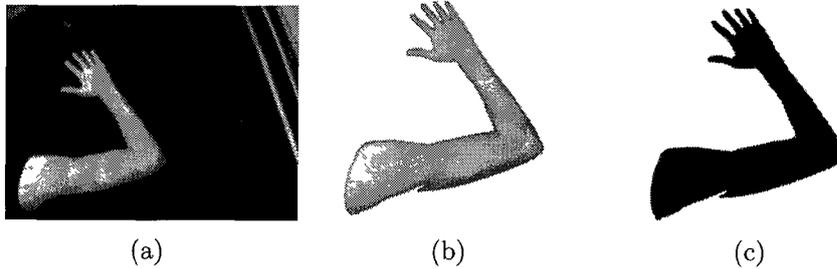


Figure 4.10: Background removal and skin detection: (a) input frame , (b) background removal and skin detection filter applied, and (c) output silhouette.

4.6.3 Noise effects

In the acquisition system used, different sources of noise impact on the captured images. The body hair on the arm, especially with a different color from the skin, can result in a very inaccurate output from the skin filter. Due to this fact, our results show that the edges of images tend to be jagged. To solve this problem, we must increase the quality of capturing and then we can simply filter the noise effects. Another noise source is the lighting. If the lighting is not uniform, hand shadow changes the brightness of the image.

4.6.4 Medial representation

As described in Chapter 2, the three major categories of the medial representation techniques are distance Skeleton, thinning Technique, and Voronoi Diagram.

Table 4.1 compares the skeletonization methods. While the distance transform preserves the geometry, the thinning method preserves the topology. The Voronoi-skeleton technique benefits from both topological and geometrical preservation.

Method	Geometrical	Topological
Distance transform	yes	no
Thinning	no	yes
Voronoi-skeleton	yes	yes

Table 4.1: Skeletonization methods comparison.

Figure 4.11 shows our results for the skeletonization of Figure 4.10(c) by: (i) distance transform skeleton, (ii) thinning technique skeleton, and (iii) Voronoi skeleton scheme. As shown in our results, the distance transform and thinning algorithm are more sensitive to noise. Meanwhile the Voronoi diagram skeleton appears to be the least sensitive to noise.

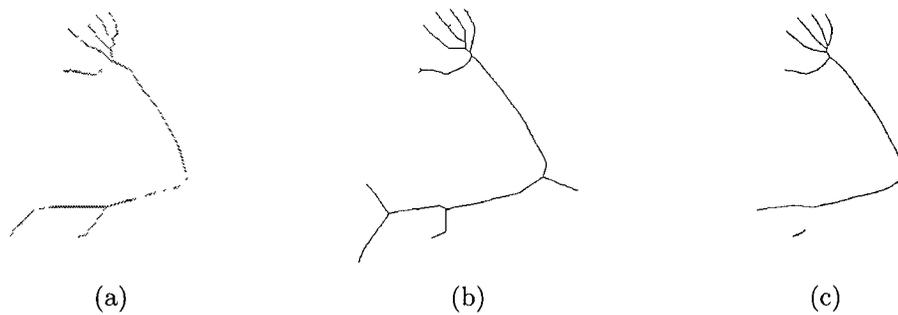


Figure 4.11: Results for the skeletonization of Figure 4.10:(a) distance transform skeleton, (b) thinning technique skeleton, and (c) Voronoi skeleton.

Table 4.2 shows the time elapsed for each skeletonization technique. The Voronoi diagram and the thinning method are more time consuming. In this regard, the distance transform method is the best choice among the others as it is more than

twice as fast as the thinning method. In fact, although the Voronoi diagram and thinning algorithms retain the topology and shape of the object, they are iterative processes, which make them expensive and time consuming.

Method	Camera 1	Camera 2	Camera 3	Camera 4
Thinning	0.9761	1.2202	1.2895	0.938
Voronoi-skeleton	4.8169	4.8191	5.4478	4.9544
distance transform	0.3831	0.5465	0.5471	0.389

Table 4.2: Time consumed for different skeletonization methods.[seconds]

After implementing all three methods of skeleton extraction, we concluded that:

1. The thinning method operates at a moderate speed but is the most sensitive to noise and often provides unreliable results.
2. The Voronoi diagram technique is the slowest method but the most accurate one. Using this technique, there is no need for any smoothing filter on the input image, which leads to the most accurate results.
3. The distance transform was found as the most applicable technique as it is more than eight times faster than Voronoi diagram. However, for using this method, smoothing the input image and pruning the skeleton are needed, which are less time consuming in total in comparison with the thinning method.

We decided to use the distant transform technique as it is the fastest. Note that, in our software, having a rough estimation of the joint locations on the medial axis is enough for finding the position of the joints on the contour. Therefore, although the distance transform method generates less accurate skeletons, it does not affect the output of the system significantly.

4.6.5 Contour detection

To find the contour of the image, we applied various methods of edge detection and compared their results. Generally, edge detection techniques can be divided into two categories: gradient and Laplacian. The gradient method searches for the edges by

looking for the maximum and minimum in the first derivative of the image (such as Roberts, Prewitt, and Canny filters). The Laplacian method looks for zero-crossings in the second derivative of the image (e.g. Marrs-Hildreth filter). In a research by Heath *et al.* [57], the Canny edge detector showed the best results for various inputs. We repeated the same experiment and the results from the Canny edge detector were less jagged and more accurate for our images. Figure 4.12 shows the results of three algorithms on our images. We opted to use the Canny gradient algorithm for edge detection.

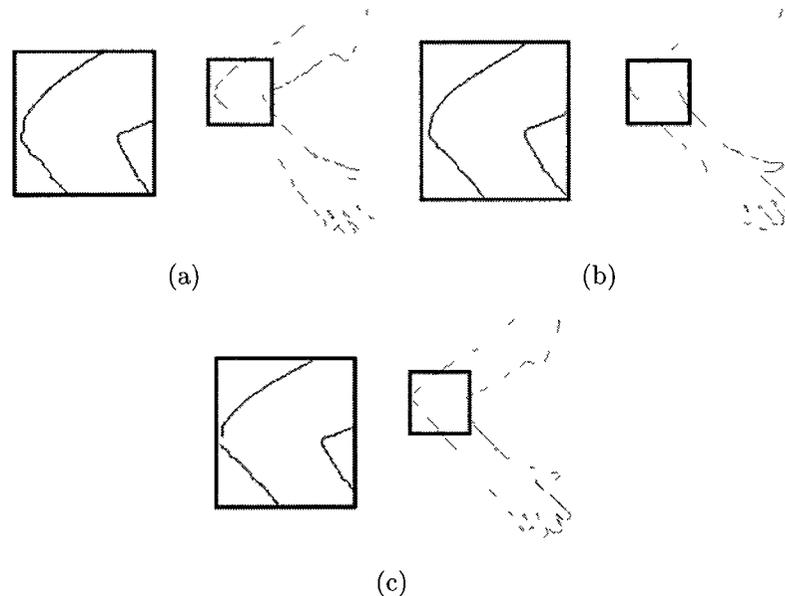


Figure 4.12: Results of the edge detection for methods: (a) Prewitt, (b) Canny and (c) Sobel.

4.7 Two dimensional joints location estimation

In this work, 2D medial axis, or skeleton, along with the contour of each image, is used to find a number of critical points in each of the camera views. The elbow holds the first set of critical points in the whole upper limb that we consider, because it can be distinguished easily when we bend the arm at the elbow. Actually, the sharp end of ulna at the elbow joint is relatively trivial to find, if the camera is located perpendicular to the plane containing the forearm and the upper arm. However, as we

work with several cameras and the hand is moving freely in the space, this assumption is not always valid. Figure 4.13 shows the elbow area and the corresponding critical points with various positions of the arm.

Obviously, locating the wrist is very difficult in images where the wrist and the hand are perfectly aligned from the lateral view, as shown in Figure 4.14. In addition, our system is unable to locate the elbow in this configuration, as no corners can be identified in the elbow area.

Another way to get around this problem is giving a routine to a patient under examination. For example fistng the hand makes it round and the wrist becomes easily distinguishable from any angle. However, this makes the examination system and procedure less flexible.

To find the location of the wrist, another method consists of computing a convex hull for the forearm and hand and studying the convexity defects [58] on a different view of the arm. The convex hull starts at the elbow area and covers the forearm and fingers. As shown in Figure 4.15, the sides of the wrist are located in the convexity defects of the contour. Two convexity defects which are closest to the elbow show the location of the wrist. In order to locate the wrist, the farthest points (WR_1 and WR_2) on the edge of the image from the convex hull are considered as the wrist area.

Figure 4.16 shows the elbow area and the located critical points for the wrist area.

On the other side of the elbow, at the end of the arm, the shoulder holds the third and final important set of points of interest. The shoulder contains several joints, as discussed in Chapter 3. The only skeletal part of the shoulder that is visible is the collarbone. The same technique that was used in the wrist and elbow location estimation is applied for the shoulder location. Figure 4.17 shows the wrist area and the corresponding critical points.

There are some special considerations in applying this algorithm. For instance, if a camera view is perpendicular to a joint, the joint location can be found in 2D with the maximum accuracy. If one part of the arm or body blocks a joint in one camera view, a part of the result will be affected. There are many cases where one of the joints cannot be distinguished, or even cases where just one point on the contour can be extracted in one of the views. In order to find the joints under these conditions,

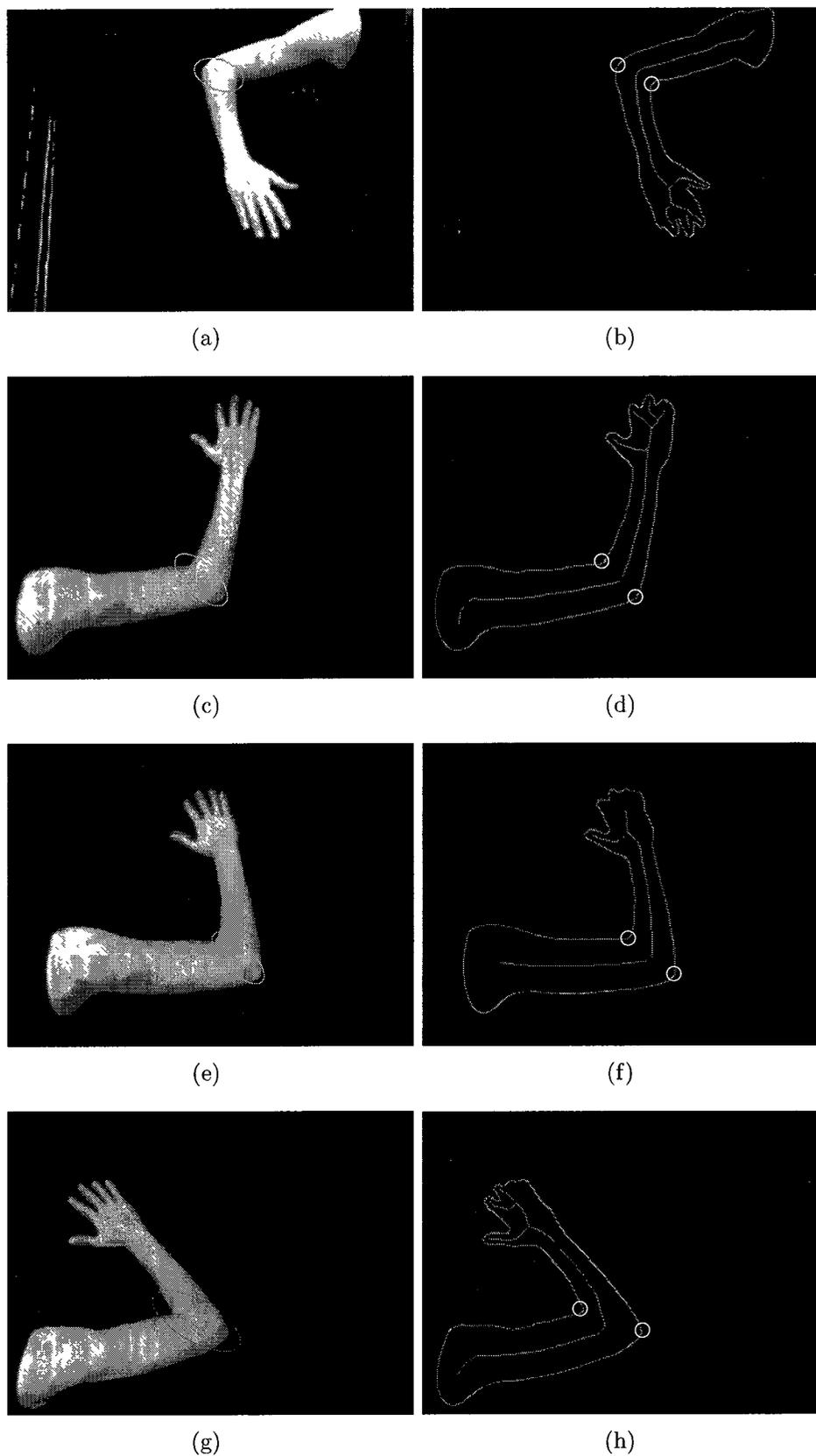


Figure 4.13: Locating the elbow joint.



Figure 4.14: Upper limb - lateral view.

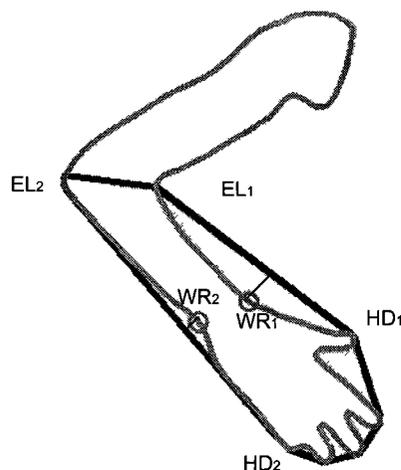


Figure 4.15: Convex hull estimated on the forearm and hand.

we propose a guided search technique that is detailed in Figures 4.18 and 4.19. If for any of the joints, the joint area on the medial line or contour can not be found, a derivative-based method is introduced that utilizes the location of the joint in the previous frame of the same view to help find the relevant region.

As shown in Figure 4.18, the first step is locating the desired point (shoulder/elbow/wrist) on the medial line. For finding the elbow, the point with the highest second derivative on the medial line is considered as the elbow location. For testing the estimated location of the elbow, the second derivative of the medial line at the elbow joint must be larger than 0.32. This lower limit was set by trial and error. The

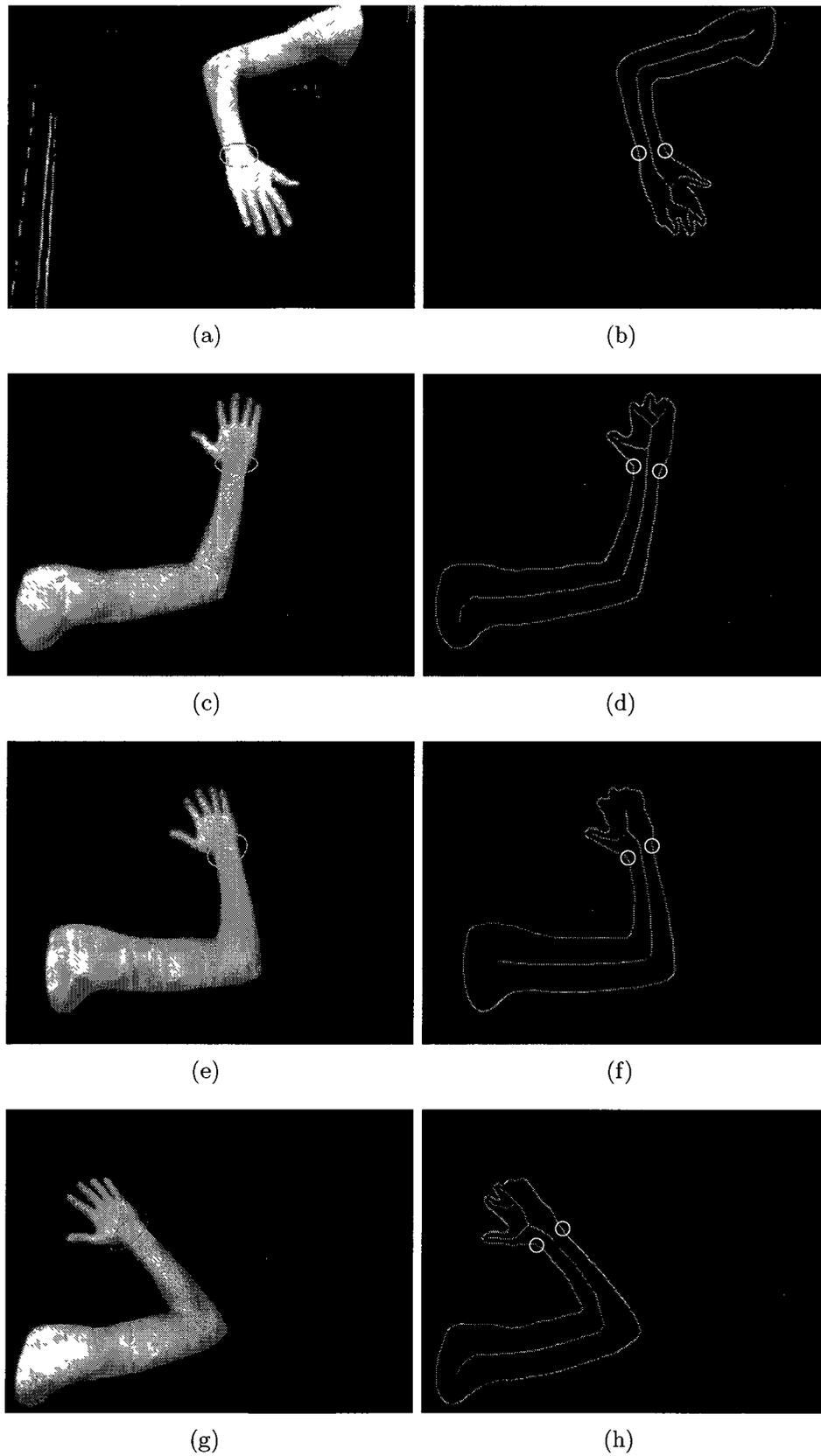


Figure 4.16: Locating the wrist joint.

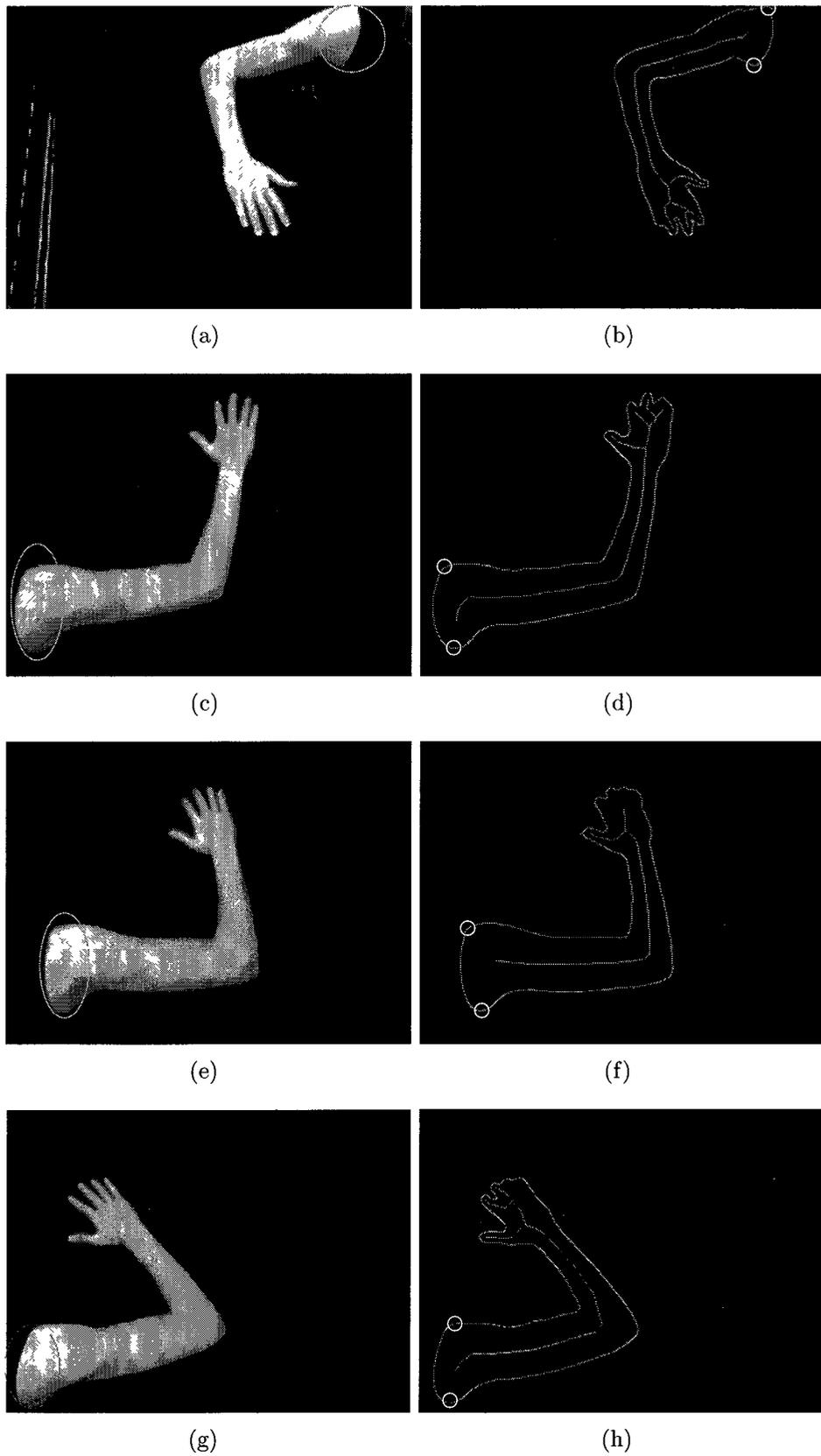


Figure 4.17: Locating the shoulder joint.

wrist and shoulder are located on the two ends of the longest branch of the medial line. Moreover if the length of the longest branch of the medial line is less than 100 pixels, the estimated locations of the wrist and shoulder are considered as invalid. As shown in Figure 4.19, if the proposed system successfully finds the desired point on the medial line, the estimated location is used to find the corresponding points on the contour. For the elbow, two points on the contour which are closest to the joint location on the medial line are considered as the joint area and reported to the 3D analysis system. If for any reasons (such as the situation in Figure 4.14, where the medial line is almost straight and the elbow can not be determined), the desired point could not be found on the medial line, the proposed system tries to locate the point using the point position extracted from the previous frame by utilizing a derivative-based locating algorithm. Whether the derivative-based locating algorithm fails to locate the desired joint on the medial line, the system tries to find the desired joint area on the contour directly using the derivate-based method.

The derivative technique finds the point with the maximum derivative which is located in a neighbourhood of the previous joint location. As shown in Figure 4.20, using the location of the desired point in the previous frame as the center, we locate a point within the radius R on the contour/medial line where the second derivative of the contour/medial curve is highest. The value of R was chosen equal to 5% of the medial line, which is a number between 10 to 20. D_0 and D_1 present the second derivative of the contour/medial line at the joint area in the previous and current frames, respectively.

If D_0 and D_1 satisfy the conditions of equation 4.2, the new point is accepted. D_{min} and Δ are two parameters which are set manually. Using these parameters we can make sure that the point found in the current frame is the best candidate for the joint location. As the arm moves steadily during the examination, the change in the joint angles is happening very smoothly. Hence, the values of D_1 and D_0 should be close to each other in two consecutive frames. The first equation checks if the values of D_1 and D_0 are close enough to each other. In the second equation, D_{min} is a minimum acceptable value for the derivative of the joint location on the contour/medial line. By increasing the value of D_{min} , the joint can not be located in more frames. If D_{min} is set too small, the identified point can be less trustworthy. For example, when the

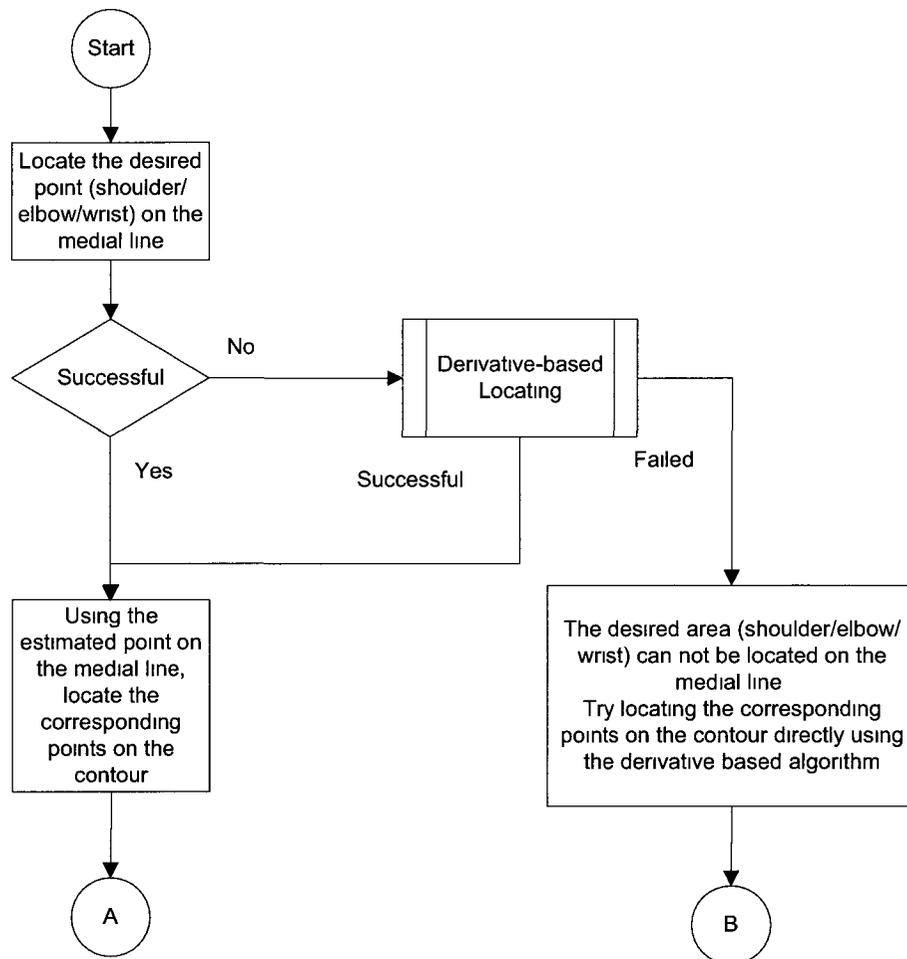


Figure 4.18: Locating the areas of interest, step 1: locating on the medial line

arm is held straight, the medial line in the elbow area becomes a straight line and the second derivative of the medial line at any point on the medial line around the elbow becomes 0. Therefore if we set D_{min} equal to 0, then any point at the elbow area can be considered as an acceptable candidate. For solving this ambiguity, we use D_{min} to set the minimum angle at the candidate points on the medial line and contour. By increasing the value of D_{min} , the minimum identifiable angle at the joints is increased and the located point is more reliable. We set the value of D_{min} by trial-and-error. The best working values of D_{min} and Δ were 0.32 and 0.078, respectively.

$$\begin{aligned}
 |D_1 - D_0| &< \Delta \\
 D_1 &> D_{min}
 \end{aligned}
 \tag{4.2}$$

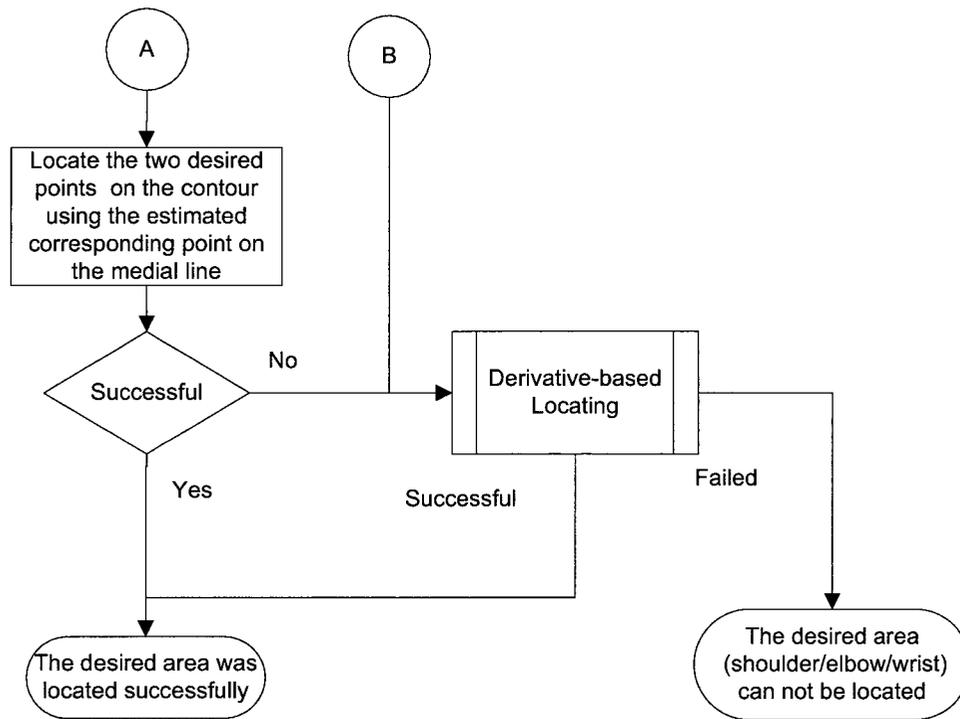


Figure 4.19: Locating the areas of interest, step 2: locating on the contour.

4.8 Three dimensional processing

As shown in Figure 4.8, after locating the joints areas in the frames, they are sent to the 3D processing system. Figure 4.21 shows the diagram for the 3D processing system proposed in this thesis. The blocks shown in this figure are discussed in this section.

Applying the technique described in the previous sections, the joint areas can be determined in all camera views. For each joint, two points on the contour of the image are extracted. Overall a total of $(\text{number of points for each joint}) \times (\text{number of joints}) \times (\text{number of camera views})$ points are located in 2D images. Since we have three joints (shoulder, elbow and wrist), we have $2 \times 3 \times X$ points, where X represents the number of camera views that the system processes and where it locates the joints successfully. The following step consists of combining all the 2D points and calculate a 3D model of the upper limb.

In Figure 4.7, we consider Camera 0 as the reference camera and transfer the extracted points on the other camera views to the reference camera view. In other words, for every point found in a given camera view, we calculate the corresponding

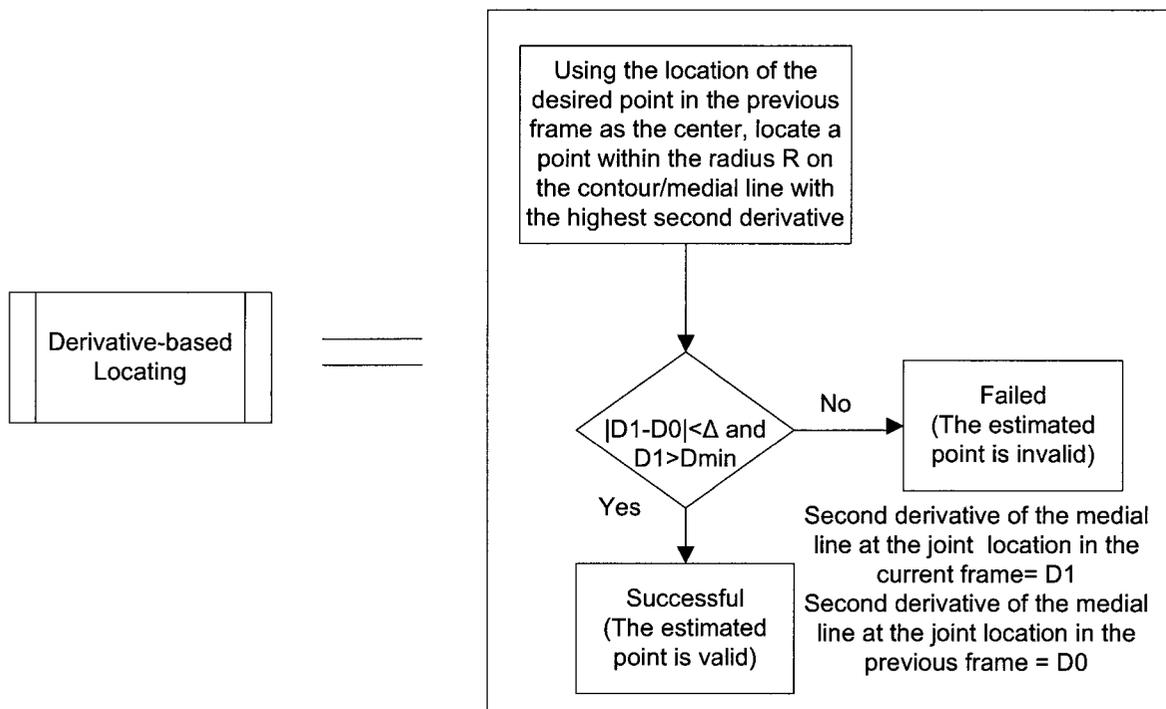


Figure 4.20: Derivative algorithm description.

epipolar line in the reference camera view. Therefore, for every joint, we have several epipolar lines and two points from the reference camera view itself.

In Figure 4.22, the lines L1 and L1' are the epipolar lines of the corresponding points from camera 1 to the reference camera view. Similarly the lines L2 and L2' are the epipolar lines of the points from camera 2 and so on. The points P0 and P0' are the estimated points in the reference camera view.

For each joint, an ellipse is fitted tangent to the corresponding epipolar lines. Moreover, this ellipse ideally must pass through the points P0 and P0'. We call this ellipse, "Perspective Ellipse (PE)". Fitting the ellipse to the data is done using the least-squares method, so that the sum of the squares of the distances to the points P0 and P0' and the squares of the distances to the lines is minimal.

The cross section of the upper limb at a joint is the intersection of a plane which is perpendicular to the medial line of the body at the joint. As seen in section 4.2.3, such a cross section can be estimated as an ellipse. This ellipse is called "Original Ellipse (OE)".

As a matter of fact, PE is the projection of OE on the reference camera plane.

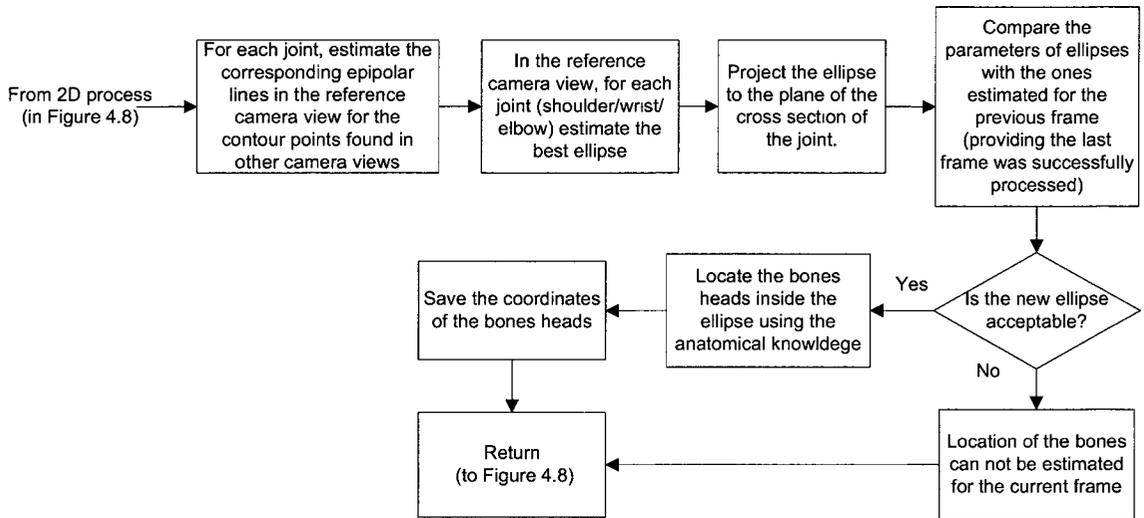


Figure 4.21: Three dimensional process.

Our goal is to estimate the original ellipse using the projected ellipse. This can be achieved by projecting PE on the plane perpendicular to the body at the joint. The result is the original ellipse.

Figures 4.23 shows the original ellipse of the wrist cross section and its perspective

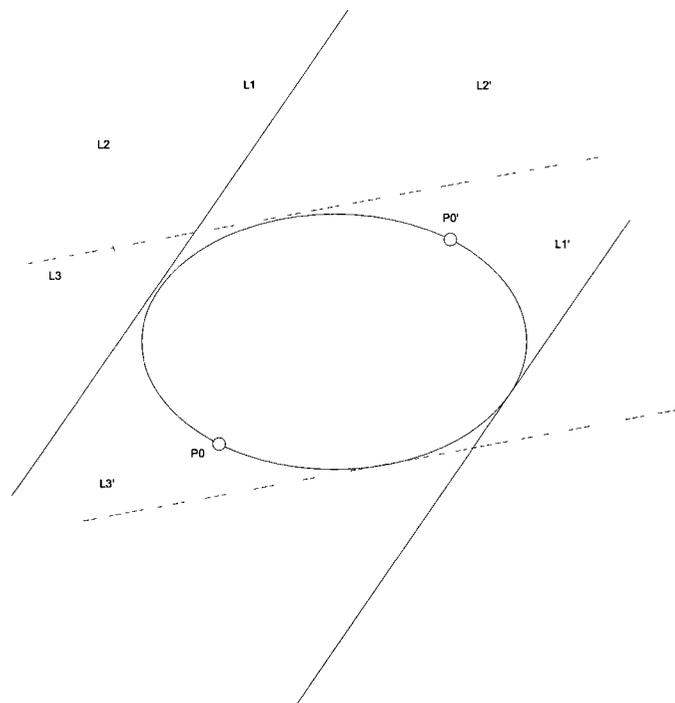


Figure 4.22: Ellipse fitting.

ellipses from different cameras views.

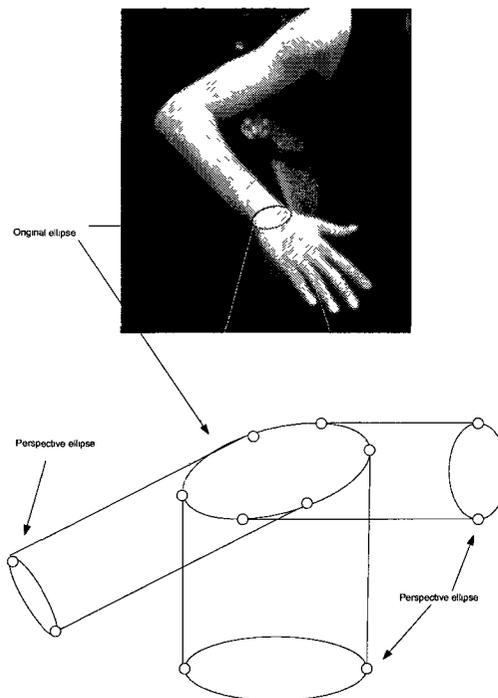


Figure 4.23: Original ellipse located on the wrist and the perspective ellipses seen from the camera views.

An ellipse parametric form equation is given by

$$\begin{aligned} X(t) &= X_c + a \cos t \cos \phi - b \sin t \sin \phi \\ Y(t) &= Y_c + a \cos t \sin \phi + b \sin t \cos \phi, \end{aligned} \quad (4.3)$$

where the parameter t varies from 0 to 2π , (X_c, Y_c) is the center of the ellipse, and ϕ is the angle between the X-axis and the major axis of the ellipse. The aforementioned parameters define a unique ellipse. For finding the parameters, five constraints are needed such as known points on the ellipse perimeter or known tangent lines to the ellipse.

By 2D processing on each camera view image as described in the last section, we locate two points for each joint, one on each side of the contour. We project all of the points from other views to the reference camera view. The perspective projection of a point in a view is a line in other views. As the points are located on the cross section of the wrist, their perspective projection are lines tangent to the ellipse. As a

result, by projecting the points from other views, in each view we obtain two points on the ellipse and several lines, which are tangent to the ellipse.

An example of a given view is shown in Figure 4.22. An ellipse is drawn which passes through the points of interest extracted in one view and is tangent to each line. To solve the system of five equations we use two points from the same view that provide two equations. Three tangent lines provide 3 more equations. This means that we need a minimum of three cameras positioned properly to have a clear view of the wrist to generate enough tangent lines. It is obvious that having more cameras will help covering the junctions with more views and decrease the error.

As shown in Figure 4.24, we consider the ellipses on the focal planes of the cameras as the conic sections of oblique cones with their apexes located on the cameras. A common conic section of these cones is the cross section of the joint where the cones intersect with the plane of cross section of the joint. Therefore knowing the camera location and focal length of the camera, the focal point position and therefore the cone parameters can be estimated. We repeat this procedure for another camera and find the corresponding ellipse and cone. The intersection of the wrist is located on

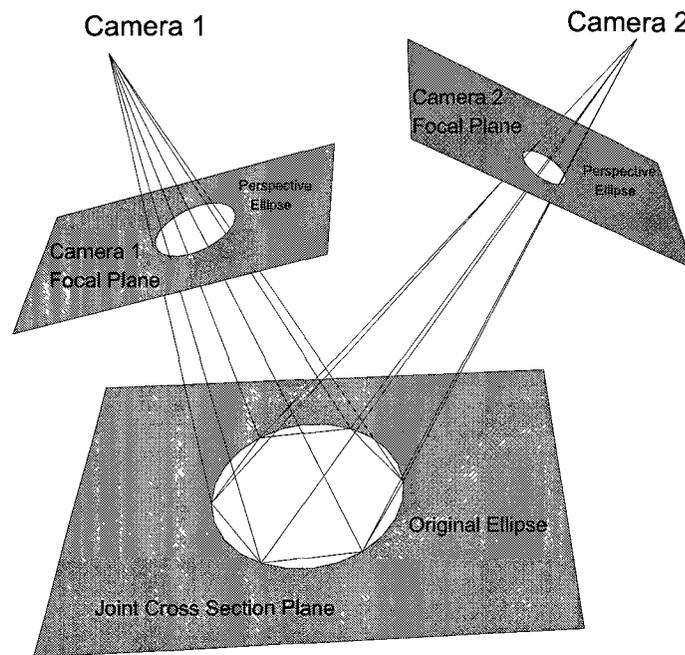


Figure 4.24: Elliptical cones intersection.

both cones.

Equation 4.4 shows the equation of an elliptical cone. Point (x_0, y_0, z_0) is the apex of the cone. For $0 \leq z - z_0 \leq c$, where the origin is at the vertex, the z-axis is the axis of the cone, the plane $z = z_0 + c$ is the plane of the base, the semi-major axis of the elliptical base is the line segment $z = z_0 + c, y = 0, 0 \leq x - x_0 \leq a$, and the semi-minor axis of the elliptical base is the line segment $z = z_0 + c, x = 0, 0 \leq y - y_0 \leq b$.

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = \frac{(z - z_0)^2}{c^2} \quad (4.4)$$

A general form of the cone equation with the arbitrary axis can be expressed as:

$$Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Fyz + 2Gx + 2Hy + 2Iz + J = 0 \quad (4.5)$$

By intersecting two cones, the equation of the intersecting ellipse can be found by solving this equation set:

$$\begin{cases} A_0x^2 + B_0y^2 + C_0z^2 + 2D_0xy + 2E_0xz + 2F_0yz + 2G_0x + 2H_0y + 2I_0z + J_0 = 0 \\ A_1x^2 + B_1y^2 + C_1z^2 + 2D_1xy + 2E_1xz + 2F_1yz + 2G_1x + 2H_1y + 2I_1z + J_1 = 0 \end{cases} \quad (4.6)$$

It is important to note that solving the above set of equations to find the intersection of two oblique circular cones is a complex problem. An alternative is therefore proposed.

In our approach for estimating the original ellipses, first we find their centre. In two dimensional processing of the images, two points in each joint areas are estimated. These two points are located on the opposite sides of the ellipse, therefore the centre of the ellipse is located in the midpoint. Considering that the centre of the original ellipse is unique, in every camera view, we find the midpoint of the extracted points for each joint using Equation 4.7.

$$\begin{cases} x_{centre} = (x_{P1} + x_{P1'})/2 \\ y_{centre} = (y_{P1} + y_{P1'})/2 \end{cases} \quad (4.7)$$

These 2D midpoints are the projection of the centre of the original ellipse on each

camera view. For each joint, the following formula gives the centre of the joint (i.e. centre of the original ellipse at the joint). Having the coordinates of the centre of the joint in two views, we find the point coordinates in 3D using triangulation. Hence, we are able to estimate the centre of the ellipses at the wrist, elbow, and shoulder in 3D.

The planes perpendicular to the body at the joints are shown in Figure 4.25. For

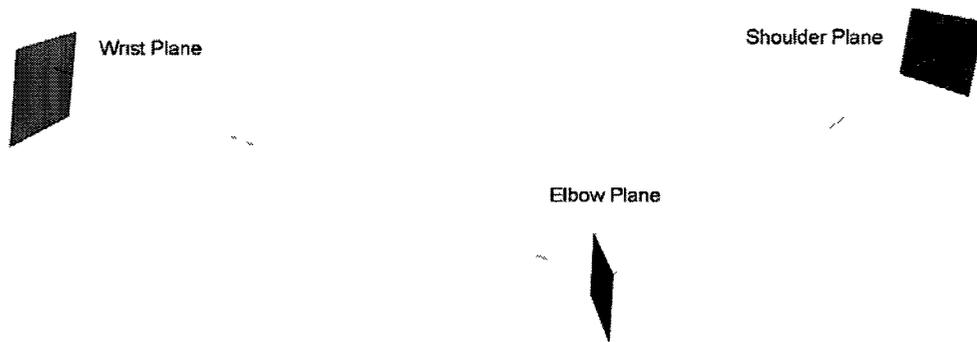


Figure 4.25: Joints cross section planes.

estimating the parameters of a plane, one point on the plane and the normal vector to each plane are needed. Since we have the centre of the original ellipses, a point on each plane is known. The vector of the planes at the wrist, elbow, and shoulder joints are shown in Figure 4.26 and can be estimated from the Equation 4.8

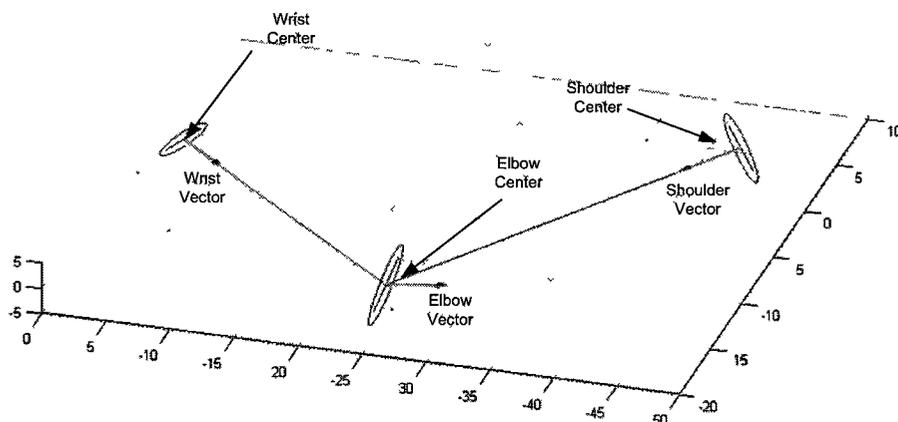


Figure 4.26: Joints cross section vectors.

$$\begin{cases} V_{WR} = C_{EL} - C_{WR} \\ V_{EL} = C_{SH} - C_{WR} \\ V_{SH} = C_{EL} - C_{SH} \end{cases} \quad (4.8)$$

where C_{EL} , C_{WR} , and C_{SH} are centres of elbow, wrist, and shoulder respectively.

Now, we have the centre and the vector of the joints, therefore the joint cross section planes are known. To find the original ellipses at the joints, we consider an oblique cone with the apex at the camera position. The intersection of the oblique cone and the corresponding joint cross section plane forms the original ellipse for each joint. In order to find the intersection, we use eight points on the perspective ellipse and draw eight lines, each passing through one point and the focal point of the camera. As shown in Figure 4.27, these lines intersect the joint cross section plane at eight points, which are located on the original ellipse. By fitting an ellipse to these points the cross section of the wrist joint is estimated.

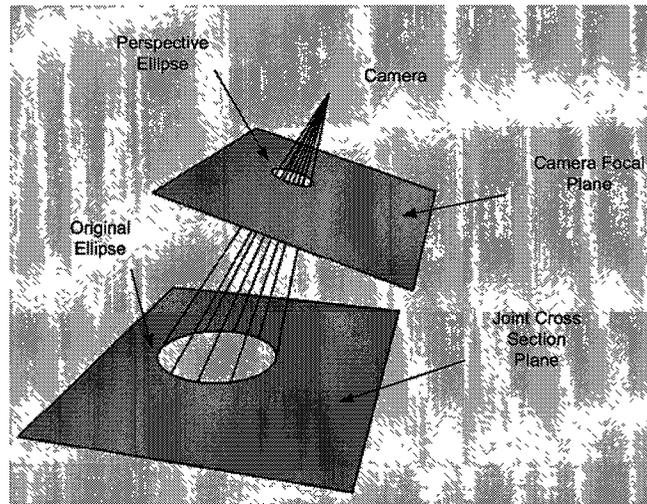


Figure 4.27: Elliptical cones and their intersection.

The major diameter of the fitted ellipse is compared with the major diameter of the last ellipse found for the joint in previous timestamps. If the major diameter is very different from the last acceptable one, the estimated ellipse is not acceptable.

4.9 Locating bone ends

In this thesis, we model the arm cross sections at the joints with ellipses. These ellipses are shown in Figures 4.28, 4.29 and 4.30 for the elbow, the wrist, and the shoulder respectively.

Figure 4.28 shows the elbow joint cross section and the fitted ellipse. We observe different bones ends in the cross section of the elbow. The points P1 and P2 show the end of the radius and ulna bones respectively. Moreover we consider the end of the humerus close to the end of the ulna. Although bending the elbow changes its outer shape, the position of the bones remains almost fixed in the elbow articulation. The point P1 is located close to the centre of the ellipse (P1'). The point P2 is very close to the intersection of the minor axis and the perimeter of the ellipse (P2'). Therefore by estimating the corresponding ellipse, the points P1 and P2 can be approximated by P1' and P2' respectively.

Figure 4.29 illustrates the wrist cross section and the corresponding ellipse. The dark discs show the bone ends. In the cross section of the wrist, the ends of ulna and radius bones are located on the major axis of the corresponding ellipse, and can be approximated as the foci of the ellipse.

According to the shoulder MRI picture shown in Figure 4.30, the head of the humerus bone is almost located in the centre of the shoulder. As we fit an ellipse on the cross section of the shoulder, we are able to approximate the end of the humerus as the centre of the ellipse.

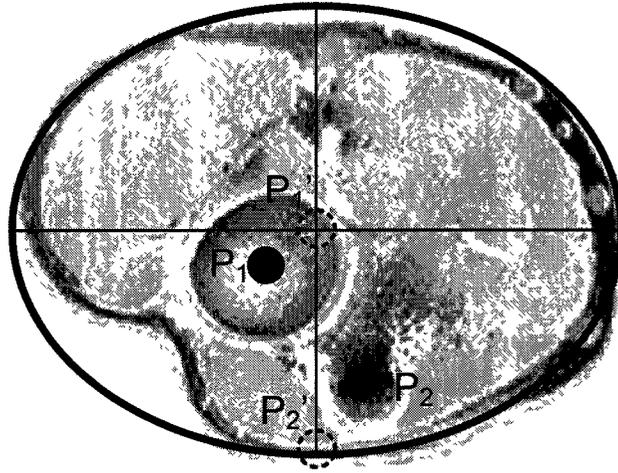


Figure 4.28: The cross section of the elbow joint and the corresponding geometry model. (Reproduced with permission, [51])

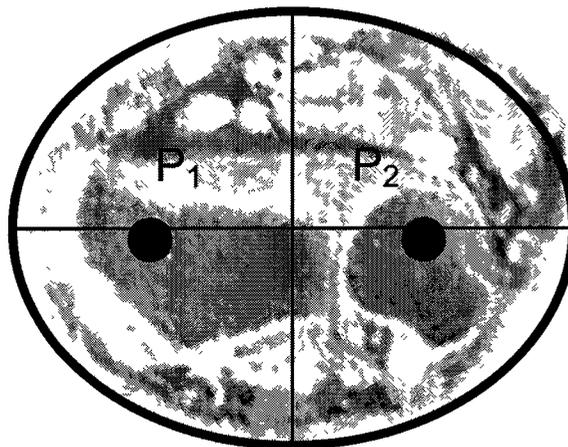


Figure 4.29: The cross section of the wrist joint and the corresponding geometry model. (Reproduced with permission, [51])

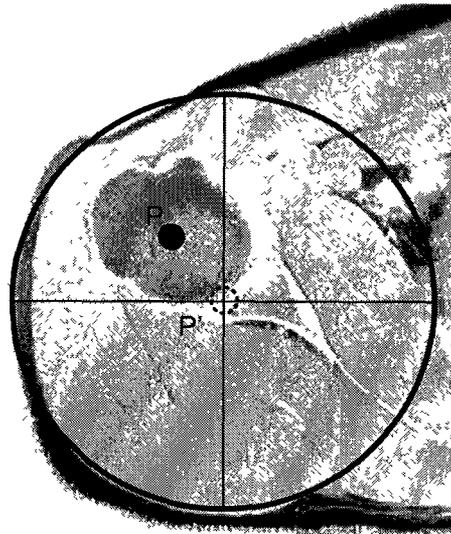


Figure 4.30: The cross section of the shoulder joint and the corresponding geometry model. (Reproduced with permission, [51])

Chapter 5

Experimental Work and Results

In this chapter, we explain the proof-of-concept and discuss the results. In the first section, we review the image acquisition system used in our experiments. In the second and third sections, we explain our method for calculating the DoF and measuring the bone lengths along with the corresponding results of the system. The discussion concludes this chapter.

5.1 System description

The acquisition system used in this work is shown in Figure 5.1. It is composed of four Point Grey Research Flea2 IEEE 1394b Firewire cameras and two Pentium IV 3.4GHz. As shown in Figure 5.1, a reconfigurable structure gives the possibility to adjust the alignment and position of the cameras. For lighting, some stand lamps were utilized. For clarity, we encircled the cameras in the illustration. We used 350 frames of video from each camera to test the proposed system. The first 10 frames were used for background detection. We dropped the next 90 frames because, during this time, the candidate was entering in the scene and they were not suitable for our experiments. Finally, the next 250 frames were used for testing the system, which hereafter, will be numbered 1 to 250.

The system was tested by a male candidate, with the specifications shown in Table 5.1. All the measurements were done by a medical doctor. The measurements were taken at the right arm with a tape measure.

We asked the candidate to stand and raise his right arm in front of him to shoulder height and let his palm face down. While he kept his arm parallel to the floor, he



Figure 5.1: Acquisition system.

	Forearm		Upper Arm
Bone	Ulna	Radius	Humerus
Measured Length	11.3	10.5	11.5

Table 5.1: The measured lengths of the major upper limb bones by a doctor [inches].

bent the elbow 90 degrees. This was the starting position (frame 1). Next, we asked him to gradually straighten his elbow while he kept his arm parallel to the floor and the upper arm stood still. This movement was recorded in frames 1 to 35. Next we asked him to rotate his palm and let his palm face left (frames 36 to 57). After that, he raised up his forearm while he kept his upper arm stood still and bent his elbow, as much as possible, vertically (frames 58 to 101). Then he moved down his forearm (102-130) and rose up and down one more time (frames 131 to 158). At the end, he was asked to rotate his wrist to let his palm face down (frames 159 to 186) and bent and opened his elbow horizontally, exactly as he did in the beginning (recorded in frames 185 to 250). After processing the videos, the joints were detected in each image and 3D coordinates were estimated. Using the 3D coordinates, the joint angles

for five degrees of freedom of our model were obtained. Moreover, the length of the bones were determined. Figure 5.2 shows sample frames of the movement sequence.

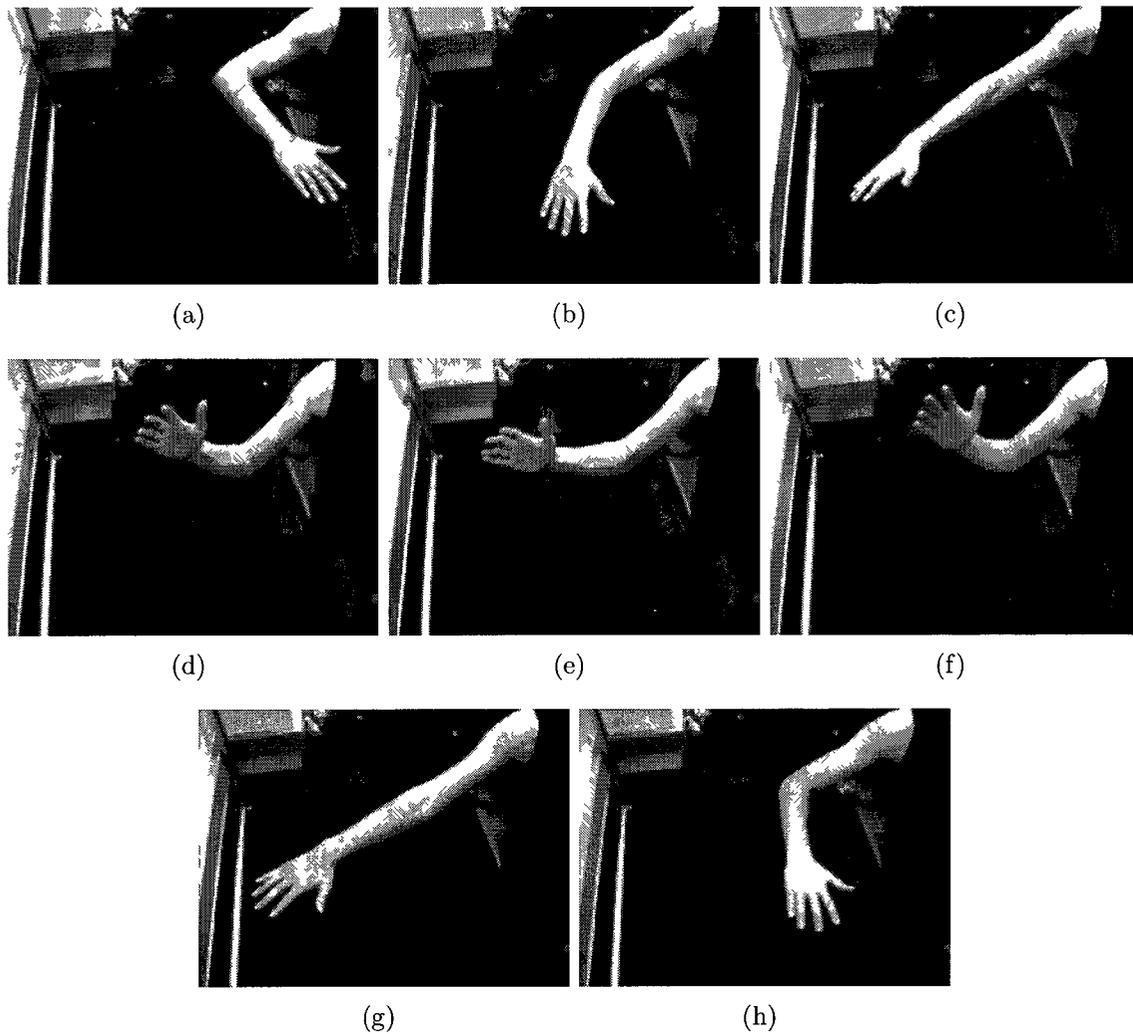


Figure 5.2: Some movement sample frames: (a) frame 1, (b) frame 30, (c) frame 50, (d) frame 80, (e) frame 130, (f) frame 155, (g) frame 200, (h) frame 225

5.2 Degrees of freedom

The model of upper limb proposed in the thesis was explained in Chapter 4. Table 5.2 summarizes the degrees of freedom for the model. The method for estimation of the DoFs will be explained in this section.

Articulation				
Shoulder			Elbow	Wrist
Pitch	Yaw	Roll	Pitch	Roll

Table 5.2: The studied degrees of freedom in each articulations.

5.2.1 Shoulder yaw

For estimating the yaw angle we project the upper arm (humerus bone) onto the floor plane, as shown in Figure 5.3.

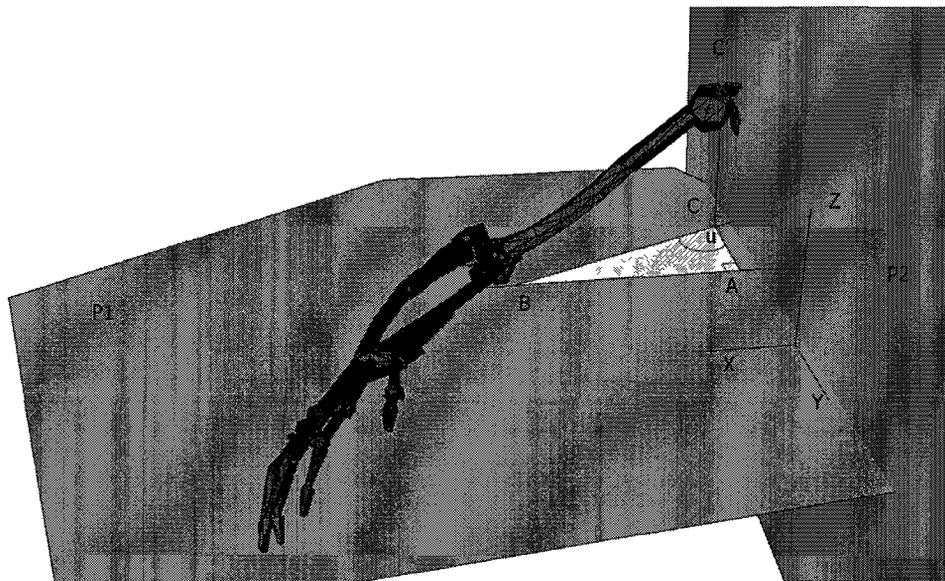


Figure 5.3: Shoulder yaw.

The vertical plane (P2) is an arbitrary plane perpendicular to both the floor (plane P1) and the subject's chest plane. The chest plane is the plane which contains the

line connecting the shoulders and is perpendicular to the floor. The points B and C' are the ends of the humerus bone. Line segment AB is on the plane P1 and is parallel with the floor. Angle u is the yaw which is calculated from the formula 5.1.

$$u = \arctan\left(\frac{AB}{AC}\right) = \arctan\left(\frac{x_{shoulder} - x_{elbow}}{y_{shoulder} - y_{elbow}}\right) \quad (5.1)$$

Figure 5.4 shows the estimated values for the yaw angle at the shoulder. The horizontal axis represents the frame number, while the vertical axis shows the angle in degrees. As the forearm is moving the upper arm moves horizontally. The estimated value for the yaw angle varies from 2 degrees to 18 degrees as the arm moves. Overall, the system has been able to track the yaw angle with small fluctuations.

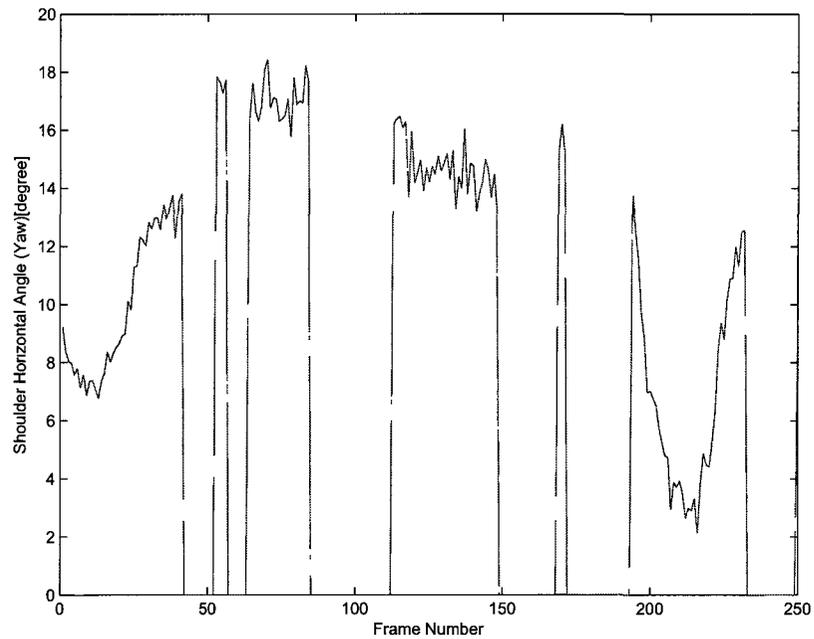


Figure 5.4: Estimated shoulder yaw angles [degrees].

For evaluating the system, 7 frames with known shoulder yaw angles were used. All the real values are from the test samples in which the candidate posed with a known value for the yaw. For each angle, the shoulder yaw angle was measured using a protractor before taking the videos. Table 5.3 shows the real and estimated values for the shoulder yaw. There is a maximum 4 degrees difference between the estimated

Real	0	10	20	30	40	50	60
Estimated	1.863	13.238	20.547	32.899	42.024	53.642	62.309
Error	1.863	3.238	0.547	2.899	2.024	3.642	2.309

Table 5.3: Shoulder yaw estimated and real values [degrees].

and real values for the shoulder yaw for each upper limb position.

5.2.2 Shoulder pitch

For estimating the pitch angle, the upper arm (humerus bone) is projected onto a vertical plane, as shown in Figure 5.5.

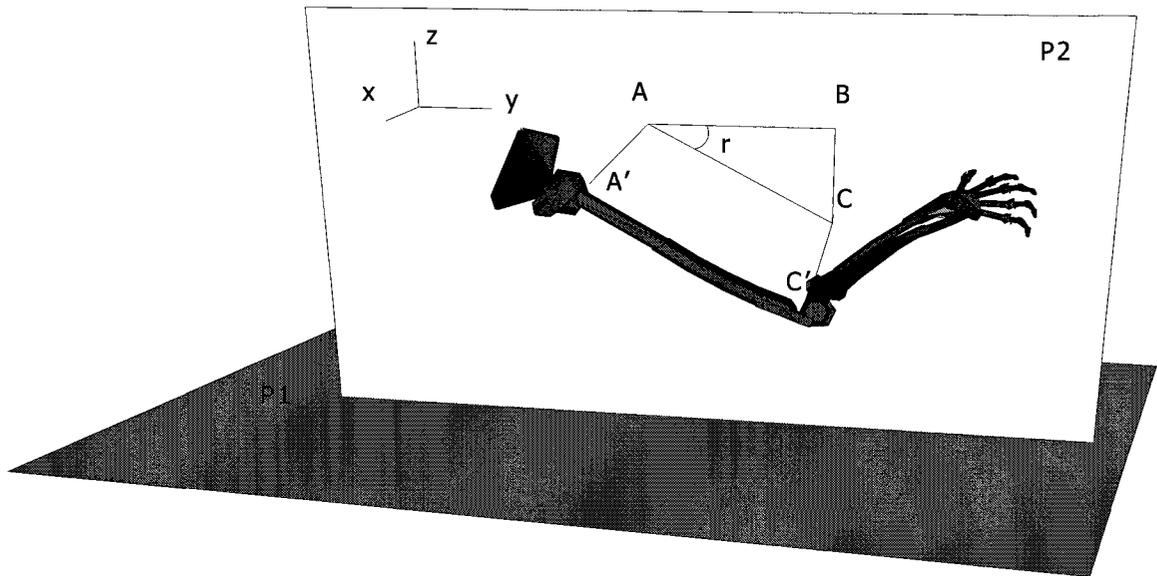


Figure 5.5: Shoulder pitch.

The points A' and C' are the ends of the humerus bone. Line segment AB is on the plane P2. 'r' represents the pitch angle which is calculated from the formula 5.2.

$$r = \arctan\left(-\frac{BC}{AB}\right) = \arctan\left(-\frac{z_{\text{shoulder}} - z_{\text{elbow}}}{y_{\text{shoulder}} - y_{\text{elbow}}}\right) \quad (5.2)$$

Figure 5.6 shows the estimated values for the pitch angle at the shoulder. The horizontal axis represents the frame number, while the vertical axis shows the angle

in degrees. As the forearm is moving the upper arm moves horizontally. The total pitch range is 14 degrees. The pitch value starts at 6 degrees because we asked the candidate to hold his upper arm almost straight. The minimum estimated pitch angle is in frame 32 which is 4.736 degrees and later it is increased up to 18.97 degrees. As the arm moves, we see these fluctuations in the pitch value.

All the gaps in the figures are the situations where our system cannot estimate the bone ends and consequently, it cannot calculate the pitch angle.

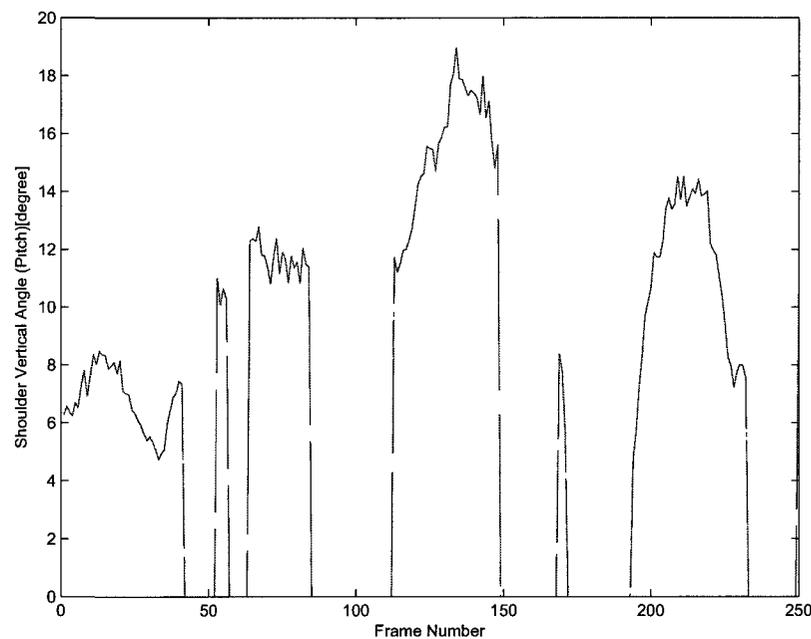


Figure 5.6: Estimated shoulder pitch angles [degrees].

For testing the system, 7 frames with known shoulder pitch angles were used. Table 5.4 shows the real and estimated values for the shoulder pitch. The pitch angle range for our test frames is 60 degrees. For taking the sample frames, the candidate stood and straightened his right arm in front of him. He bent the elbow, 100 degrees. He tilted his upper arm vertically from -30 degrees to +30 degrees. The pitch angle has been estimated with a good accuracy. The difference between the real and estimated value is always less than 3 degrees.

Real	-30	-20	-10	0	10	20	30
Estimated	-32.816	-20.254	-11.039	-1.667	9.52	18.907	31.051
Error	-2.816	-0.254	-1.039	-1.667	-0.48	-1.093	0.051

Table 5.4: Shoulder pitch estimated and real values [degrees].

5.2.3 Shoulder roll

Figure 5.7 shows the scheme for finding the shoulder roll. Plane P1 is the floor and plane P2 which has the r angle related to the plane P1, is perpendicular to the humerus bone at the elbow. In general, the forearm is not in plane P2. In the case shown in Figure 5.7, the elbow angle is more than 90 degrees and the forearm is located behind the plane P2. The line segment BC is the projection of the line segment BC' onto the plane P2, and s represents the roll angle at the shoulder.

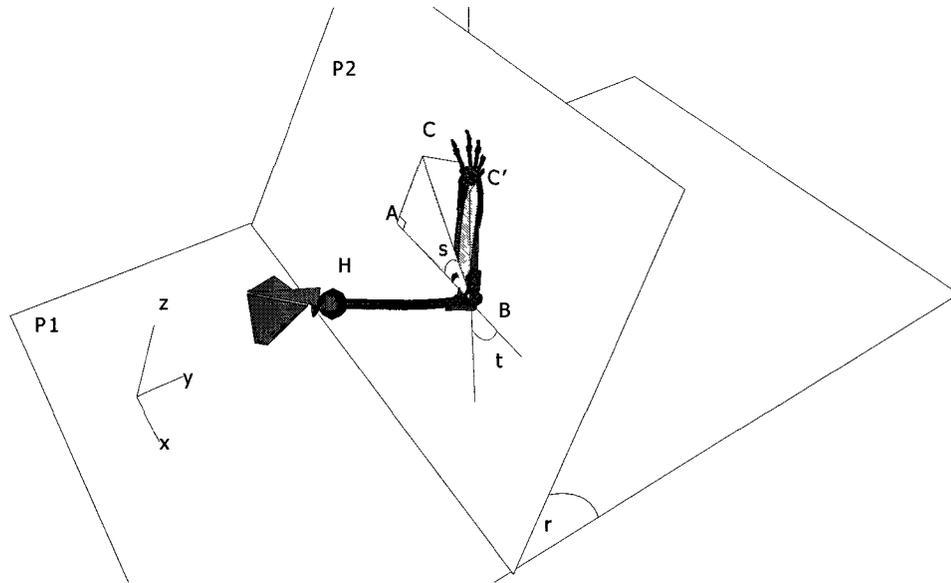


Figure 5.7: Shoulder roll.

The straight line equation that represents humerus (HB) can be expressed as:

$$\frac{x - x_{sh}}{x_{el} - x_{sh}} = \frac{y - y_{sh}}{y_{el} - y_{sh}} = \frac{z - z_{sh}}{z_{el} - z_{sh}} \quad (5.3)$$

where (x_{sh}, y_{sh}, z_{sh}) , (x_{el}, y_{el}, z_{el}) , and (x_{wr}, y_{wr}, z_{wr}) represent the coordinates of the

middle of the shoulder, the middle of the elbow and the middle of the wrist respectively.

The plane P2 which is perpendicular to the humerus at the elbow is:

$$a(x - x_{el}) + b(y - y_{el}) + c(z - z_{el}) = 0 \quad (5.4)$$

where:

$$\begin{aligned} a &= (x_{el} - x_{sh}) \\ b &= (y_{el} - y_{sh}) \\ c &= (z_{el} - z_{sh}) \end{aligned} \quad (5.5)$$

The projection of point C' on the plane P2 is the point C with the following coordinates:

$$\begin{aligned} x_C &= x_{wr} - a \frac{ax_{wr} + by_{wr} + cz_{wr} + d}{a^2 + b^2 + c^2} \\ y_C &= y_{wr} - b \frac{ax_{wr} + by_{wr} + cz_{wr} + d}{a^2 + b^2 + c^2} \\ z_C &= z_{wr} - c \frac{ax_{wr} + by_{wr} + cz_{wr} + d}{a^2 + b^2 + c^2} \end{aligned} \quad (5.6)$$

In order to calculate the angle s , we need to find the coordinates of the point A. Figure 5.8 shows the bones in a closer look. We pass a plane, P3, through the point C such that it is perpendicular to the planes P1 and P2. The length of the line segment AB is equal to the distance between the point B and the plane P3. In Figure 5.8 the vectors of the planes P1, P2, and P3, are denoted by v_1 , v_2 , and v_3 respectively. The plane P1 is parallel to the XY plane, therefore the vector v_1 is $z = 0$. The vector v_2 calculated from Equation 5.4, which can be expressed as vector (a, b, c) . Since plane P3 is perpendicular to both planes P1 and P2, the vector of the plane P3 can be calculated by the cross product of the vectors of the planes P1 and P2 using the following equation:

$$v_3 = v_1 \times v_2 = (0, 0, 1) \times (a, b, c) = (-b, a, 0) \quad (5.7)$$

Therefore, the equation of the plane P3 is

$$-b(x - x_C) + a(y - y_C) = 0 \quad (5.8)$$

The distance between the point B and the plane P3 is calculated as:

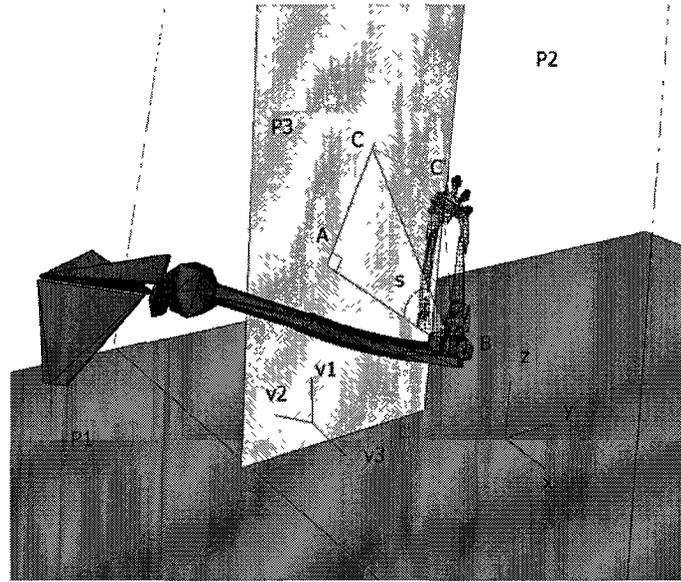


Figure 5.8: Shoulder roll.

$$d = \frac{|bx_B - ay_B + bx_C - ay_C|}{\sqrt{b^2 + a^2}} = \frac{|b(x_B + x_C) - a(y_B + y_C)|}{\sqrt{b^2 + a^2}} \quad (5.9)$$

Finally, the roll angle can be calculated as:

$$s = \arccos\left(\frac{AB}{BC}\right) = \arccos\left(\frac{d}{\sqrt{(x_B - x_C)^2 + (y_B - y_C)^2}}\right) \quad (5.10)$$

Figure 5.9 shows the estimated values for the rolling angle at the shoulder.

Table 5.5 shows the values calculated for the shoulder rolling. The angle of elbow was 90 degrees during the test.

5.2.4 Elbow pitch

As we explained in Chapter 4, the first joint that is located in our method is the elbow. Locating the elbow is the most important part in our method because, without finding the elbow, we cannot locate the other joints. As we discussed in Chapter 4, it is not possible to find the elbow in all the upper arm positions. Here we find the elbow pitch and study for what elbow angles our system can estimate the bone lengths and the other degrees of freedom.

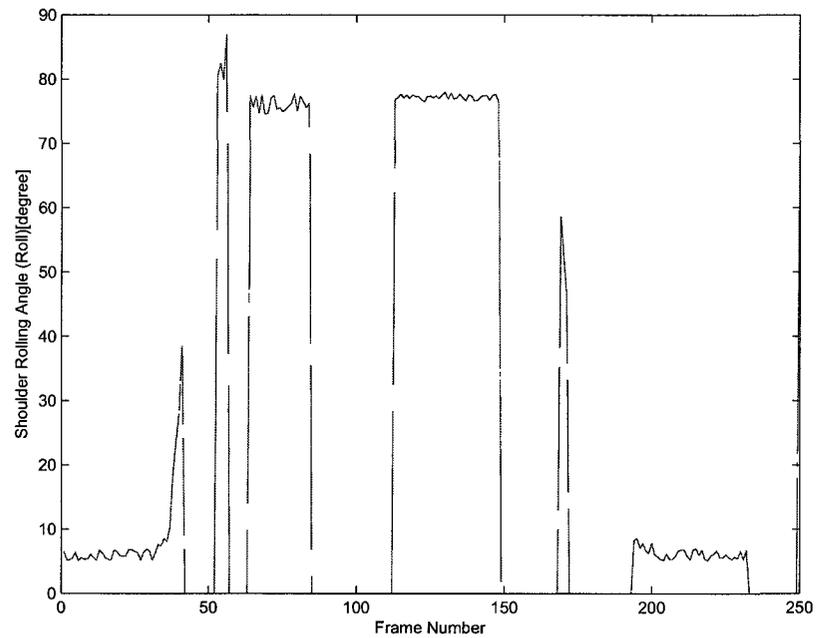


Figure 5.9: Estimated shoulder rolling angles [degrees].

Figure 5.10 shows the scheme of the upper limb for estimating the elbow angle.

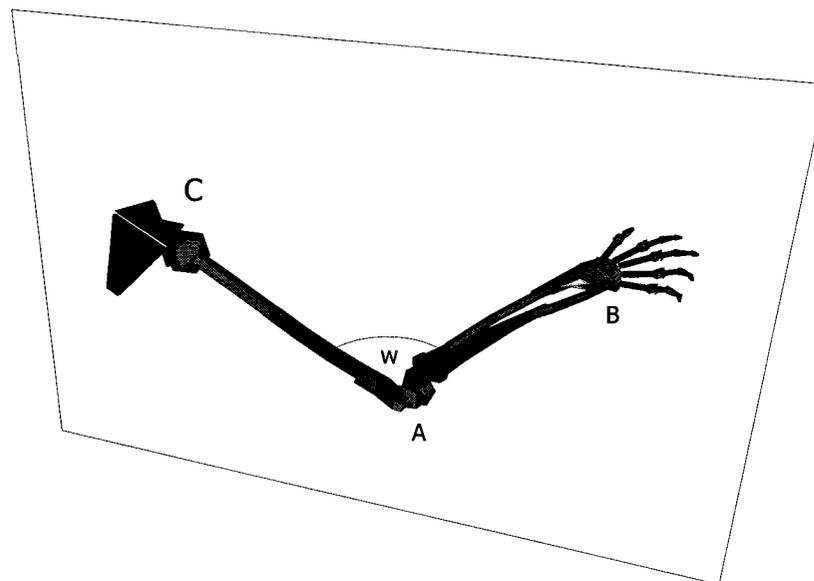


Figure 5.10: Elbow pitch.

Real	-10	0	10	20	30	40
Estimated	-12.545	3.897	11.621	18.432	28.026	40.388
Error	-2.545	3.897	1.621	1.568	1.974	0.388
Real	50	60	70	80	90	
Estimated	52.422	63.001	72.989	85.222	94.31	
Error	2.422	3.001	2.989	5.222	4.31	

Table 5.5: Shoulder roll estimated and real values [degrees].

The elbow angle can be calculated using the following equation.

$$w = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \quad (5.11)$$

where a, b and c are the Euclidean distances between the points A,B, and C. The values of a, b, and c are calculated using Equations 5.12.

$$\begin{aligned} a &= BC = \sqrt{(x_B - x_C)^2 + (y_B - y_C)^2 + (z_B - z_C)^2} \\ b &= AC = \sqrt{(x_A - x_C)^2 + (y_A - y_C)^2 + (z_A - z_C)^2} \\ c &= AB = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \end{aligned} \quad (5.12)$$

Tables 5.6 to 5.8 show the estimated angles at the elbow joint for various elbow pitch angles. It is observed that the elbow angle and shoulder pitch angle are almost not related to each other.

Shoulder pitch=0	Elbow Angle [degrees]						
Real	80	90	100	110	120	130	140
Estimated	74.639	86.942	104.589	113.869	124.365	134.951	145.325
Error	5.361	3.058	4.589	3.869	4.365	4.951	5.325

Table 5.6: The elbow angle [degrees] where the shoulder pitch is equal to 0 degree.

Shoulder pitch=30	Elbow Angle [degrees]						
Real	80	90	100	110	120	130	140
Estimated	82.828	92.067	104.762	113.727	124.894	134.551	145.437
Error	2.828	2.067	4.762	3.727	4.894	4.551	5.437

Table 5.7: The elbow angle [degrees] where the shoulder pitch is equal to 30 degrees.

Shoulder pitch =45	Elbow Angle						
Real	80	90	100	110	120	130	140
Estimated	76.84	91.648	101.01	110.683	123.191	132.371	144.332
Error	3.16	1.648	1.01	0.683	3.191	2.371	4.332

Table 5.8: The elbow angle [degrees] where the shoulder pitch is equal to 45 degrees.

Figure 5.11 shows the estimated elbow angle for each frame over the sequence of the subject movements described in Section 5.1.

As shown in Figure 5.11, the elbow pitch angle cannot be estimated for several series of the frames. The main reasons that our system is not able to calculate the pitch angle in these cases are as follows:

1) The system is unable to locate the elbow in two or more camera views. In our experiment the system is unable to locate the elbow when the elbow angle is not in the range of 70 to 160 degrees in two or more camera views.

2) The wrist or shoulder cannot be identified in the videos when one part of the arm or hand is occluding the other part in the camera views.

5.2.5 Wrist roll

The wrist roll angle is the rotation angle of the two forearm bones, in other words, the angle between the major axis of the ellipses at the wrist and the elbow. Figure 5.12 shows the angle of the wrist when the elbow is located horizontally.

For finding this angle, we calculate two vectors V_{EL} and V_{WR} from Equation 5.13. The vector V_{EL} is the vector of the major axis of the estimated ellipse for the elbow

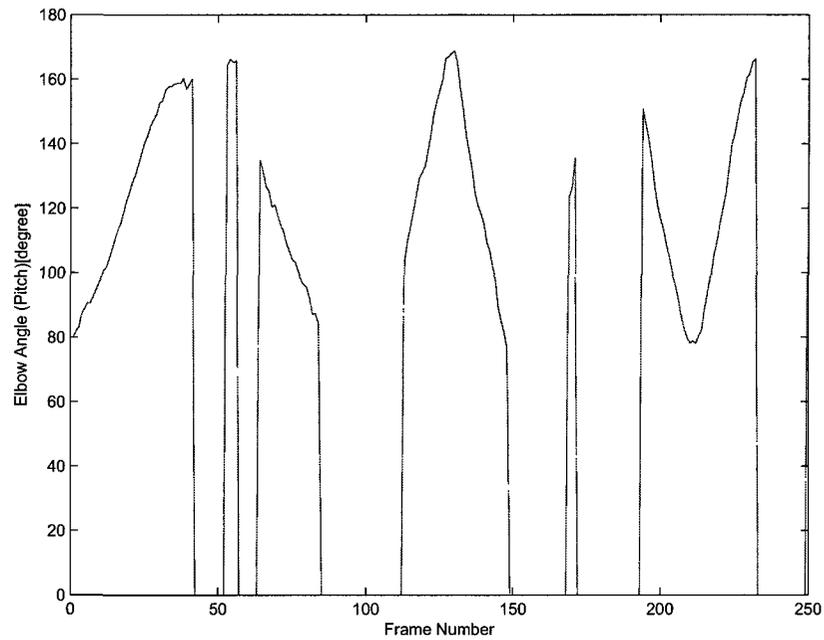


Figure 5.11: Estimated elbow pitch angles [degrees].

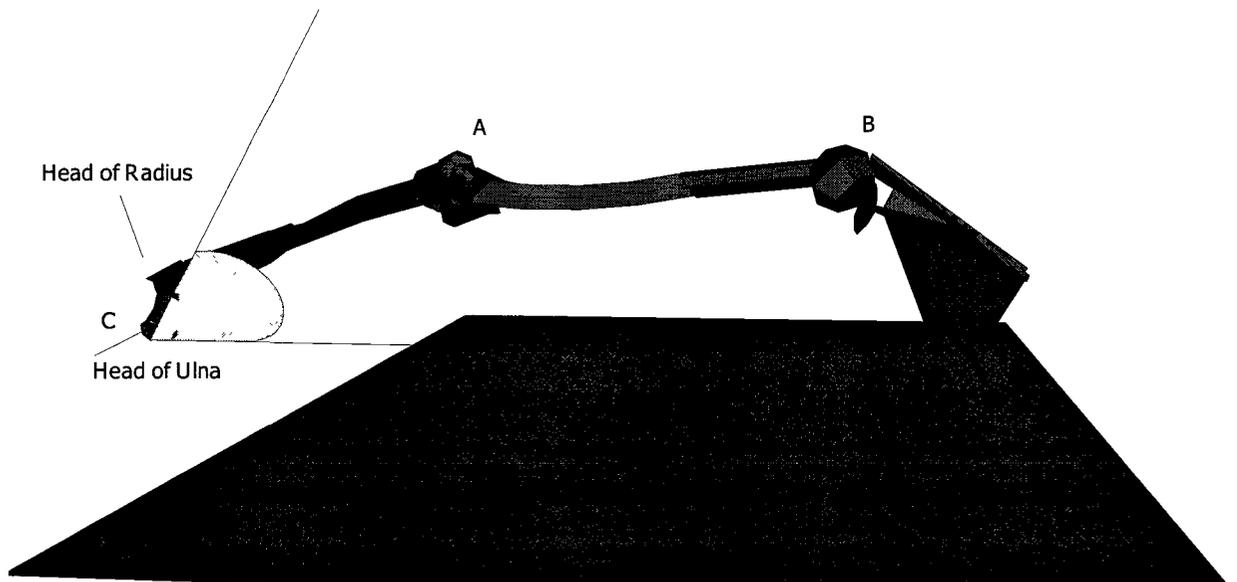


Figure 5.12: Wrist roll.

while the vector V_{WR} is the vector of the major axis of the estimated ellipse for the wrist. As shown in Figure 5.13, the points $EL_0 : (x_0, y_0, z_0)$ and $EL_1 : (x_1, y_1, z_1)$ are the foci of the estimated wrist ellipse. Similarly the points $WR_0 : (x_2, y_2, z_2)$ and

$WR_1 : (x_3, y_3, z_3)$ are the foci of the estimated elbow ellipse. The wrist roll angle is denoted by g .

$$\begin{aligned} V_{EL} &= (x_1 - x_0, y_1 - y_0, z_1 - z_0) \\ V_{WR} &= (x_3 - x_2, y_3 - y_2, z_3 - z_2) \\ g &= \arccos \left(\frac{V_{EL} \cdot V_{WR}}{|V_{EL}| |V_{WR}|} \right) \end{aligned} \quad (5.13)$$

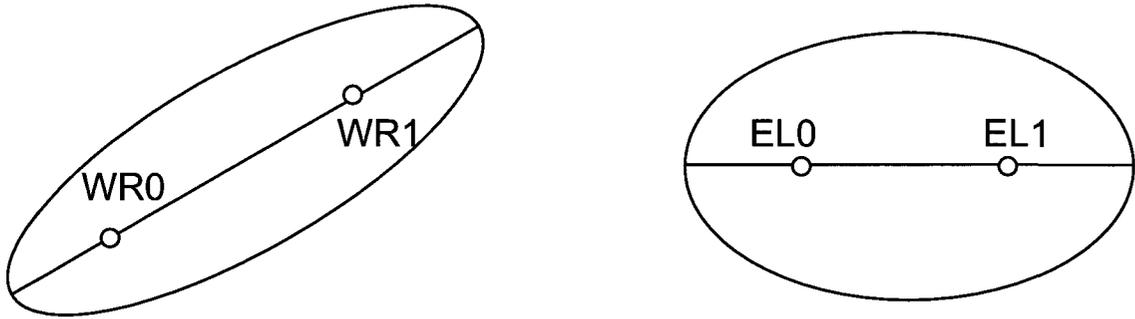


Figure 5.13: Wrist and elbow estimated ellipses.

Figure 5.14 shows the estimated wrist roll angles for the captured videos.

Table 5.9 shows the actual and estimated results for wrist roll. It can be seen that the system was not able to find the angle with a good precision. In our system we rely on the two curves on two sides of the wrist, which is difficult to track especially in low resolution images. Using higher resolution cameras can improve the accuracy of the system.

Real	-30	0	30	60	90
Estimated	-28.51	5.87	25.44	53.5	81.66
Error	1.49	5.87	4.56	7.5	8.34

Table 5.9: The actual and estimated results for wrist angles [degrees].

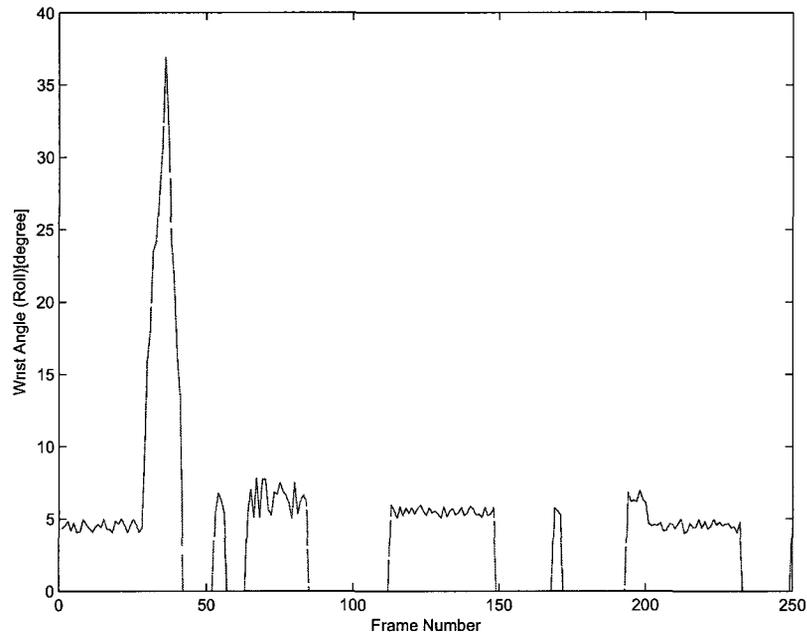


Figure 5.14: Estimated wrist roll angles [degrees].

5.3 Bone lengths

The humerus, ulna, and radius estimated lengths for various elbow angles and the estimation error for each one are presented in this section. In Chapter 4 we showed how our system estimates the end of the bones. By finding the Euclidean distance between the bone ends, the length of the bones is estimated.

5.3.1 Humerus

The actual length of the humerus was measured between the epicondyle and head of the humerus using a measuring tape by a medical doctor. For estimating the humerus bone length, we calculated the Euclidean distance between the ends of humerus in the shoulder and in the elbow.

Figure 5.15 shows the estimated length for humerus throughout the video sequence. The horizontal axis shows the corresponding frame numbers.

Figure 5.16 illustrates the estimation error for the length of the humerus bone. The error is defined by Equation 5.14.

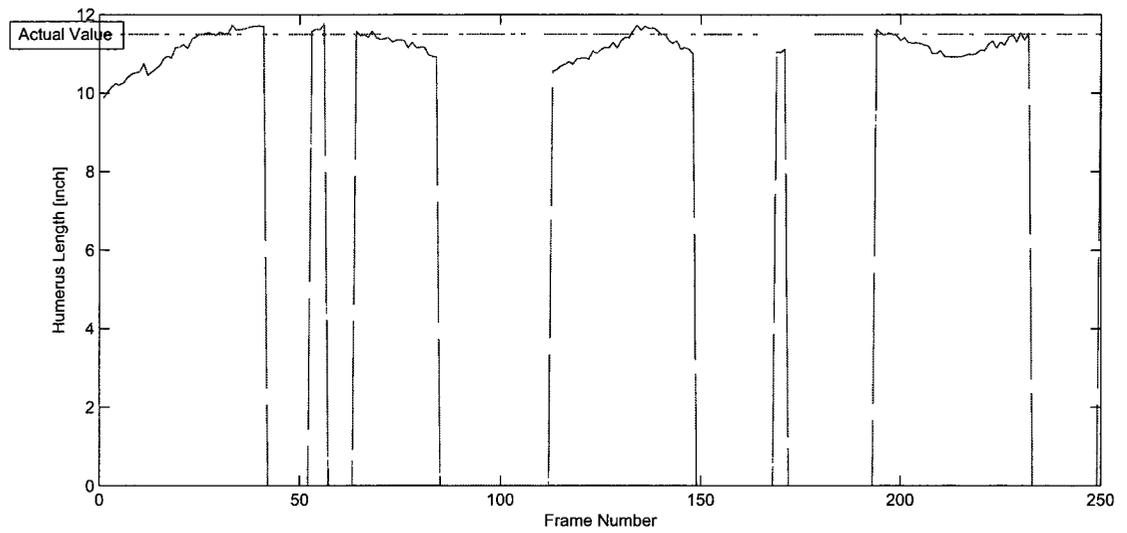


Figure 5.15: Estimated length of humerus bone [inches].

$$Error = \frac{|L_{estimated} - L_{real}|}{L_{real}} \times 100 \quad (5.14)$$

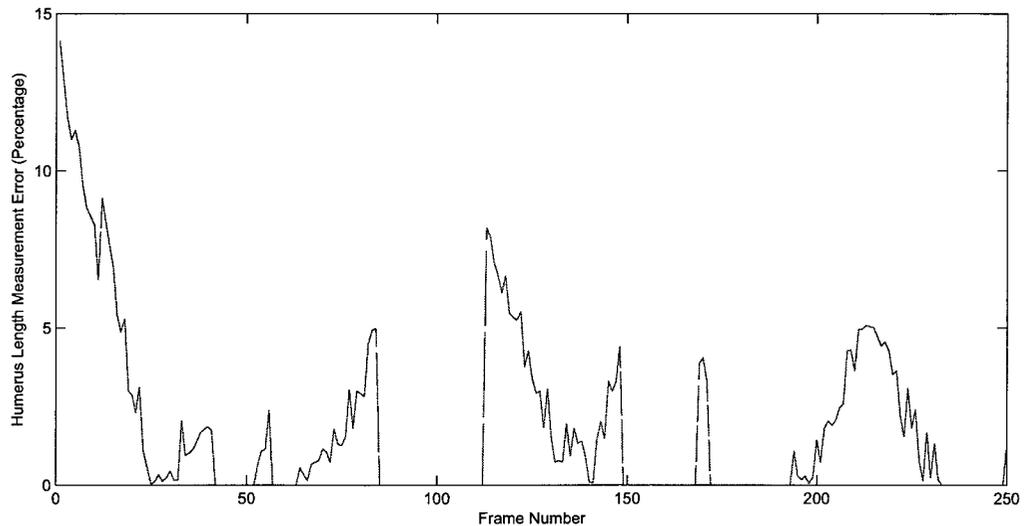


Figure 5.16: Absolute error for the estimated humerus bone lengths.

From the results shown in Figure 5.15 and Figure 5.16 we can conclude:

- Apart from the beginning frames, the length of the humerus bone was estimated

with more than 90 percent accuracy.

- By studying Figures 5.11 and 5.15 we conclude that for the angles close to 80 degrees, the length of the humerus was estimated with the minimum accuracy.
- For the angles around 160 degrees, the humerus bone was estimated with more than 97 percent accuracy.

5.3.2 Radius

In anatomy science [50], the length of the radius is measured between the tip of the styloid process and the articular surface at the top of radius head. According to Table 5.1, the real length of the radius for the candidate is 10.5 inches. For estimating the radius bone length, we calculated the Euclidean distance between the ends of radius in the wrist and in the elbow. The method for locating the end of the bones in the estimated ellipses for the joints was explained in Chapter 4. Figure 5.17 shows the estimated length for the radius bone over the video sequence.

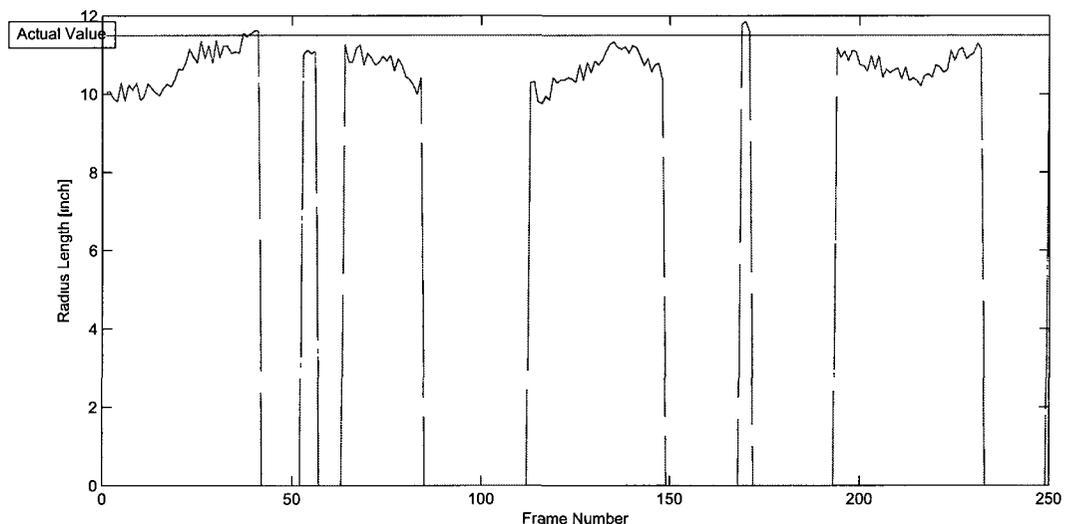


Figure 5.17: Estimated length of radius bone.

The estimation error for the radius bone length is shown in Figure 5.18.

From the results shown in Figure 5.17 and Figure 5.18 it can be concluded that in most upper limb positions, the length of the radius bone is estimated with more

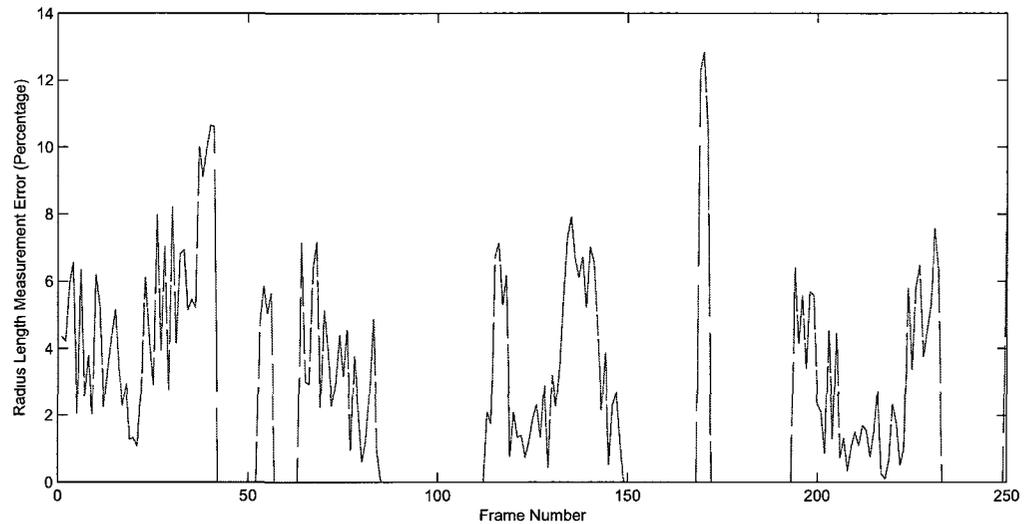


Figure 5.18: Absolute error for the estimated radius bone lengths.

than 90 percent accuracy.

5.3.3 Ulna

The length of ulna is measured from the tip of the styloid process at the lower end to superior surface of the olecranon process. The ulna length for the candidate was 11.3 inches. For estimating the ulna bone ends similar to the previous bones we calculated the Euclidean distance between the end of the bones located in the elbow and wrist cross sections. Figure 5.19 shows the estimated length for ulna.

The estimation error for the ulna bone length is shown in Figure 5.20. According to the results in Figures 5.19 and 5.20, our system reaches more than 88 percent accuracy for estimating the ulna bone length. As it was stated in Chapter 4, it is difficult to find the end of the ulna bone at the wrist area because it is located in a flat position. This causes the fluctuations in the estimated radius bone length.

5.3.4 Average error

Table 5.10 shows the average estimated length for each bone. The average values are calculated over all the values that the system estimated. In this calculation, we opted

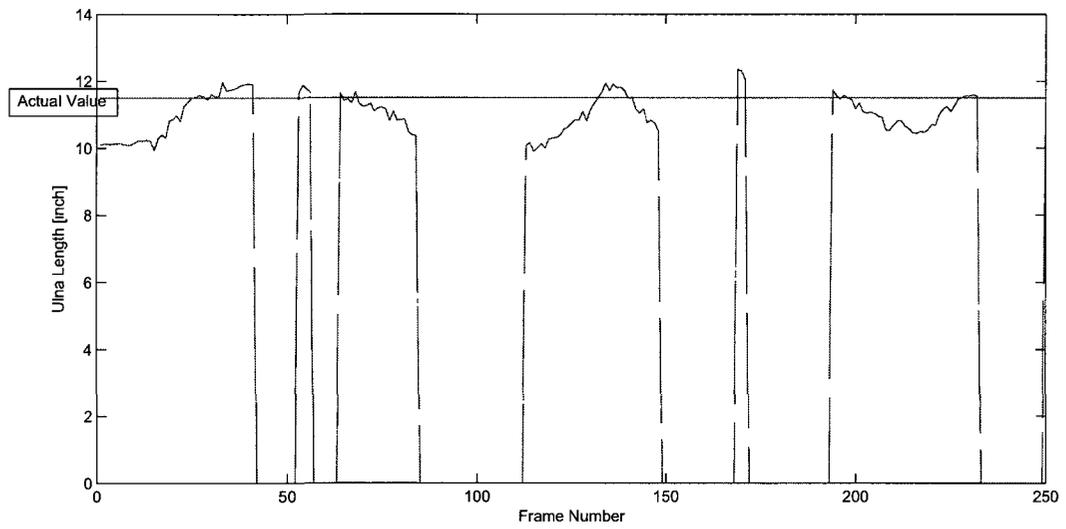


Figure 5.19: Estimated length of ulna bone.

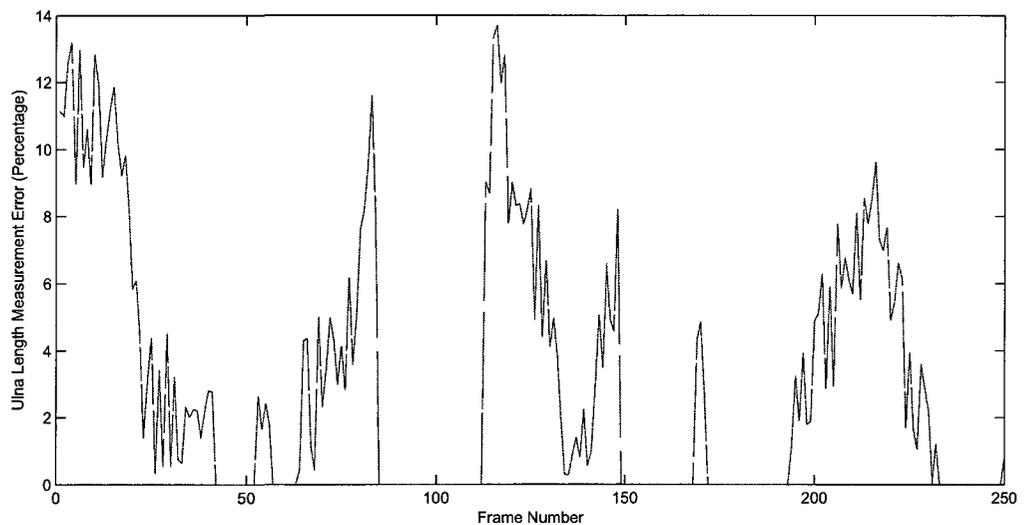


Figure 5.20: Estimation error for the ulna bone length.

out the states where the system was unable to locate the bone ends and estimate the length of the bones which are shown as gaps in Figures 5.15, 5.17 and 5.19.

By comparing the measured values of length for the bones in Table 5.1 and the estimated values for the bones in Table 5.10 it can be concluded that the system has estimated the bone length within a half inch accuracy over the movement sequence.

	Ulna	Radius	Humerus
Mean	11.039	10.736	11.192
Variance	0.362	0.210	0.1515

Table 5.10: The calculated bones lengths [inches].

5.3.5 Discussion

In Figures 5.15-5.20 a number of gaps exist in the results. These gaps belong to the frames for which the system was unable to determine one of the bone ends. It means that the system marked the results of belonging to these ranges as invalid. In total, the angles for five DoF were calculated in this thesis. Among them, the wrist roll was estimated with the least accuracy because of the small size of the wrist and the fact that calculating the wrist roll relies on two ellipse estimations for the wrist and elbow areas, which cause the errors in calculations to add up.

The humerus is located between the elbow joint and the shoulder. Its end point on the elbow is located on the minor axis of the estimated ellipse at the elbow, near the end of the ulna. The proposed system was able to estimate the humerus length more precisely than the radius length and the ulna length. This is due to the fact that the shoulder does not move very much in comparison with the other bones. In the forearm, the length of the radius is estimated with less error. In fact one head of radius is presumably located in the middle of the elbow which can be found easily for different figures of the elbow when the candidate bends his elbow. The other end of the radius is at the wrist which is trackable because of its position in the arc under the thumb.

Unlike the radius, the ulna is located under the little finger in a relatively flat position that sometimes is less observable. As a result, the variance on the ulna length estimations is about 1.5 times the variance of the measured radius length.

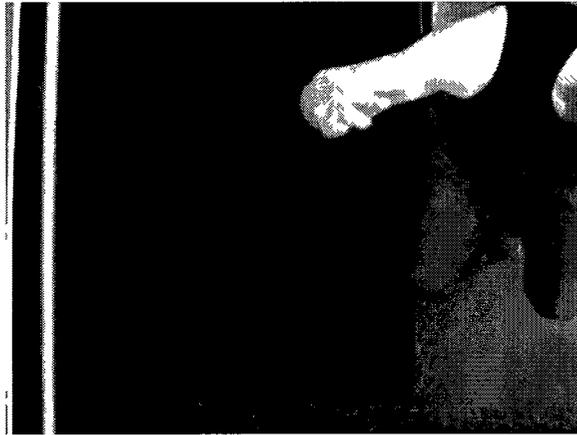
As discussed in Chapter 4, for some frames, the outputs of the system were invalid, mainly because of the effect of shadow or improper lighting. In several frames, especially when the elbow angle is less than 80 degrees, the forearm occludes the upper arm in some camera views and the detected points at the elbow joint are less accurate. This causes the system to be unable to estimate the ellipse at the elbow. A similar scenario happens for the wrist, as the fingers may occlude the wrist in a

particular camera view. Figure 5.21 illustrates two examples of these cases.

As stated in Chapter 4, the calibration system has a relatively high accuracy with an error of less than 0.5 pixel. Therefore the main source of error of the system is not originating from calibration errors.

Figures 5.22 to 5.25 show the camera views for two frames as the input, and the upper limb 3D illustration of arm bones as the output of our system. This figure illustrates the estimated ellipses and lines for articulations and bones, respectively.

In an interview with a medical doctor, he declared that almost 10 degrees of error for the joint range of motion assessments is acceptable for musculoskeletal examinations. As a result, the accuracy of our system is suitable for practical medical applications. For the bone length measurement, no specific accuracy were found in textbooks. The medical doctor recommended less than half of an inch of error for the bone length which is equal to an accuracy of 95% or higher. Therefore, the desired accuracy is close to the accuracy of the system.



(a)



(b)

Figure 5.21: Arm positions for which the system does not respond properly: (a) forearm occludes the elbow and upper arm , (b) the elbow and the wrist are not distinguishable from the lateral view.

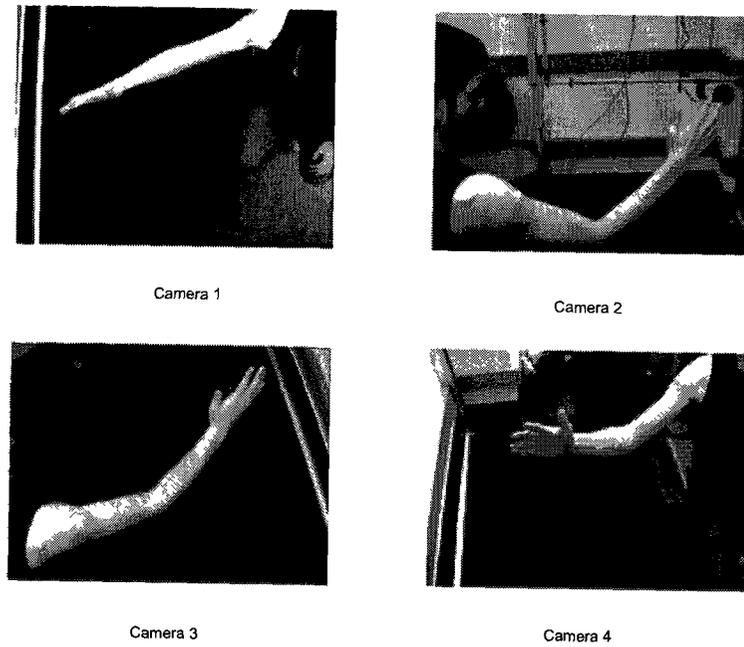


Figure 5.22: Camera views (case I).

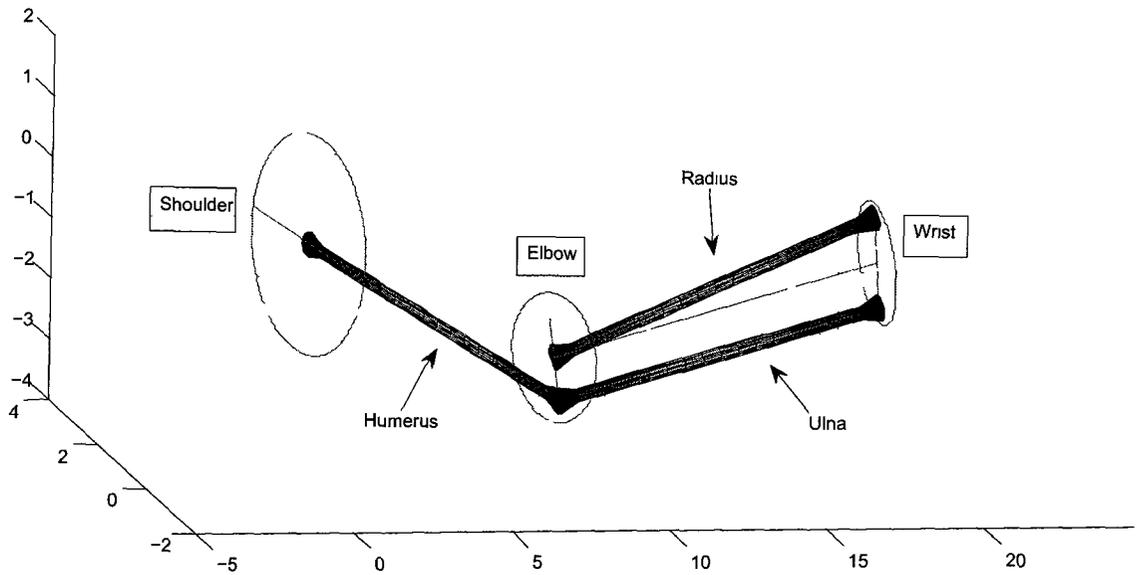


Figure 5.23: Upper limb 3D illustration of arm bones for case I, seen from the right side.

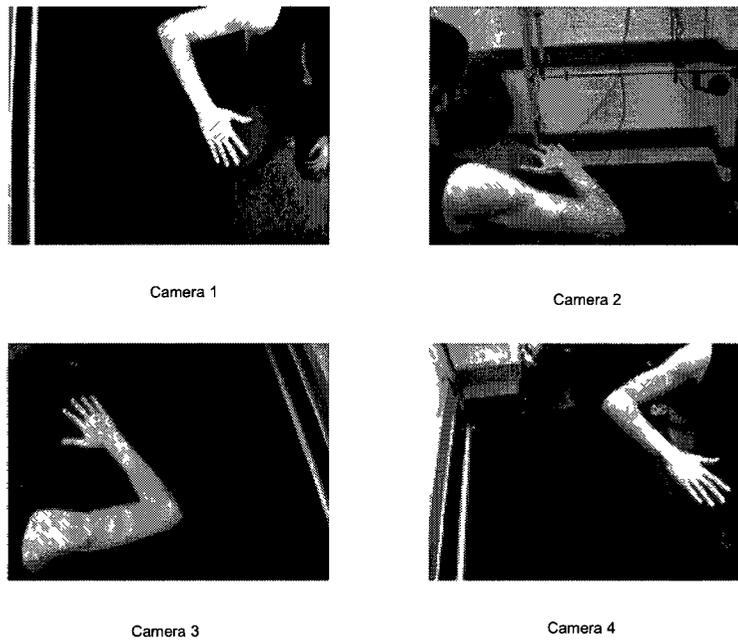


Figure 5.24: Camera views (case II).

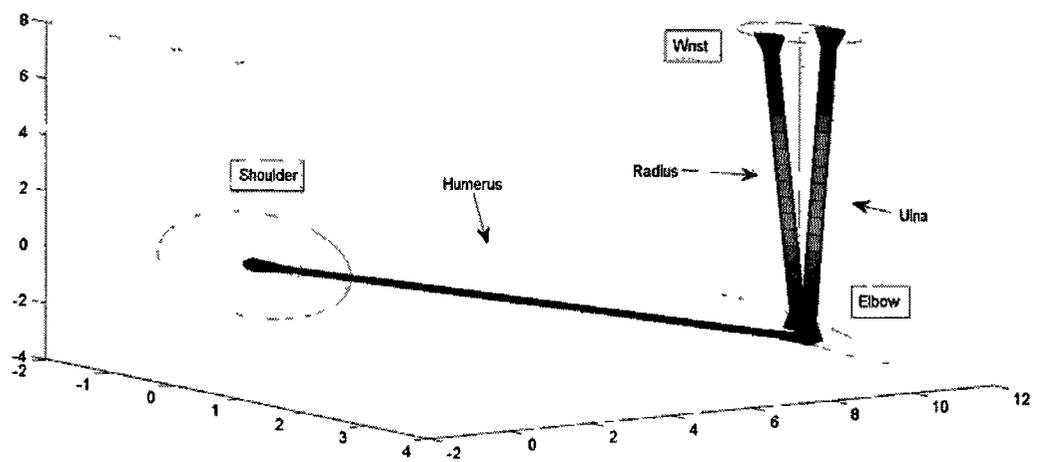


Figure 5.25: Upper limb 3D illustration of arm bones for case II, seen from above.

Chapter 6

Conclusions and Contributions

The goal of this work was the development of a computer vision system for musculoskeletal disorders examination in clinical settings. The system was required to be easy to use and the output of the system to be shown in 3D, while providing quantitative measurements of all human upper limb joint angles and bone lengths. To achieve this goal, we chose a markerless motion capture approach which would be more convenient for the doctors and patients in comparison with marker-based systems which have the hassle of putting on special suits with the markers attached on them. With the proposed system, a patient just stands between the cameras and starts to move his or her arm following a predefined pattern of motion. Since there is no need for attaching any markers on the patient's body, the whole process is fast. It should be noted that there is no need to change the settings of our system for each patient.

6.1 Summary

This work introduced a novel markerless motion capture approach system for quantitative analysis of the human upper limb. The proposed markerless system provides a number of benefits over marker-based systems. Since it only uses passive vision for monitoring, the hassle of putting on the suits with the markers on them is avoided. The advantage of the proposed method is that it uses an innovative technique for estimating the joints angles and length of the bones. The system finds and tracks the rolling movements in the wrist without any need for tracking the rotation of the hand or the fingers.

This research work provided a computer vision system for analysis of arm images.

The system is able to find the location of the joints and estimate the angles for five degrees of freedom. Moreover it is able to find the length of the bones. The system utilized a multi-camera setup to cover the arm from various angles. The video streams from the cameras are analyzed separately before being merged. The output is the angles of the five degrees of freedom at the joints and the length of the bones in the upper arm and forearm. The system manages to find the bone ends at the wrist; as a result it is able to find the rotation of the wrist without tracking the movement of the hand. By finding the joints in 2D, only those points are used for 3D conversion instead of converting the complete body into a 3D representation.

A major potential application for this system is for musculoskeletal examination in clinics. The system accuracy can be improved by using more cameras. Also using higher resolution cameras would give the system higher accuracy. Another application for this system is in sports such as tennis, where the wrong style for moving the hand can cause long-term injuries.

6.2 Limitations

While designing the proposed system, efforts were made to eliminate the effect of skin color by applying a suitable skin filter. The proposed system has been validated using the samples of one subject. For a better validation of the system, it should be tested using the samples taken from several subjects with different skin colors or different body types.

Another limitation of the system is the reliability of the results. For example, the results for the bone length measurements can vary up to 10% which, is not suitable for precise bone length measurement. This aspect could be improved by using more cameras. An intelligence system, such as a neural network based system, can be trained to choose the best camera views and generate the most accurate results.

6.3 Contributions

This thesis provided two major contributions to the field of markerless motion capture applications.

- 1) A new method that shows promise. The proposed method is based on the

combination of the skeleton and contour analysis to extract all the usable information from relatively low resolution images. Relying on the skeletonization method for finding the skeleton of the body can be performed in 2D and 3D. Because of the low resolution input images, it was impossible to achieve good results by 3D reconstruction. In addition, 3D systems need a larger number of cameras in order to produce good resolution outputs which are needed for modeling purposes and skeleton extraction. The results of our work support the development of motion capture systems using low resolution input and a limited number of cameras.

2) A model to track the upper limb movements without modeling the muscles. Our model is based on anatomical knowledge of the joints shape and structure which is less complex than the methods which model the muscle movements. The proposed model is a joint-based model which is not dependent on the muscle relaxation and contraction. Therefore its development and processing load is less complicated.

6.4 Future work

In the following section, a number of opportunities are outlined in order to extend this Master's thesis work.

- **Co-working with other systems:**

As mentioned before, the proposed method does not rely on the shape of the hand for finding the wrist. In other words, it does not need to track the movements of the hand or finger locations in order to detect the rotation of the wrist. As a result, the proposed method can be combined with any method that works on the fingers and the results of our scheme are independent from the results of that method which can serve in a collaborative manner.

- **Lower limb structure modeling:**

Considering a tree structure for the human body, the proposed approach can be applied for analyzing every branch of that tree, for example to lower limb. In fact using the structure information of the body, a global system can divide the body into several branches and analyze every branch independently. In each branch, by locating the joints and matching the ellipse or other proper geometrical shapes, it can find the exact position and the state of the joints and

reconstruct the skeleton in 3D. By combining the results, the complete human skeleton can be determined from the images.

- **Muscle relaxation and contraction modeling:**

One of the interesting research areas might be applying our system to perform muscle relaxation and contraction modeling. One approach is dividing the length of the muscle in several slices. Each slice can be treated as a bone and our proposed method can be applied for them. Corresponding fitted ellipses show the estimated shape of the muscle at each section. Dividing the muscle in a larger number of sections gives a better model of contraction and relaxation.

- **Virtual reality applications:**

While a major application of human motion capture is in computer character animation or virtual reality applications, the proposed approach can be extended to these applications in video games and movie making industries. The position information gathered by our system are sent to the computers. Finally, this information is used to map the movements of a human actor to an animated character.

List of References

- [1] S. Bériault, P. Payeur, and G. Comeau, “Flexible multi-camera network calibration for human gesture monitoring,” in *IEEE International Workshop on Robotic and Sensors Environments*, pp. 1–6, 2007.
- [2] F. J. Perales, “Human motion analysis & synthesis using computer vision and graphics techniques. state of art and applications.,” in *Proceedings of the world multiconference on systemics, cybernetics and informatics (SCI2001)*, 2001.
- [3] L. Lucchetti, A. Cappozzo, A. Cappello, and U. D. Croce, “Skin movement artefact assessment and compensation in the estimation of knee-joint kinematics,” *Journal of Biomechanics*, vol. 31, no. 11, pp. 977–984, 1998.
- [4] S. Yabukami, H. Kikuchi, M. Yamaguchi, K. I. Arai, K. Takahaschi, A. Itagaki, and N. Wako, “Motion capture system of magnetic markers using three-axial magnetic field sensor,” *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 3646–3648, 2000.
- [5] S. Hashi, M. Toyoda, S. Yabukami, K. Ishiyama, Y. Okazaki, and K. I. Arai, “Wireless magnetic motion capture system - compensatory tracking of positional error caused by mutual inductance -,” *IEEE Transactions on Magnetism*, vol. 43, no. 6, pp. 2364–2366, 2007.
- [6] “Vicon motion capture system,” www.vicon.com, 2006.
- [7] B. Holmberg and H. Lanshammar, “Possibilities in using skin texture based image registration for human movement,” in *Proc., Ninth International Symposium on the 3-D Analysis of Human Movement*, 2006.
- [8] R. B. Davis, S. Ounpuu, D. Tyburski, and J. R. Gage, “A gait analysis data collection and reduction technique,” *Human Movement Science*, vol. 10, no. 5, pp. 575–587, 1991.
- [9] G. K. Seidel, D. M. Marchinda, M. Dijkers, and R. W. Soutas-Little, “Hip joint center location from palpable bony landmarks—a cadaver study,” *Journal of Biomechanics*, vol. 28, no. 8, pp. 995–998, 1995.
- [10] M. H. Schwartz and A. Rozumalski, “A new method for estimating joint parameters from motion data,” *Journal of Biomechanics*, vol. 38, no. 1, pp. 107–116, 2005.

- [11] K. Halvorsen, M. Lesser, and A. Lundberg, "A new method for estimating the axis of rotation and the center of rotation," *Journal of biomechanics*, vol. 32 issue 11, pp. 1221–1227, 1 November 1999.
- [12] A. Marshall and B. Shaw, "Modeling the survival of hip fracture patients using a conditional phase-type distribution," in *CBMS '08: Proceedings of the 2008 21st IEEE International Symposium on Computer-Based Medical Systems*, (Washington, DC, USA), pp. 518–523, IEEE Computer Society, 2008.
- [13] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.
- [14] A. Leardini, A. Cappozzo, F. Catani, S. Toksvig-Larsen, A. Petitto, V. Sforza, G. Cassanelli, and S. Giannini, "Validation of a functional method for the estimation of hip joint centre location," *Journal of Biomechanics*, vol. 32, no. 1, pp. 99–103, 1999.
- [15] A. L. Bell, D. R. Pedersen, and R. A. Brand, "A comparison of the accuracy of several hip center location prediction methods," *Journal of Biomechanics*, vol. 23, no. 6, pp. 617–621, 1990.
- [16] A. Cappozzo, F. Catani, U. D. Croce, and A. Leardini, "Position and orientation in space of bones during movement: Experimental artefacts," *Clinical Biomechanics*, vol. 11, no. 2, pp. 90–100, 1996.
- [17] M. Descarreaux, J.-S. Blouin, and N. Teasdale, "A non-invasive technique for measurement of cervical vertebral angle: Report of a preliminary study," *European Spine Journal*, vol. 12, pp. 314–319, 2003.
- [18] Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with silhouettes," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 716–721 vol.2, 1999.
- [19] K. M. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time part 2: Applications to human modeling and markerless motion tracking," *International Journal of Computer Vision*, vol. 63, no. 3, pp. 225–245, 2005.
- [20] P. D. Leva, "Joint center longitudinal positions computed from a selected subset of chandlers data," *Journal of Biomechanics*, vol. 29, no. 9, pp. 1231–1233, 1996.
- [21] L. Mundermann, S. Corazzo, and T. Andriacchi, "Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–6, 2007.
- [22] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, Jan. 1995.
- [23] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, "Articulated body posture estimation from multi-camera voxel data," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, (Kauai, Hawaii), p. 455, Dec. 2001.

- [24] T. Krosshaug and R. Bahr, “A model-based image-matching technique for three dimensional reconstruction of human motion from uncalibrated video sequences,” *Journal of Biomechanics*, vol. 38, no. 4, pp. 919–929, 2005.
- [25] J. K. Aggarwal and Q. Cai, “Human motion analysis: A review,” *Computer Vision and Image Understanding*, vol. 73, pp. 428–440, 1999.
- [26] V. Ferrari, M. Marín-Jiménez, and A. Zisserman, “2D human pose estimation in TV shows,” in *Dagstuhl post-proceedings*, (Berlin, Heidelberg), pp. 128–147, Springer-Verlag, 2009.
- [27] C. Bregler, J. Malik, and K. Pullen, “Twist based acquisition and tracking of animal and human kinematics,” *International Journal of Computer Vision*, vol. 56, no. 3, pp. 179–194, 2004.
- [28] I. A. Kakadiaris and D. N. Metaxas, “Model-based estimation of 3D human motion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1453–1459, 2000.
- [29] L. Sigal, S. Bhatia, S. Roth, M. Black, and M. Isard, “Tracking loose-limbed people,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-421 – I-428 Vol.1, June 2004.
- [30] J. O’Rourke and N. I. Badler, “Model-based image analysis of human motion using constraint propagation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 522–536, 1980.
- [31] H. Sidenbladh, M. J. Black, and D. J. Fleet, “Stochastic tracking of 3D human figures using 2D image motion,” in *Proc., European Conference on Computer Vision (ECCV)*, (Dublin, Ireland), pp. 702–718, June 2000.
- [32] C. Barron and I. A. Kakadiaris, “Estimating anthropometry and pose from a single uncalibrated image,” *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 269–284, 2001.
- [33] R. Plankers and P. Fua, “Tracking and modeling people in video sequences,” *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 285–302, 2001.
- [34] A. Agarwal and B. Triggs, “Recovering 3D human pose from monocular images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 44–58, 2006.
- [35] H. Delingette, “General object reconstruction based on simplex meshes,” *International Journal of Computer Vision*, vol. 32, no. 2, pp. 111–146, 1999.
- [36] A. Kelemen, G. Székely, and G. Gerig, “Elastic model-based segmentation of 3-d neuroradiological data sets,” *IEEE Trans. Medical Imaging*, vol. 18, pp. 828–839, Oct. 1999.
- [37] A. Bottino and A. Laurentini, “A silhouette-based technique for the reconstruction of human movement,” *Computer Vision and Image Understanding*, vol. 83, no. 1, pp. 79–95, 2001.

- [38] G. K. Cheung, S. Baker, and T. Kanade, "Shape-from silhouette of articulated objects and its use for human body kinematics estimation and motion capture," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, (Madison, WI), pp. 77–84, June 2003.
- [39] R. Ogniewicz and M. Ilg, "Voronoi skeletons: Theory and applications," in *Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 63–69, 1992.
- [40] H. Blum, "A transformation for extracting new descriptors of shape," in *Models for the Perception of Speech and Visual Form* (W. Wathen-Dunn, ed.), pp. 362–380, Cambridge: MIT Press, 1967.
- [41] G. Borgefors, "Distance transformations in arbitrary dimensions," *Computer Vision, Graphics, and Image Processing*, vol. 27, no. 3, pp. 321–345, 1984.
- [42] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780–785, July 1997.
- [43] D. Ramanan and D. Forsyth, "Finding and tracking people from the bottom up," vol. 2, pp. II–467 – II–474 vol.2, jun. 2003.
- [44] T. J. Roberts, S. J. McKenna, and I. W. Ricketts, "Human tracking using 3D surface colour distributions," *Image and Vision Computing*, vol. 24, no. 12, pp. 1332–1342, 2006.
- [45] J. Deutscher and I. Reid, "Articulated body motion capture by stochastic search," *Int. J. Comput. Vision*, vol. 61, no. 2, pp. 185–205, 2005.
- [46] K. Grauman, G. Shakhnarovich, and T. Darrell, "Inferring 3D structure with a statistical image-based shape model," in *Proceedings of Ninth IEEE International Conference on Computer Vision, 2003.*, pp. 641–647 vol.1, 2003.
- [47] C. Taylor, "Reconstruction of articulated objects from point correspondences in a single uncalibrated image," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 677–684 vol.1, 2000.
- [48] F. Caillette, *Real-Time Markerless 3-D Human Body Tracking*. PhD thesis, University of Manchester, 2005.
- [49] Wikipedia, www.wikipedia.com.
- [50] C. M. Hall and L. T. Brody, *Therapeutic exercise: Moving toward function*. Lippincott Williams & Wilkins, 2005.
- [51] *e-MRI, Hoa D.*, www.imaios.com.
- [52] A. Tozeren, *Human body dynamics : classical mechanics and human movement*. Springer, 1999.
- [53] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, November 2000.

- [54] P. KadewTraKuPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *2nd European Workshop on Advanced Video-Based Surveillance Systems*, 2001.
- [55] M. Jones, M. J. Jones, and J. M. Rehg, "Statistical color models with application to skin detection," in *International Journal of Computer Vision*, pp. 274–280, 1999.
- [56] R. Gonzalez and R. Woods, *Digital image processing*. Reading, Massachusetts: Addison Wesley Publishing Company, 1992.
- [57] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer, "Comparison of edge detectors: A methodology and initial study," *Comput. Vis. Image Underst.*, vol. 69, no. 1, pp. 38–54, 1998.
- [58] K. Homma and E.-i. Takenaka, "An image processing method for feature extraction of space-occupying lesions," *J Nucl Med*, vol. 26, no. 12, pp. 1472–1477, 1985.

Appendix A

Application specifications

The application was written in Visual studio 2008. The OpenCV library was used for the image processing. In our software, the input video format is standard AVI with a 640x480 resolution and 15frames/sec. The computers were both Pentium IV 3.4GHz with 3 IEEE 1394b Firewire ports each. Four Point Grey Research Flea2 IEEE 1394b Firewire cameras were used for video capturing. The main files of our software are shown in Table A.1.

Figure A.1 shows the application main window. The user interface is very easy to use. The operator can set the number of cameras to be involved in the calculations. The application can process a single frame, the frames within a specific time range, or all of the frames. The application was developed using Visual C++ 2008. We also utilized OpenCV internal functions such as Canny filter in developing our code. The skin filter was implemented according proposed by We developed our own code for background removal and other parts of preprocessing algorithm. In addition, our code includes the two dimensional joints location estimation algorithm, three dimensional

Skin.cpp	Skin detection filter
skeleton.cpp	Skeletonization algorithms
fAnalysis.cpp	Frame extraction and analysis
ContourProcess.cpp	All the processes related to the contours and convex hull
ArmParts.cpp	Articulations and bones detection methods
3Dconversion.cpp	Converting from 3D to 2D and vice versa

Table A.1: Main parts of the application.

Processing, and range of motion estimation procedures which were entirely developed by us. A sample code from our application is shown in figure A.2

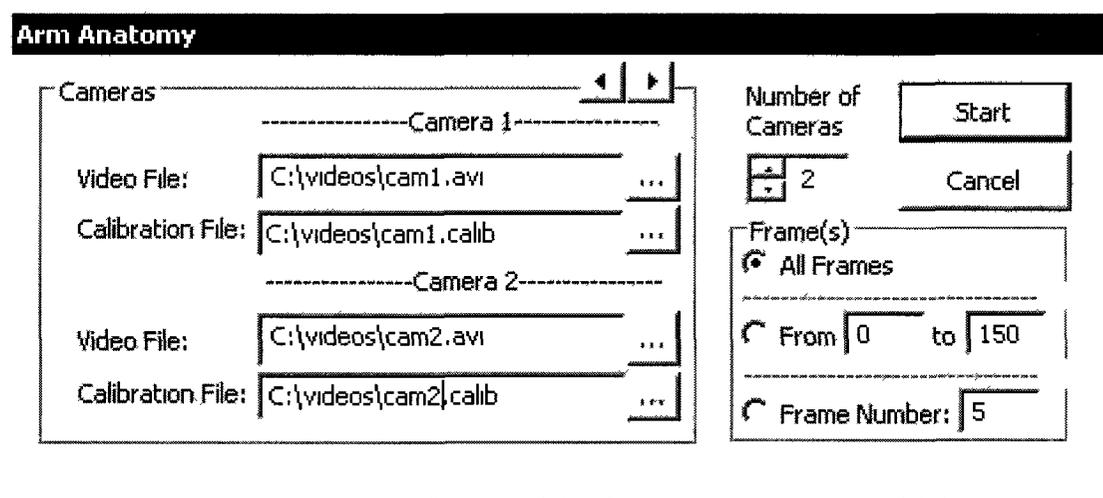


Figure A.1: Application main window.

```

int findjoints(CvSeq* skel,int p,CvSize size1){
//finds closest joint point to the point index: p

    IplImage* image = cvCreateImage( size1, 8, 1 );
    IplImage* imagep = cvCreateImage( size1, 32, 1 );
    cvDrawContours( image, skel, cvScalarAll(255), cvScalarAll(255), 0, 1, 8 );
    cvThreshold( image, image, 254, 255, 0 );
    CvPoint jointPointR,A;
    A= *((CvPoint*)cvGetSeqElem ( skel, p ));
    int s0,d,s;
    float cdata[9] = {-.1,-.1,-.1,
        -.1, 1,-.1,
        -.1,-.1,-.1} ;
    CvMat* pKernel = cvCreateMat(3,3,CV_32F);
    pKernel->data.fl=cdata;
    cvFilter2D(image,imagep,pKernel);
    cvConvertScale(imagep, image, 1);
    cvReleaseImage( &imagep );
    uchar *data = ( uchar* )image->imageData;
    int g=skel->total;
    for (int y=0;y<image->height;y++){
        for (int x=0;x<image->width;x++){
            d=data[x+y*image->widthStep];
            if ((d<204) && (d>0)) {
                jointPointR.x=x;
                jointPointR.y=y;
                s=indexFinder(jointPointR,skel);
                if (s!=-1){
                    if (shortestPath(skel, jointPointR,A)->total<g){
                        g=shortestPath(skel, jointPointR,A)->total;
                        s0=s;
                    }
                }
            }
        }
    }

    cvReleaseImage( &image );

    return s0;
}

```

Figure A.2: Application sample code.