Model Free Human Pose Estimation with Application to the Classification of Abnormal Human Movement and the Detection of Hidden Loads

Benjamin A. Smith

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Michael Roan, Chairman
Martin E. Johnson
Dennis W. Hong
Thurmon E. Lockhart
Krishnan Ramu

July 21, 2010
Blacksburg, VA

Keywords: Pattern Recognition, Gait Analysis, Computer Vision, Surveillance

Model Free Human Pose Estimation with Application to the Classification of Abnormal Human Movement and the Detection of Hidden Loads

Benjamin A. Smith

Dr. Michael J. Roan, Chairman

Department of Mechanical Engineering

(ABSTRACT)

The extraction and analysis of human gait characteristics using image sequences are an important area of research. Recently, the focus of this research area has turned to computer vision as an unobtrusive way to analyze human motions. The applications for such a system are wide ranging in many disciplines. For example, it has been shown that visual systems can be used to identify people by their gait, estimate a subject's kinematic configuration and identify abnormal motion. The focus of this thesis is a system that accurately classifies observed motions without the use of an explicit spatial or temporal model. The visual detection of hidden loads through passive visual analysis of gait is presented as a test of the system.

The major contributions of this thesis are in two areas. The first is a neural network based scheme that classifies walking styles based on simple image metrics obtained from a single, monocular gray scale image sequence. The powerful neural network classifier utilized in this system provides an efficient, robust and highly accurate classification using these image metrics. This eliminates the need for more complex and difficult to obtain measures that are required by many of the currently human visual analysis systems. This system uses computer vision and pattern recognition techniques combined with physiological knowledge of human gait to estimate an observed subject's hip angle. The hip angle is then used to calculate a normality index of the gait. The hip angle estimate and normality index are then used as inputs to a neural network. It is shown through experiment that this system provides an accurate classification of four different walking styles observed by a single camera. Secondly, a computer vision based approach is presented that provides an accurate pose estimate without the use of an explicit spatial or temporal model. A hybrid fuzzy neural network is used to assign contour points of a silhouette to kinematically relevant groups. These labeled points are used to estimate the joint locations of the subject. The joint angles are shown to be good estimates as compared to ground truth angles provided by a motion capture system. The effectiveness of the system to distinguish between subtle gait differences is demonstrated by detecting the presence of hidden loads when carried by walking people.

To my parents Brian and Terriann and my brother James

Acknowledgements

I would like to thank Dr. Michael J. Roan for providing me with the opportunity to work on several interesting projects in visual surveillance and human analysis and for constant support, encouragement, motivation, and guidance in research as well as the valuable, detailed and often critical feedback on my writing and presentation skills. I will always recall the time I spent at Virginia Tech Vibrations and Acoustics Lab (VAL) very fondly for the various experiences I had both in and out of the lab. I would also like to extend my appreciation to Dr. Martin E. Johnson, Dr. Dennis W. Hong, Dr. Thurmon E. Lockhart, and Dr. Krishnan Ramu for serving on my advisory committee. Additionally I extend my thanks to Gail Coe for all the help and support she has provided to me in my years in the VAL.

I would like to thank the Office of Naval Research for funding this research. Next, I would like to thank my fellow researchers in the VAL laboratory. It has been a great pleasure working with them for the past several years. I would especially like to thank Tim Wiltgen, Minhyung Lee and Kamal Idrisi for their input and invaluable support over the years. I would also like to thank the various past and current members; Mark Sumner, Dan Mennit, Bonsung Koo, Jeong Im, Phillip Gillet, Mike Milo, Phil Chin and all the previous and current members of the VAL for all their support and help. I wish them all great success in their future academic and professional lives.

Thank you to all the members of Virginia Tech paintball for the great times both on and off the field. I will always remember building up the team and making it what it is today. I will always remember the good times I had with all the members of, the "A" team. I especially thank Casey Smith, Dean Wilson, Bobby Bowen and Alex Gunn for their help and friendship over the years.

The final word of acknowledgement is reserved for my family. Thanks to my parents Brian and Terriann for instilling me with the skills and ethics that I needed to succeed in my academic endeavors. I also thank them for their unconditional love and unending support. Thanks to my brother James for helping me whenever I needed it. I owe everything I achieve in life to my families' love, support and incredible patience.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Locomotion Analysis Background

Analysis of human walking movements, or gait, has been an ongoing area of research since the advent of the still camera in 1896 [7]. Since that time many researchers have investigated the dynamics of human gait in order to fully understand and describe the complicated process of upright bipedal motion. Stemming from this, a number of areas have emerged that attempt to exploit the analysis of this motion. These areas are highly researched and include clinical gait analysis [46] for rehabilitation and diagnosis, biometric gait analysis for automatic person identification [14 - 21, 24, 30] and passive human analysis for surveillance [6, 80]. These areas dominate the field of gait research at present. Although these fields analyze human gait, each one uses different techniques and systems. Clinical gait analysis focuses on collection of gait data in controlled environments using motion capture systems that are intrusive. Data acquisition must be implemented in special facilities that are used specifically for motion capture. Goals of passive human gait analysis are much different. The goal is to unobtrusively observe and analyze an individual's gait in a variety of different areas and scenarios in these systems. Because of the limitations, gait analysis for use in these systems is based on visual data capture. In other words, video of walking subjects is used to analyze gait. The complicated nature of this analysis necessitates the use of sophisticated computer vision systems to generate variables that describe gait motion.

Due to the wide range of applications, passive visual analysis of walking people is an area of intense research in the field of automated visual surveillance. Currently, these systems rely on complex image metrics [6, 85, 95, 122, 125, 132, 148], complicated classification schemes [96, 118, 124, 131, 162, 176, 176, 206, 207, 210, 257], multiple cameras [117, 130, 142, 215, 226] and computationally intensive model matching techniques [78, 91, 92, 93, 120, 150, 165, 198, 199, 228, 250, 261]. Traditionally, vision based gait analysis systems have focused on person identification [6, 14, 15, 16, 20, 21, 22, 23, 24, 24, 92, 97, 123], however, recent research has been done in many alternative applications [6, 92]. For example, systems attempt to classify activities [18, 30, 42, 80, 118, 124, 171, 207], gait abnormalities [45, 46, 111] or estimate the pose of individuals [25, 35, 36, 85, 96, 112, 113, 114, 115, 174, 196]. Of these applications, gait abnormality classification has many unique challenges and a broad array of applications. Numerous clinical studies have shown that accurate gait analysis can be used to diagnosis diseases [55], quantify the effects of obesity [53, 54, 55, 56, 57, 69, 72] and detect the carriage of external loads [49, 50, 51, 52, 59, 61, 64, 65]. Many of these systems rely on high resolution motion capture systems to measure key kinematic variables, such as joint angles. However, a visually based measurement system offers many advantages in these types of systems such as eliminating the need for complex motion capture systems as well as allowing for unobtrusive, cost effective and simple to implement gait analysis.

1.2 Current Human Locomotion Analysis Systems

Currently, most visual gait analysis systems are either based on spatial models, temporal fitting or direct image feature matching. Temporally based analysis systems attempt to relate temporal information contained in an image sequence to the observed gait dynamics. These models view the human body as an articulated kinematic chain that generates an observed image sequence. The changes from frame to frame are considered to be driven by the underlying structure of the observed subject. The changes in the image are then related to kinematic gait parameters [24, 30, 44, 79, 228, 232, 248] or spatiotemporal trajectories of relevant parts in the kinematic chain [170, 172, 241].

Of the spatial model based analysis systems, the W4 system [80] was one of the first to accurately analyze humans in a passive surveillance application. This system was able to track and analyze groups of people, as well as single people, within a scene. While [80] focusing on the tracking of people, and not the analysis of their motions, their system still allows for the tracking of individual body parts that can be used in conjunction with other analysis schemes to classify observed motions. In [80] shape analysis of the silhouette is used to locate and track individual body parts of an observed walking subject. Some of the more recent work in this area is that of Zhou et al. [78] and McIntosh et al. [119]. In Zhou et al. [78], they utilize a simple human model to estimate kinematic configuration from human silhouettes. They perform a dynamic Hough transform [88] to find the contour of a human torso in the image sequence. The region around the detected torso is windowed and normalized. The extracted silhouette is then converted to an edge image and compared to 6 different model prototypes using the chamfer distance transform [264]. The model configuration is initialized based on the level of matching between each prototype and the edge image. The model parameters are adjusted to minimize the error between the input silhouette and the chosen prototype. These parameters are used for the entirety of the gait cycle. An HMM framework is used to estimate the configuration of the model from one frame to the next. This pose is adjusted if the difference between the silhouette and the model is too large. The configuration is changed such that the amount of error between the silhouette and model is minimized. The testing done shows resilience to image noise and silhouette error, however, all results are qualitative in nature. No comparisons to ground truth kinematic measurements are shown. In McIntosh et al. [119], a human body model is learned from data using a deformable organism framework. Shape contexts [122] and normalized texture information of input images are used to initialize the limb locations of the model. The shape of each limb is then optimized based on the input image. Once the limb shape is determined, joint position estimates are refined from their initial values. Finally a local gradient search is performed to further refine limb position estimates. This is done hierarchically from limb to limb using a probabilistic framework. Qualitative as well as quantitative results are given using the Mobo [159] database. The results are comparisons of the model to manually segmented images as well as visual comparisons between the images and the calculated model locations. Again, they do not perform comparisons to ground truth kinematic data.

While model based approaches remain the most popular for visual gait analysis, systems which use subspace analysis techniques to analysis image sequences are also prevalent. These systems attempt to reduce the high dimensional image space (often thousands of dimensions) to just a few highly relevant dimensions. These methods often rely on analysis techniques, such as Principle component analysis [81], to reduce the image to a few highly descriptive values. The most recent of these methods is that of Elgammal and Lee [79]. They use a non-linear subspace embedding technique known as Local Linear Embedding to form a low dimensional gait manifold from an input image sequence. This reduces each image to a point on a three dimensional manifold. The relationship between this manifold and the actual body configuration of the observed person is then learned using a Radial Basis Function. The system is limited in its testing and it is not shown to accurately determine gait parameters directly from image sequences.

The final class of visual gait analysis systems relating to this work is those which attempt to directly relate image metrics to human body configurations. These systems often rely on relating edge or gradient information taken from an input image to a learned representation of the kinematic configuration. Two recent works in this area are Mori and Malik [122] and Knossow et al. [121]. In [122], shape contexts are used to initialize limb configuration of observed subjects. Points located on the edges of the image are then classified based on their distance from each of the proposed limb locations. These points are used in conjunction with a 2-D kinematic chain and deformation model to refine the limb position estimates. Using Taylor's method keypoints obtained from limb locations and a scaled orthographic projection model for the camera are used to estimate the 3-D configuration of the subject. Qualitative results are given that compare estimated kinematic configuration to the corresponding images of the subject. In [121], the extremal contours of the edge image of a subject are compared to the contours generated by a projection of a model onto the image plane. This method attempts to match multiple synchronous images of a human subject to a human model in two dimensions using edge information. The edges are matched using the chamfer distance transform and a best fit is obtained both for model configuration and model orientation. The results is an estimation of the 3-D configuration of the observed subject. Results are given compared to ground truth data for a number of body configurations, however, the subjects are performing exaggerated movements human locomotion analysis is not done.

1.3 Data Collection Methods

There are various sensor modalities used in the analysis of human movements. Each of these modalities varies in terms of implementation, cost and intrusiveness. The most popular types of captures systems used in the analysis of human movements are three-dimensional motion capture, single two-dimensional camera and multiple, calibrated two-dimensional cameras.

1.3.1 Three Dimensional Motion Capture

Commercial three-dimensional motion capture systems are the most commonly used in the field of biomechanics for analyzing human movements. These systems consist of multiple infrared cameras used in conjunction with reflective markers placed on important kinematic regions of an observed subject's body, often joint locations. The infrared cameras are calibrated so that they can calculate the exact three-dimensional position of each reflective marker. The positions of each marker are then used together to determine the exact configuration of the subject's body. While these types of system are intrusive, expensive and difficult to implement they provide the most accurate results. In many cases the positions calculated using these systems can be treated as ground truth locations.

1.3.2 Single Camera Systems

The simplest to implement human analysis systems are those that use a single two-dimensional camera. These systems are unobtrusive, cheap and easy to implement. However, the processing required for these systems are the most difficult due to the fact that they provide the least amount of information of any of the sensor modalities used in human movement analysis. Despite these issues, single camera systems are the most researched due to the ease of data collection that these types of systems provide.

1.3.3 Multiple Camera Systems

The use of multiple camera systems has arisen from the desire to obtain accurate three dimensional models of humans without the use of intrusive motion capture equipment. These systems utilize multiple synchronized and calibrated cameras to obtain multiple viewpoints of an observed subject simultaneously. These images are then used to form volumes in the three dimensional space by localizing a specific region of the object in multiple image planes. The volumes, or voxels, are then fit to three dimensional models of the human structure. These systems provide a large amount of data and result in improved three dimensional performance when compared to single camera systems. However, they require multiple calibrated cameras and the subject must be visible from a specific number of cameras at all times. This limits the application of these systems to a restricted environment.

1.4 Contribution

Although much research has been done in the field of computer vision for the analysis of human gait, very little work has been done in the area of global gait analysis for the detection of gait abnormalities. The majority of the work done in this area utilizes motion capture data [6], assumes that all joint locations are marked [17], utilizes synthetic image data, or is very limited in scope and effectiveness [21, 25, 27, 18, 43] (a limited number of grossly different human motions such as various gymnastic movements [21]). In this work, a system is proposed that determines if an observed subject walking normally or moving abnormally. The system presented in this thesis extracts gait variables from video in a novel way. A variety of computer vision techniques are used to estimate traditional gait descriptors (such as hip angle and height) from images. The assumption that human gait motion behaves as a damped harmonic oscillator is exploited in a pattern recognition scheme that is used to classify gait as normal or abnormal. The experimental results show that this approach yields an accurate classification of normal and abnormal gait when presented with a number of different walking subjects. These results demonstrate not only the validity of the human gait as an oscillator assumption, but also the validity of using video to analyze human gait to classify a variety of different abnormalities.

Additionally, a system is described that detects the conveyance of hidden loads by walking subjects. This is uniquely accomplished using spatial information without the use of an explicit human model. Silhouettes are extracted from individual images to generate a series of boundary edge points. These points are assigned to a relevant kinematic family (ie. shank, thigh, etc.) using a hybrid fuzzy neural network. The points of each part family are used to estimate the joint locations of the observed subject. This allows for joint angles to be estimated in each frame independently from preceding frames. This system does not rely on any type of temporally based image model or the fitting of an explicit spatial model, something that has yet to be shown in current literature. This allows for accurate kinematic estimation that does not rely on multiple cameras, computationally expensive algorithms or calibrated cameras. The system's effectiveness is shown through comparisons to ground truth motion capture data and in an application of hidden load detection.

1.5 Organization

The remainder of the thesis is organized as follows. Chapter 2 is a review of current human pose estimation and gait analysis techniques. Chapter 3 reviews the basics of biomechanics and outlines the effects of hidden loads on human gait. Chapter 4 discusses neural network design and training as well as fuzzy logic and fuzzy neural networks. The computer vision techniques that are used in this thesis are outlined in Chapter 5. Chapter 6 describes the system developed to classify abnormal gait and determine the type of gait observed. The visual hidden load detection system is presented in Chapter 7 and the conclusions and discussion of future work is covered in Chapter 8.

Chapter 2: State of the Art in Human Activity Analysis

2.1 Visual Human Activity Analysis

The visual analysis of humans has been a highly researched area in the past 10 years. This type of analysis requires the use of many of the most complex areas of computer vision. Chief among these are object detection, tracking and localization. However, human analysis remains a challenging, unique problem due to non-rigid objects (humans), varied and complex motions as well as appearance and size changes from subject to subject. Further complicating these issues are the requirements of the higher-level analysis systems that analyze the poses obtained by the visual systems. The requirements for not only high accuracy, but limited assumptions make the visual analysis of human subjects one of the most challenging in computer vision.

2.1.1 Common Human Analysis System Assumptions

Due to the problems associated with automated human visual analysis stated above, many systems rely on a range of assumptions to make the problem more manageable. Additionally, scenes are often further complicated by background clutter, dynamic lighting conditions, camera noise and changing motion. Each of these areas can be attributed to the background, camera, motion and subject. These assumptions are further detailed in the following sections.

2.1.1.1 Image Background Assumptions

The assumptions which deal with the area surrounding the subject in an image are commonly referred to as background assumptions. These types of assumptions are the most common in human analysis systems and can be created to deal with problems such as clutter and lighting. The first of these assumptions is that of a static background. The exclusion of moving objects simplifies the modeling of background regions in the image, which in turn, allows for the reliable extraction of a subject's silhouette in an image [13]. An extension of this assumption is that of stable lighting conditions. Within an image,

Figure 2-1: Example of cluttered background

dynamic lighting conditions can be difficult to model and lead to a large variation in individual pixel values in the image [13]. Another assumption which is commonly used in edge based analysis systems is that of an uncluttered background. This allows for the use of edge features that can be reliably attributed to the subject and not erroneous portions of the image. Another assumption commonly used in human analysis systems is that the

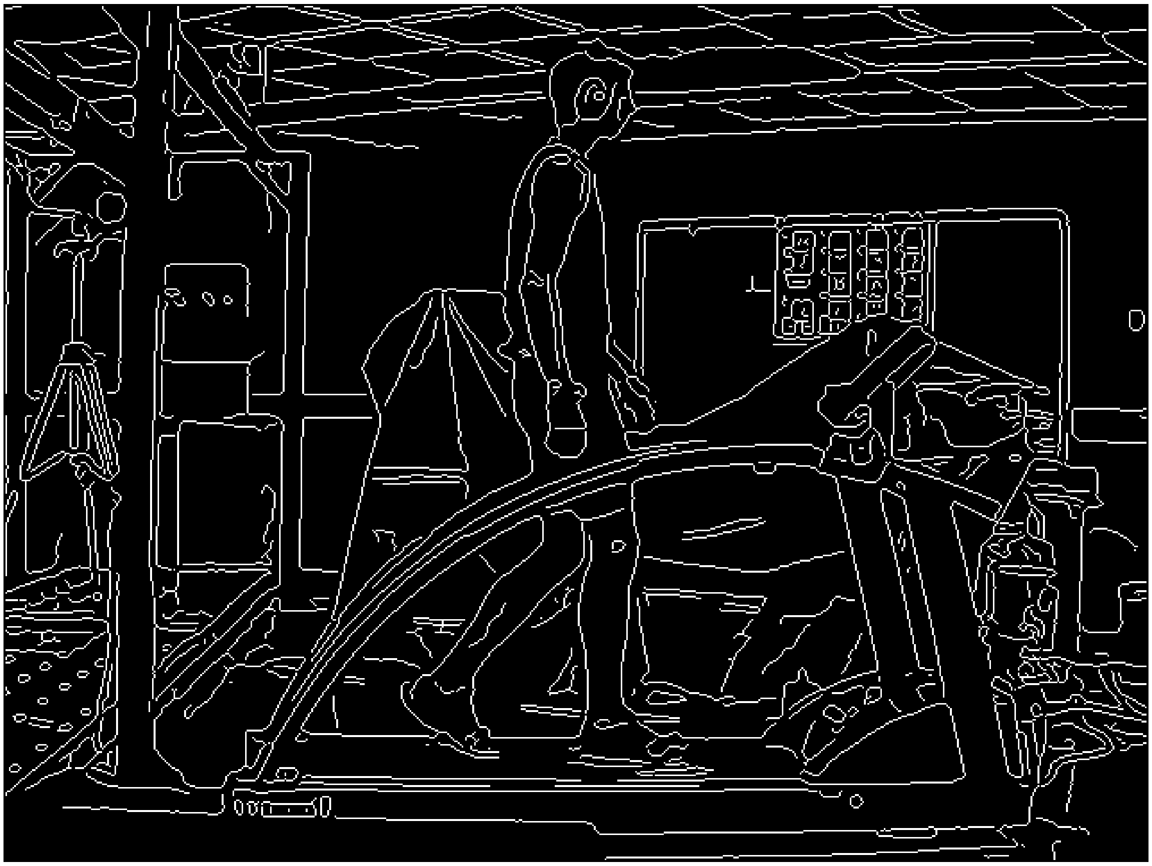


Figure 2-2: Example of cluttered edge image

image contains only one subject. This greatly simplifies analysis because all image features can be attributed to one subject without the need for assigning them to multiple subjects in a scene [92].

2.1.1.2 Camera Assumptions

The most common assumption that relates to the image capture device is that the camera is static. Camera motion can create large variations in the observed images that make modeling a scene difficult, if not impossible. Camera calibration is another common assumption that is used in systems that find the pose of a subject in three dimensions. A calibrated camera allows the user to know the position of the camera and regions within the image. This can be used to project the observed 2-D pose into a three dimensional space. These assumptions are also used in multiple camera systems that rely on known camera positions to form voxels from the pixel locations in each of the camera's images.

2.1.1.3 Assumptions of Observed Motion

Motion assumptions are made to reduce the complexity of dynamics that must be modeled in the system. The assumption that the subject is visible in all frames eliminates the need for detecting the subject in each frame. Many systems assume that the frame rate of the video is sufficient enough to prevent large jumps of the subject from frame to frame, this allows for motion limitations that greatly reduce the search space in each image. The final assumption deals with the motion of the subject. This is that the motion being observed is simple and predictable by a previously defined model. The assumption of known motion allows for reliable tracking and improves the accuracy of spatial pose estimation in many systems.

2.1.2 Types of Human Activity Analysis Systems

Human visual analysis systems utilize a variety of techniques. The most common types of systems use spatial models, tracking methodologies or manifold methods to estimate the observed subject's parameters.

2.1.2.1 Spatial Model Based Analysis Systems

A popular method used to calculate the pose of human subjects is spatial modeling. These models form an approximate representation of the body and attempt to fit this representation to the observed image data. These models can be used directly in the image space as 2-D models, in conjunction with depth or probabilistic information

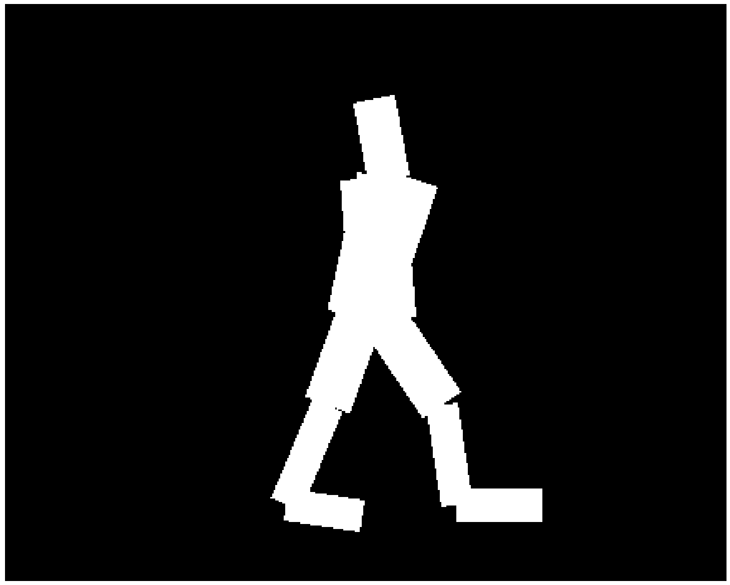


Figure 2-3: Spatial Humanoid Model

to fit 3-D models to 2-D images. An array of cameras can be used to fit a 3-D model.

Many different types of 2-D human models have been proposed in the literature [78, 91, 92, 93, 120, 150, 165, 198, 199, 228, 250, 261]. Often times these models assume that the human structure is a collection of articulating rigid parts. Recent literature has shown that template based [78, 161, 235] and 2D image blob models [115, 162, 163, 259] can be effective in estimating the pose of observed subjects. Often times, these models are limited to parallel or front views in order to reduce the effect of small out-of-plane rotations. Some systems attempt to handle out of plane rotations by using depth in their model in addition to the 2-D image information [265]. A major limitation of 2-D models have is self-occlusion and differentiating between left and right limbs.

In order to overcome the problems and limitations of 2-D body models, three dimensional models have become popular. These models often use multiple view points (provided by multiple cameras) [117, 131, 193, 197, 266], stereo vision [228], or strong prior information [78, 125, 226] in order to find a subject's pose in 3-D. Many of these methods expand on template based 2-D

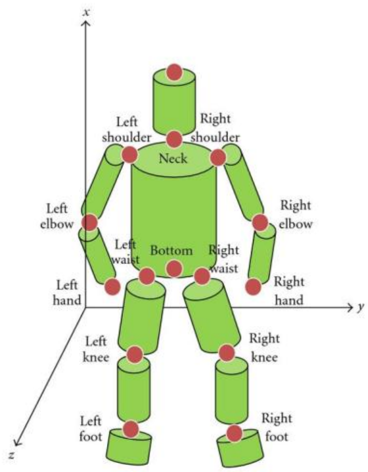


Figure 2-4: Three dimensional humanoid model, taken from [271] Y. Wang and K. Cheng 2010.

models and create a 3-D model consisting of volumetric parts [130]. These models assume that the subject is a collection of articulating rigid parts in the same manner of the 2-D models. While these types of systems often provide much more information, the models are ill-constrained which makes it difficult to accurately determine the 3-D pose of a subject.

2.1.2.2 Human Tracking Methods for Activity Analysis

Tracking methods attempt to find the location of body parts by estimating the subject's motions from frame to frame in an image sequence. This can be done in a variety of ways. A popular method is to assume the initial pose and position of the subject is known and track the proceeding frames in the sequence [6, 30, 92, 131, 164, 165, 193, 209, 226, 228, 237, 248] to recover the subject's pose throughout the image sequence. While these systems perform well when the motion is known, they require a specific type of motion and the initial position and pose to be known. This is not always the case in human visual analysis. Another type of tracking method is those that attempt to find the motion information for each limb from a video. These motion trajectories are fit to known motion models for each limb.

2.1.2.3 Manifold Methods for Activity Analysis and Viewpoint Estimation

Manifold methods attempt to condense an image sequence to a reduced dimensional manifold [79]. These methods use dimensionality reduction techniques to find the underlying motion manifold of observed image sequences. These dimensionally reduced sequences are then used to learn a representative manifold structure in the manifold subspace. The manifolds for various observed motions can then be classified in

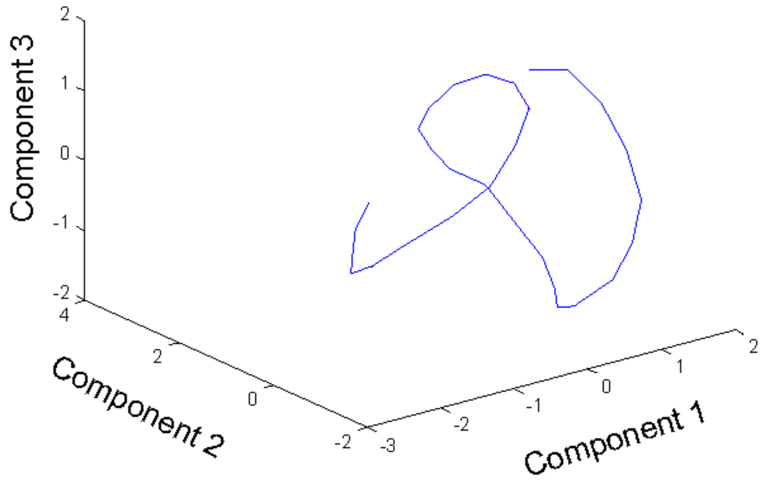


Figure 2-5: Gait motion manifold using Locally Linear Embedding (LLE)

the manifold space. These types of systems have shown that they can be generalized to multiple view-points [79], however, they have difficulty analyzing unseen subjects or motions.

2.1.3 Types of Human Body Models Used for Activity Analysis

The different types of body models that are used to estimate pose have distinct differences in construction and how they are fit to the observed images. Model types can be two or three dimensional and are fit to images using different features based on the application. The two predominant types of models used in recent works have been Kinematic Trees and Part Based systems.

2.1.3.1 Kinematic Tree Type Human Body Models

The Kinematic Tree based representations assume that each part of the body combines to form a tree structure. The pose of the subject is parameterized by $\mathbf{X} = [x_r, \theta_r, \theta_1, \theta_2, \dots, \theta_N]$, where $x_r \in R^3$ is the position of the root; θ_r is the orientation of the root in the world frame of reference and θ_i is the orientation of limb *i* of reference.

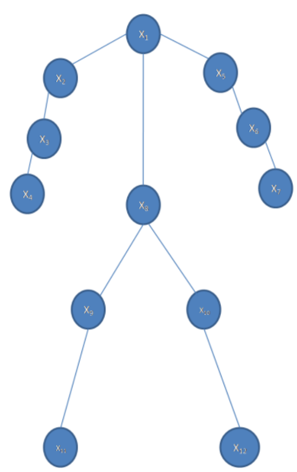


Figure 2-6: Human kinematic tree

This often leads to a high dimensional pose space because the joint orientations are characterized by two or three angles.

2.1.3.2 Part Based Human Body Models

In contrast to Kinematic Trees, part based systems treat each body part individually and link each part of the body together with a function which enforces body configuration. These types of models use a simple representation of each body part such as rectangles [79, 169], trapezoids [195, 196], or templates [149]. The positions of each part of the model are fit to the image in terms of either compatibility [172, 174] or probabilistically [122, 169, 196, 197, 248]. The models are

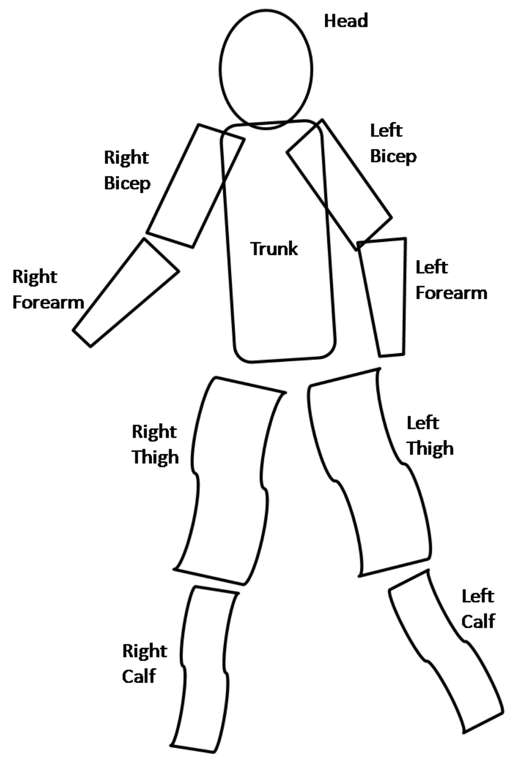


Figure 2-7: Part based human model

created using kinematic constraints which enforce the position of each limb relative to the others.

2.1.4 Image Features Used in Human Activity Analysis

The dimensionality of images requires that features be extracted before attempting to estimate the pose of a subject. These features are based on the system. Many different types of image features are used in pose estimation systems. The most common types are described below.

2.1.4.1 Image Silhouettes

Silhouettes are one of the most popular features used in human visual analysis systems. This is due to their relatively straightforward computation and ability to quickly localize the subject within an image. Silhouette calculations are more robust to image clutter and edge noise; however the level of information they provide is limited.

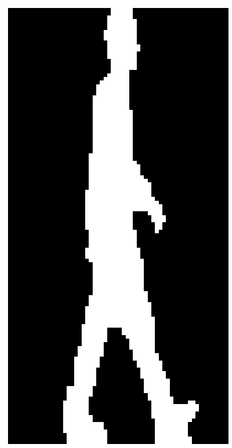


Figure 2-8: Human silhouette

Silhouettes are often formed by extracting the foreground from an image. The area of the foreground is then assumed to be the silhouette. An example of this is Chroma-keying [133] where subjects perform actions in front of a green or blue screen that can easily be edited out. More complicated scenes require sophisticated analysis techniques. The most common class of background subtraction techniques are based on statistics. These types of systems use a sequence of images to learn a model for each pixel in the image, these are often Gaussian. Once the model parameters are learned, each pixel's membership to the background can then be calculated. In order to handle dynamic background conditions, many recent systems use Gaussian mixture models [216] or nonparametric models [13] to calculate a pixel’s multi-modal intensity distribution. After background subtraction, the resulting silhouette can contain high amounts of clutter and false positives due to pixel variations. These errors are corrected using morphological filtering operations.

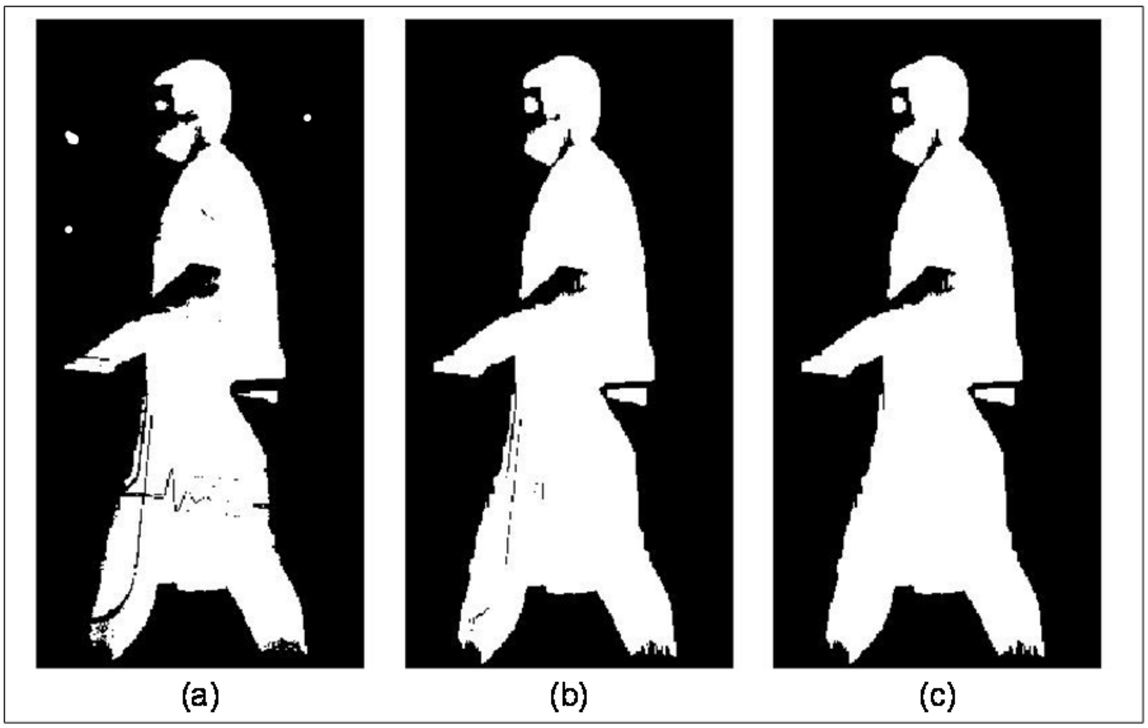


Figure 2-9: Morphological filtering of a silhouette, (a) input silhouette, (b) dilation filtered image, (c) final silhouette

These operations serve to remove noise and provide a more consistent silhouette. Several methods that use color thresholding [90] have also shown promising results when using a color camera.

2.1.4.2 Color Features

The addition of color or texture information of the image allows for more sophisticated analysis than that provided by silhouettes. Color information is often used after a subject has been found in an image to provide a more accurate pose estimate. Additionally, many systems use color images for skin-color detection and segmentation. These systems rely on the assumption that skin has a distinct color and texture and certain areas of the body will always have skin exposed [126, 127]. These models are often based on probability of a certain pixel being skin. This formulation is given in equation 2.1.

$$p(skin|I_{x,y}) = \frac{p(I_{x,y}|skin)p(skin)}{p(I_{x,y}|skin)p(skin) + p(I_{x,y}|\sim skin)p(\sim skin)} \quad (2.1)$$

This probability can then be compared to a threshold to determine if the pixel in question contains skin or not. The parameters from equation 2.1 are often learned from exemplar data.

While silhouettes provide a large amount of information and are calculated using proven techniques, they are still susceptible to many problems. Chief among these is if background colors appear on foreground objects because there is no way to distinguish between foreground and background in these cases. Shadows are hard to remove from images and cause false positives on the silhouette. A large number of objects in the image can also make distinguishing between separate foreground objects difficult or impossible.

2.1.4.3 Image Edge Features

Edge features are commonly used due to their semi-invariance to viewpoint, lighting and texture. They can also provide insight into the interior of objects that silhouettes are unable to provide. Edge features can be calculated a variety of ways, however, most are based on spatial image derivatives. Directional filters can be used in order to obtain edge orientation in addition to location. Edge features tend to be highly

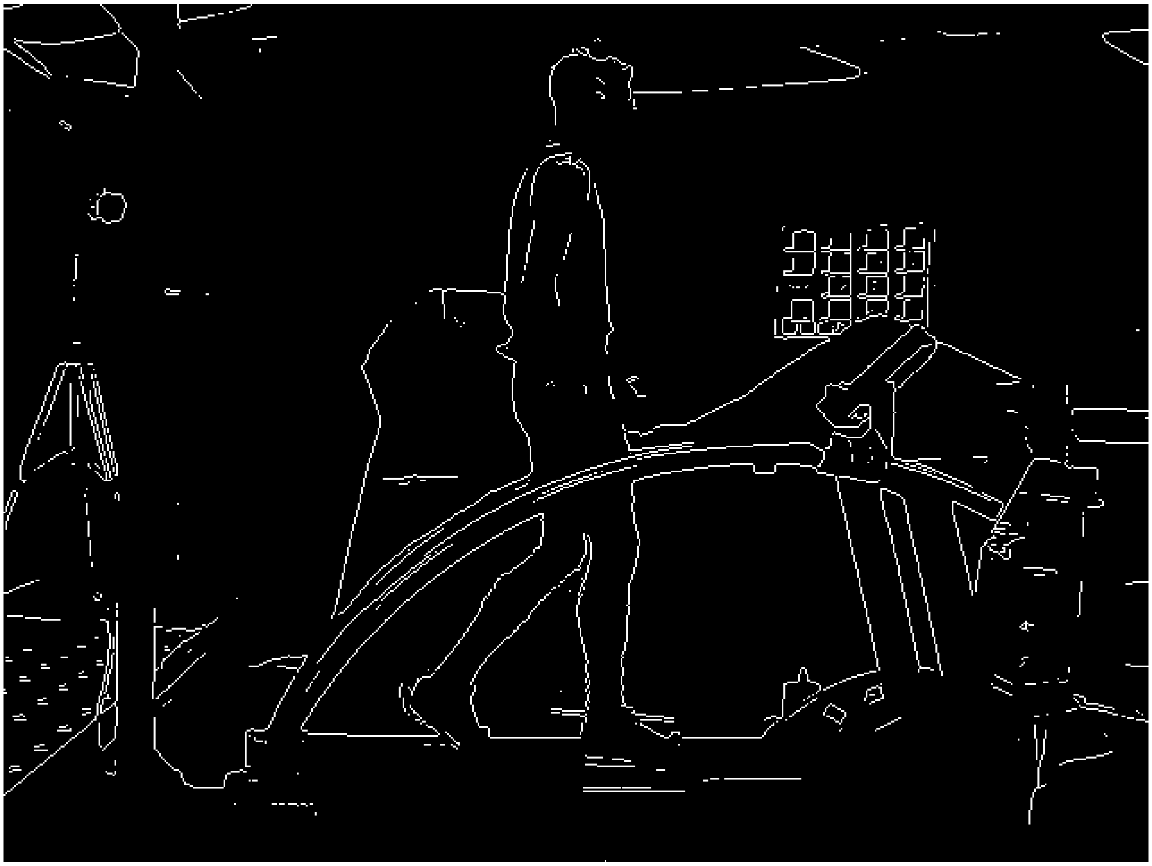


Figure 2-10: Edge image

local. Because of this many systems use spatial filtering [86] to remedy these problems.

2.1.4.4 Human Contours

Contours are the full or partial outline of an object. These can be either static or dynamic. Static contours are just another representation of a silhouette and provide no additional information to the silhouette. Dynamic contours, or active contours, allow the contour to deform according to the observed image edges. These deformations are controlled by energy functions based on image edges and a smoothness parameter. The deformations can also be based on prior knowledge of an object. These types of systems are referred to as deformable templates [77]. These types of models are limited in the analysis of articulated objects and provide a poor estimation of pose when using a single

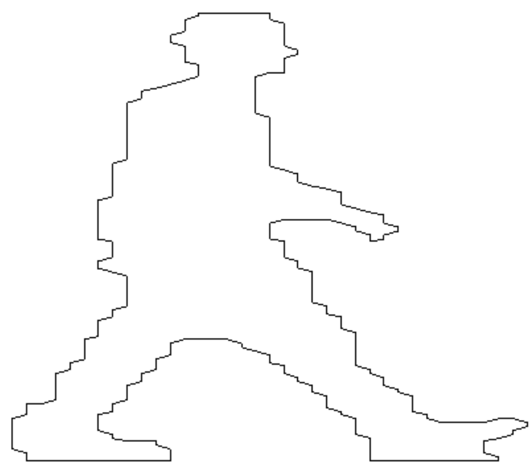


Figure 2-11: Human contour

camera. However, promising results have been shown when used in multiple camera systems [184].

2.1.4.5 Ridge Features

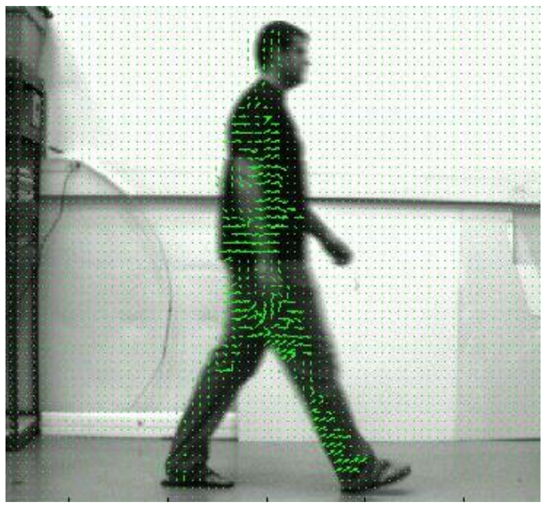
Ridges are the second spatial derivative of an image. These derivatives account for elongated structures within an image and have shown to be effective in modeling limbs [192]. Ridges are particularly effective at finding parallel edge structures within an image. This makes them ideal for limb detection. Noise can be a significant problem since they are a high order derivative. The filters are also highly sensitive to orientation and scale.

2.1.4.6 Image Sequence Optical Flow

Optical flow is the estimation of motion taking place within the image. This is accomplished by finding matching image regions between consecutive frames within a sequence. This results in an horizontal and vertical velocity for each pixel. The equation used to calculate optical flow is given in equation 2.2.

$$\frac{dI(x,y,t)}{dt} = \frac{\partial I(x,y,t)}{\partial x}\frac{dx}{dt} + \frac{\partial I(x,y,t)}{\partial y}\frac{dy}{dt} + \frac{\partial I(x,y,t)}{\partial t} = \mathbf{0} \qquad (2.2)$$

Solving equation 2.2 for $\left[\frac{dx}{dt}, \frac{dy}{dt}\right]$ results in an optical flow field. In order to help alleviate the noise in the optical flow field locally constant motion is assumed. While this method



(a)

Figure 2-12: Optical flow field for an image

does help reduce the noise it can cause problems when analyzing articulated objects because the constant motion assumption is frequently violated. Due to these issues, optical flow is often used in conjunction with other techniques in human pose estimation.

2.1.4.7 Three Dimensional Voxels

Voxels are a three dimensional pixel generated using multiple viewpoints []. Typically this is done by find corresponding pixels in the images of a number of calibrated cameras [142]. A three dimensional silhouette can then be found by projecting finding a silhouette for each image. If the projection of a voxel falls within the silhouette for all viewpoints, then that voxel is assumed to be within the volume of the three dimensional silhouette.

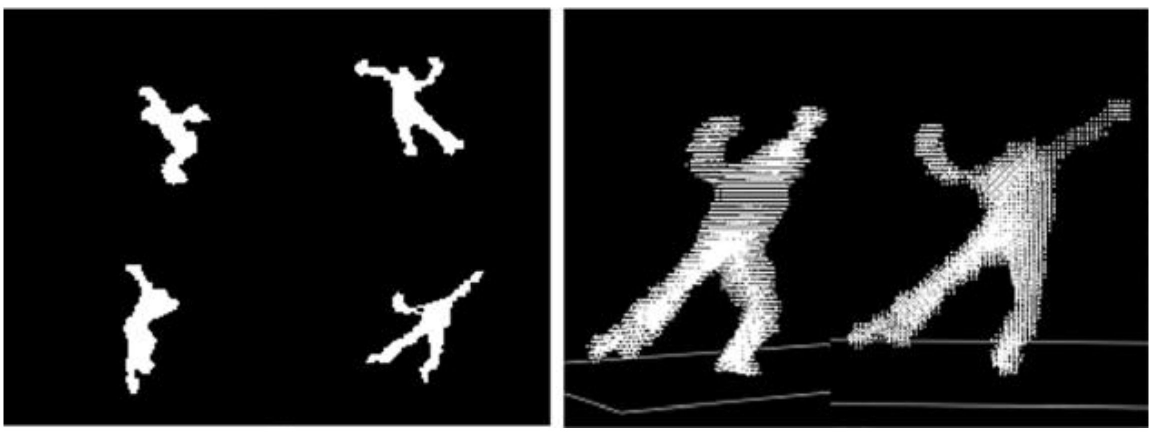


Figure 2-13: Example of three dimensional silhouette, taken from [270] I. Mikic and M. Trivedi 2003.

A voxel is considered to be in the background if any of its projections are located outside of the silhouette for its associated viewpoint. While voxel based features are quite powerful in human pose analysis they still rely on an accurate calculation of a silhouette and therefore suffer from the same limitations as traditional 2-D silhouettes. Also, the need for multiple calibrated cameras makes the image data difficult to obtain.

2.1.4.8 Shape Contexts

Shape context is an image descriptor that is based on binary edges. For each of the edge points the distance and angle to each of the remaining edge points are calculated. A log-polar histogram is then formed using these distances and angles. The process is then repeated for each of the selected edge points. The collection of these histograms is typically used as the descriptor for the object. This metric can be computationally expensive but has invariance to object rotation and translation.

2.1.5 Human Pose Estimation for Activity Analysis

Many of the models, techniques and image metrics described above are used to estimate the pose of an observed human subject. Although a wide variety of techniques have emerged most can be grouped into one of two categories: discriminative or generative. While these methods each use similar features and models, they attempt to solve the problem of human pose estimation in two different and distinct ways.

2.1.5.1 Discriminative Models

Discriminative approaches desire to find a direct mapping between the image feature space and the two or three dimensional pose space. These often use a single image [113, 179, 181, 189, 206] and either silhouettes or edges as image features. A probabilistic mapping is learned using nearest neighbor search [189], regression [206] or specialized mappings [179]. These methods can provide accurate results given an adequate training set. The main drawback to this type of approach is that they rely heavily on the quality of the image features. This makes them very sensitive to image noise and errors in feature calculation.

2.1.5.2 Generative Pose Estimation

Generative approaches to human pose estimation are the most popular methods used. These approaches rely on kinematic tree or spatial models. The pose is defined by a set of parameters which represent the position and limb orientation of the subject. The pose is estimated calculating the likelihood of a configuration from the image features. While these methods are robust to noise they require a very high dimensional search space that can be difficult to implement. These methods often must be manually initialized and rely heavily on tracking to retain accuracy throughout an image sequence.

2.1.6 Current Evaluation Methods for Human Activity Analysis

Currently many pose estimation systems are evaluated qualitatively due to the lack of reliable "ground truth" pose data. Qualitative methods are widely used and evaluation relies on visual inspection of results [6, 92, 174]. While this type of evaluation can provide a feel for the general performance of an algorithm, the accuracy is often far from what is needed for many human gait analysis applications.

Recently some quantitative evaluation of human pose estimation systems have been done. For many 2-D human pose systems the algorithm's performance is compared to manually labeled data [170, 173]. However, these types of performance metrics often use pixels as the distance metric and can vary greatly when comparing results. A system that uses motion capture data to compare the performance of the pose estimation, such as that done in this thesis, is desirable. This facilitates the direct comparison between ground truth kinematic position data and the pose estimated by the system.

2.2 Biomechanical Analysis of the Kinematics for Loaded Gait

Little work has been done to study the effect of external load carriage on human gait. The current studies are limited to general effects and rely solely on 3-D motion capture systems. These studies have shown that differences between loaded and unloaded gait exist in various single pack loads [49, 50, 51, 52, 59, 61, 64, 65]. The work in this thesis attempts to find differences due to double pack loading without the use of 3-D motion capture systems.

2.2.1 Three Dimensional Motion Capture Systems

Three dimensional motion capture systems utilize some number of specialized cameras in order to exactly find the position of reflective markers that are placed on a subject's body. These systems provide highly accurate, highly sampled data that can be used to exactly calculate a joint's position. While these systems provide high quality data, they remain intrusive and highly restrictive in their implementation.

2.2.2 Traditional Gait Analysis Techniques

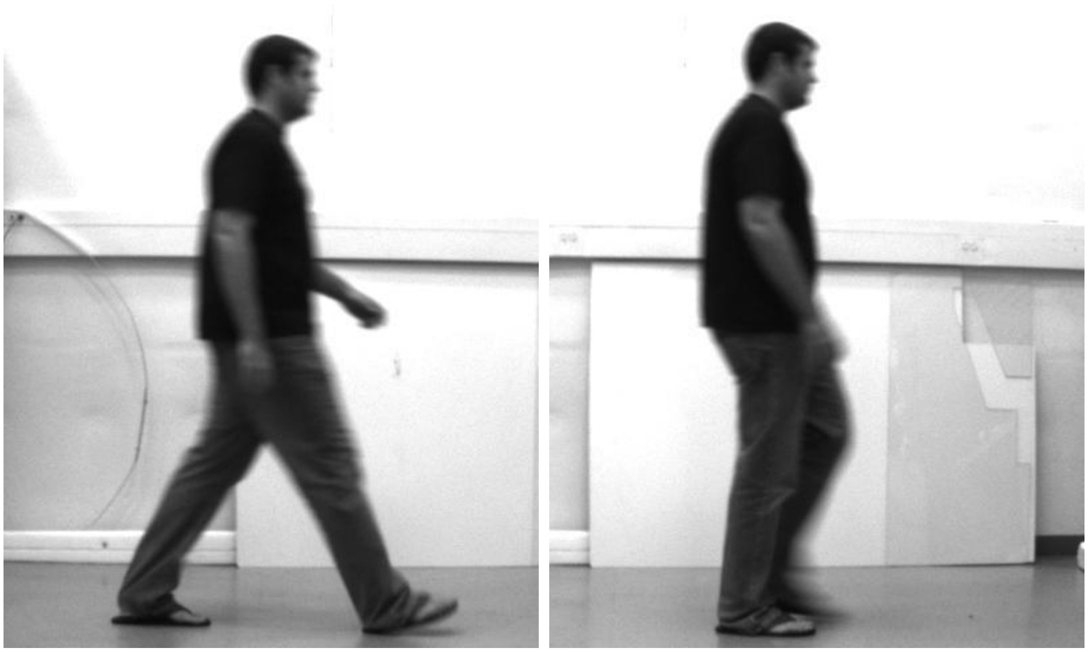
Many current biomechanical gait analysis systems rely on linear spatio-temporal measurements such as stance times, stride frequency and trunk inclination. While these metrics are easy to calculate from motion capture or accelerometer data the information they provide is rather limited. Recently, more complicated methods have been investigated such as continuous relative phase (CRP) analysis. This technique has been used to evaluate limb coordination in several studies [66]. Phase analysis can be used to show coordination between two oscillating components [106] which can give insight into the control of a system. This makes phase analysis very useful for investigating complex effects on human movement.

Chapter 3: Biomechanics of Human Locomotion

The process of a human moving from one place to another is defined as locomotion [7]. Because such a large number of different motions can arise during locomotion, the movement from one place to another at a regular speed is the focus of explanation for this section. This type of motion is governed by the rhythmic displacement of body parts that result in a constant movement in the desired direction. Bipedal motion requires a very specific rhythmic pattern to avoid a number of instabilities and unnecessary energy expenditures. Because of this particular rhythmic motion, gross similarities are found in human gait between individuals that are dissimilar in proportion, ethnicity, or any other attribute. This rhythmic motion can be broken down into a number of phases that dominate the motion.

3.1 Human Gait Dynamics

Due to the repeatability and predictability that must take place during human gait the motion can be characterized by a few key phases and movements. The legs oscillate underneath the upper body at a certain and repeatable frequency in order to maintain a certain velocity. The movement parallel to the plane of progression is very large in relation to the movement perpendicular to the sagittal plane. In addition the body undergoes a vertical displacement as the legs oscillate. With respect to these motions, inferences may be drawn from the fact that an efficient system always attempts to minimize the amount of energy expended to accomplish a task. Using these inferences, the hypothesis that all of these motions will be incorporated in a fashion that minimizes energy expenditure while maximizing the amount of distance traveled per cycle can be made [7]. Stemming from the above hypothesis, human walking is defined as a repeatable process in which a human body is supported in an upright fashion by each leg. While being supported by one leg, the opposite leg is moving forward underneath the body in the direction of motion so that once the moving leg is as far forward as the cycle allows, it becomes the support leg and the trailing leg now swings forward. This cycle is repeated to accomplish forward motion of the entire body. Two specific factors must be in place for the gait cycle to be accomplished. A ground reaction must take place beneath each foot and each leg has to continue its periodic motion uninterrupted for the duration of the walking cycle [7]. Because of these tenants, there are specific body motions that must universally take place in each individual's gait. One important fact is that the body must accelerate and decelerate in the direction of motion during the gait cycle. This comes from the way the support and swing legs move underneath the body, when the support leg is behind the body the dynamics allow the body to accelerate at a slightly greater rate, when the support leg is located in front of the body there is an associated deceleration [7]. The structure of the body itself causes a vertical movement associated with where the support leg is located. When the leg is in front of or behind the upper body a triangle is formed and the height is therefore reduced as compared to when the support leg is directly underneath the upper body. An example of this is shown in Figure 3.1.



(a) (b)

Figure 3-1: Comparison of double stance phase and single stance phase, notice how the configuration of the legs results in a change in vertical height from one phase as compared to the other

As can be seen in Figure 3.1, the physical structure of the body coupled with the motions necessary to move the body forward efficiently result in vertical motion. This may also be observed quantitatively using a commercially available motion tracking system to observe and record the three-dimensional location of points on the body with respect to time. Figure 3.2 shows the vertical motion of the head of a subject during normal walking. The head predictably moves in a cyclic up and down motion.

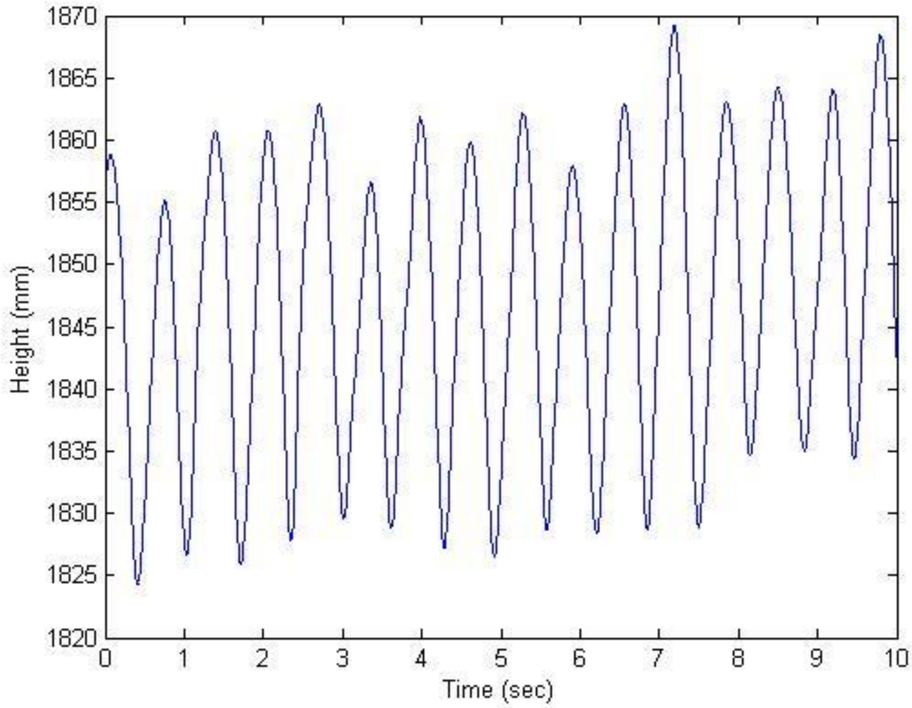


Figure 3-2: Vertical motion of the top of the head measured during normal walking; notice that the motion is dominated by a roughly sinusoidal component

In addition to the vertical motion of the subject, another necessary motion that most take place repeatedly throughout a period of gait is the angular variation between the two legs as viewed in the two-dimensional sagittal plane. The sagittal plane is the plane parallel to the direction of motion. This angular variation takes place when the legs move from the double stance phase to the single stance phase as seen when the subject transitions from (a) to (b) in Figure 3.1. This angle can be calculated using position data similar to that used to find the vertical motion of the head, however the knee positions and hip location in the sagittal plane is used to calculate the angular data. An example of this hip angle data is given in Figure 3.3.

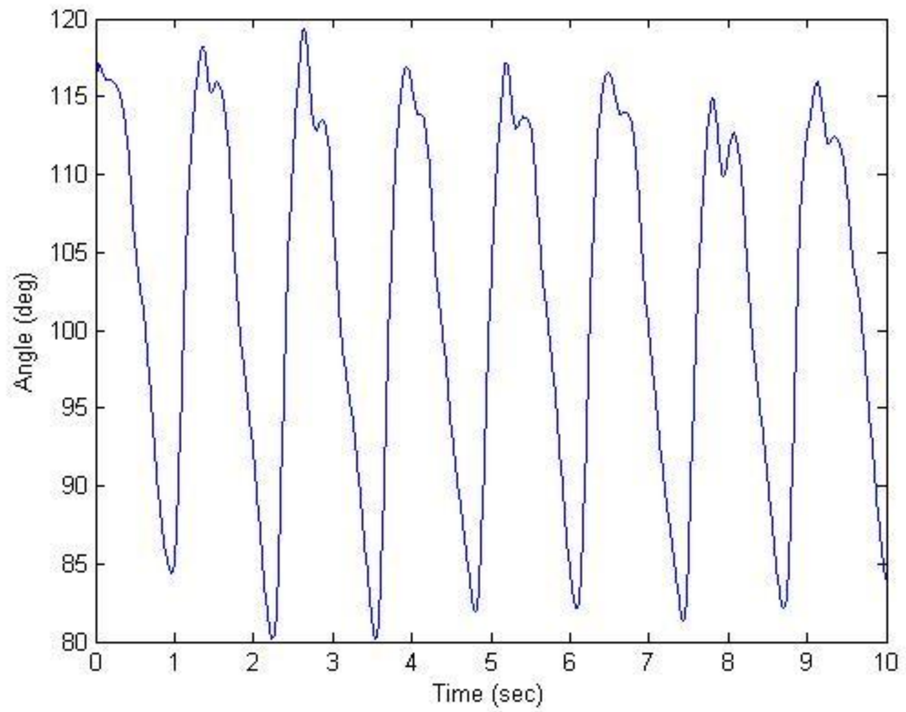


Figure 3-3: Hip angle of a subject calculated using motion capture data, the angle is degrees down for the plane parallel to the ground at the height of the hip, notice how the motion is dominated by a sinusoidal component with subtle, consistent preturbances

Figure 3.3 shows that when a natural unimpeded gait cycle is being performed, the hip angle changes in a predictable and repeatable manner. This phenomenon is not limited only to the hip angle and can be generalized to the knee and ankle angles as well. Although there are slight differences in the motion between individuals, the majority of the motion taking place is similar for each individual during normal gait. Figures 3.4 and 3.5 show the knee and ankle angles for one subject walking normally.

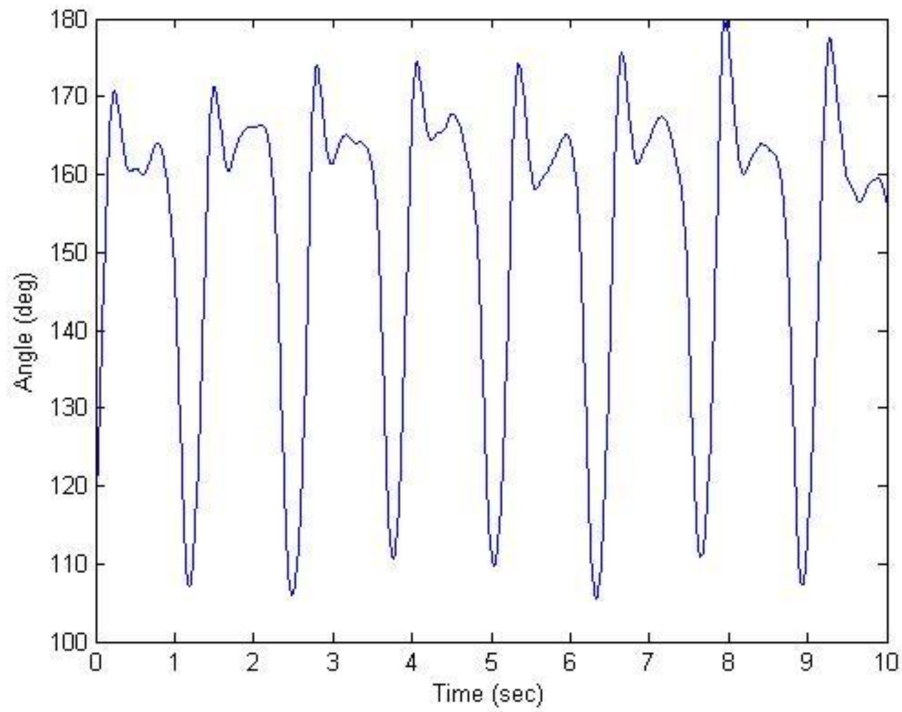


Figure 3-4: Knee angle, 180 degrees is a straight leg, notice the consistent oscillatory motion present during normal walking

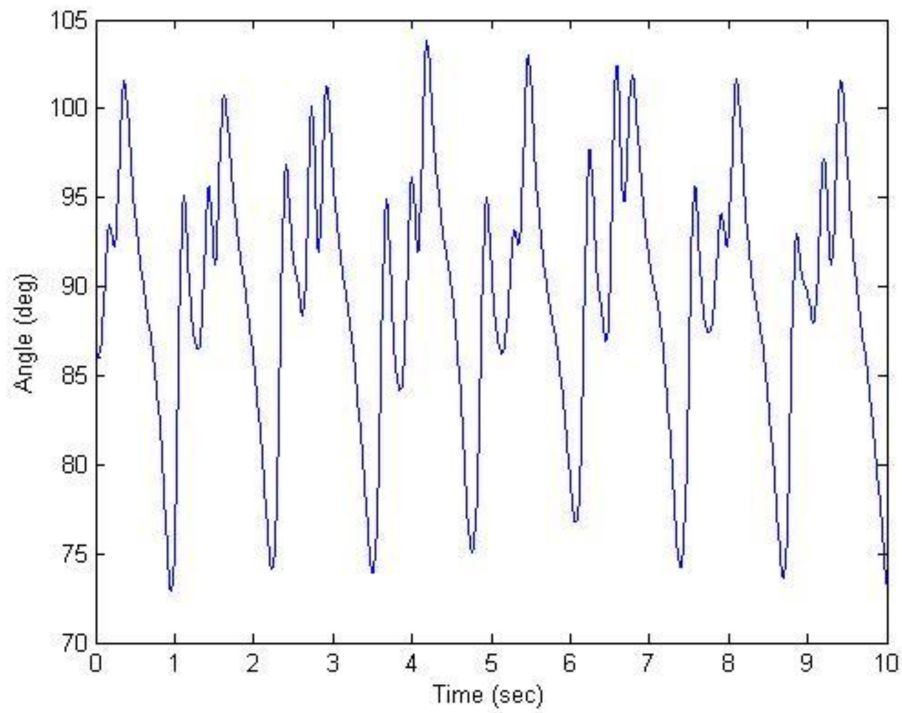


Figure 3-5: Ankle angle of a subject during normal walking

The acceleration and deceleration experienced by the body during normal walking is another similarity found between individual's gait cycles. This can be seen when observing normal human gait and can be quantitatively found using a motion tracking system. This is accomplished by tracking the upper body motion in the sagittal plane and taking the necessary temporal derivatives to arrive at acceleration. The upper body acceleration of a subject walking normally can be found in Figure 3.6.

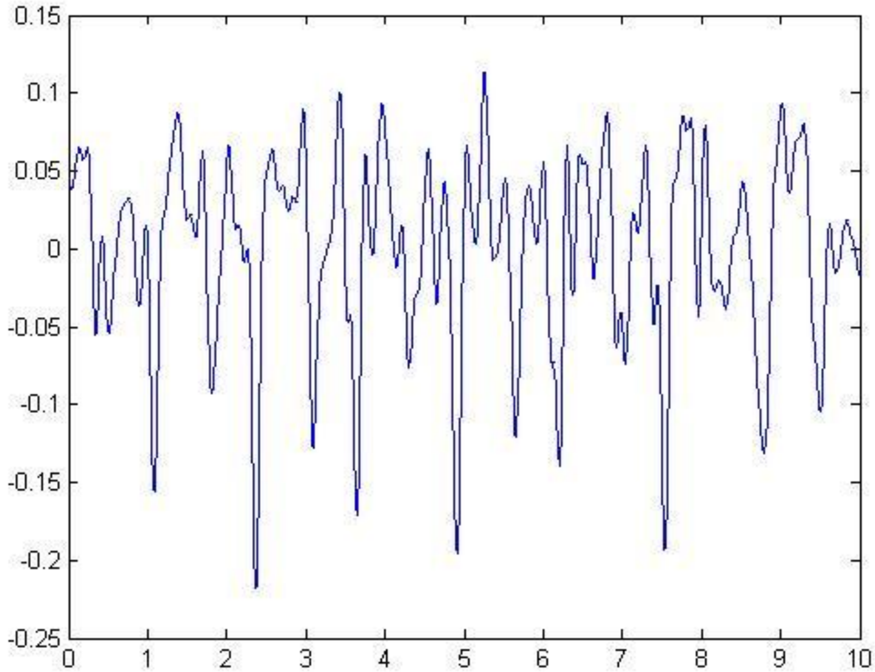


Figure 3-6: Upper body acceleration of a subject walking normally on a treadmill. Calculated using the second derivative of the shoulder position measured in the direction of movement. Notice small acceleration changes present in each cycle

Again, as seen in Figure 3.6, the data shows a repeatable cyclic nature for the subject. The data shown in Figures 3.2-3.6 demonstrates that a human subject walking normally experiences a predictable and repeatable cycle of motion. These overall repeatable motions can be found and measured quantitatively by a motion capture system. Due to the nature of normal gait, the majority of these motions can be generalized for an entire population. This is because the complex motion which must take place in each individual's gait is governed by the same goals and limitations. Figures 3.7-3.10 found below show various descriptors of the human gait for four individuals; these descriptors are hip angle, knee angle and ankle angle. Two cycles are shown for each individual in order to show detail of the individual waveforms generated for each cycle.

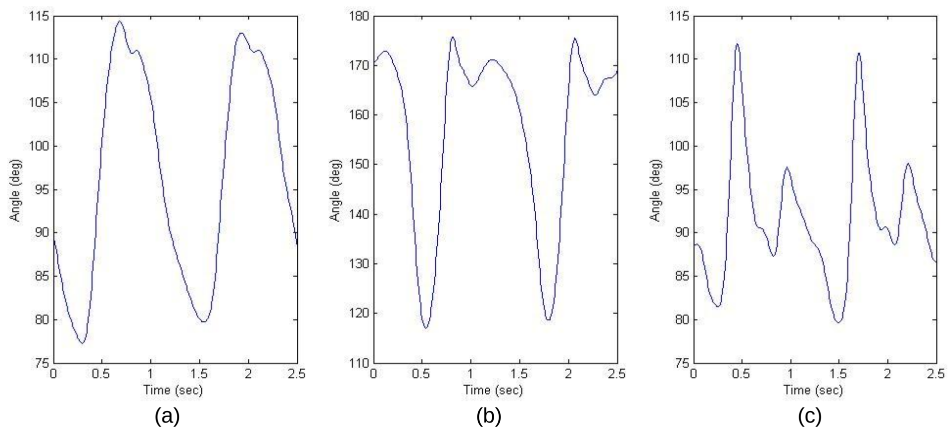


Figure 3-7: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle

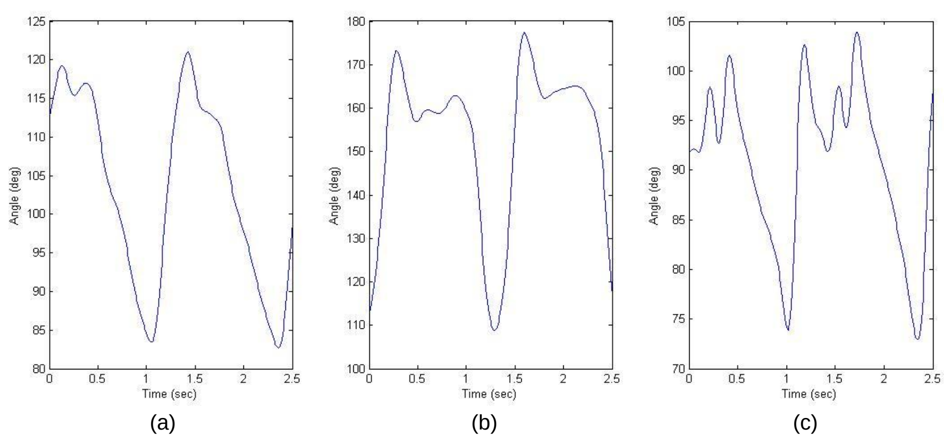


Figure 3-8: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle

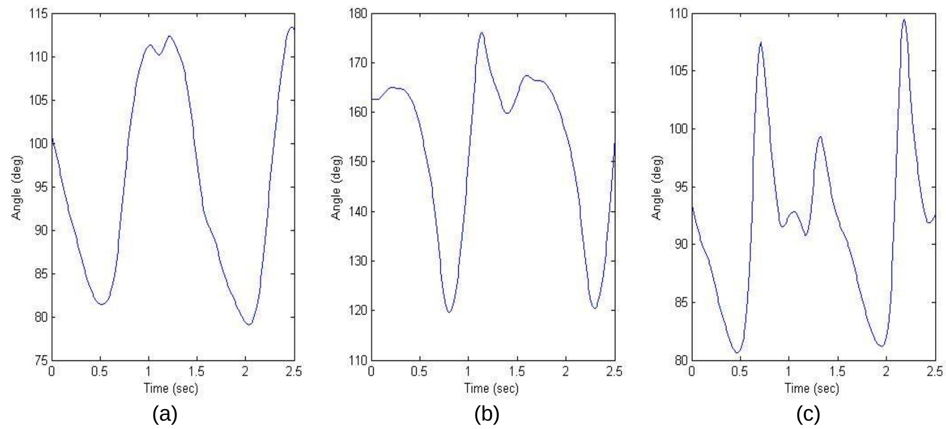


Figure 3-9: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle

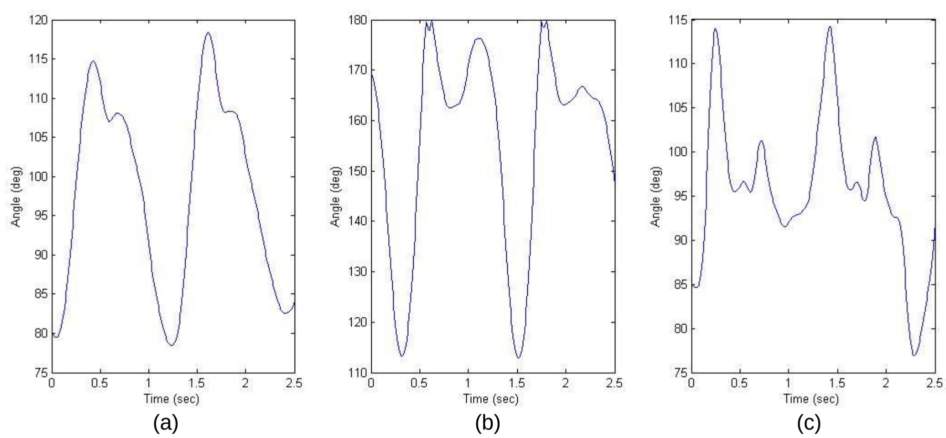


Figure 3-10: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle

As can be seen in Figures 3.7-3.10, small differences between individual subjects and their gait cycles can be found. However, the overall motion remains similar for each subject throughout the period of movement as is described above. From this observation the hypothesis arises that if an individual is walking abnormally the compensation for these differences will result in a change in the gait cycle. These changes result in similar differences to the oscillatory motion that characterizes gait for all individuals.

3.2 Gait Kinematics during Load Conveyance

Load carrying is a common cause of injuries to the knee and lower back [48, 49, 51]. This has motivated previous studies that characterize the human effects of load carrying including the effect on gait patterns. According to these studies, the duration of the double stance increased with increased loads, while the single stance duration decreased in 10 healthy males [49] and in 15 boys [50]. Significant gait differences were observed between loaded and unloaded walking. These differences were dependant on whether the load was a backpack or double-pack. With a backpack load, forward leaning of the trunk is a natural behavior to help keep the center of mass over the feet. It was found that the forward inclination considerably increased with load weight to minimize energy cost [51]. This minimization of energy expenditure resulted in the decrease of vertical positions at the knee and ankle with the added weight [52]. Several published studies indicated that pelvic rotation reduced and ankle rotation increased in the sagittal plane under loaded conditions [49, 51]. These studies also showed that knee flexion after impact was greater when carrying loads in order to absorb increased impact forces.

Excessive body weight has also been linked to large number of health problems such as cardiovascular disease, stroke, hypertension, and diabetes as well as numerous gait related injuries [53]. A limited amount of work has been done to investigate the injury related gait kinematics of overweight individuals. The kinematic deviations include slower velocity, shorter step length, increased double support time, decreased knee range of motion, and larger ground reaction forces compared to normal weight individuals [54-57].

Little work has been done to study the effects of external load carriage. Previous studies focus on external load carriage by normal weight individuals using mostly linear spatial-temporal measurements such as double stance time, stride frequency and trunk inclination [75]. Recently, Haddad and Emmerik [59] examined the intralimb and interlimb adaptations with a unilateral leg load. Continuous relative phase (CRP) analysis was used to evaluate limb coordination. The advantage of using phase analysis is that it can convert four variables (two positions and two velocities) into one measurement. Phase analysis can also be used to show coordination between two oscillating components [76] which can give insight into the control of a system. This makes phase analysis very useful for investigating human movement and its complexity using a reduced set of measurements or metrics. The path length of the phase portrait has been used to determine the effectiveness of the postural control system in controlling the lower body stability and steadiness similar to distance measures of center of pressure (COP) in sway balance control tests [60, 77]. To the best of our knowledge, gait adaptations due to external loads have not been compared using measures extracted from phase portraits in order to quantify gait differences.

Chapter 4: Traditional and Fuzzy Neural Network Systems for Classification and Prediction

A neural network is a data processing algorithm that consists of a variety of interconnected processing elements known as nodes. These nodes all work in conjunction with each other based on a connectionist approach to arrive at the output of the system. Each node utilizes a relatively simple mathematical processing function that is used to determine the output of the individual node. This output is then processed in conjunction with all the other nodes of the network to produce a globally complex processing algorithm. The figure below shows a conceptual diagram of a neural network with each node represented as a "black box" of sorts with some input that results in an output based on the node's mathematical governing function.

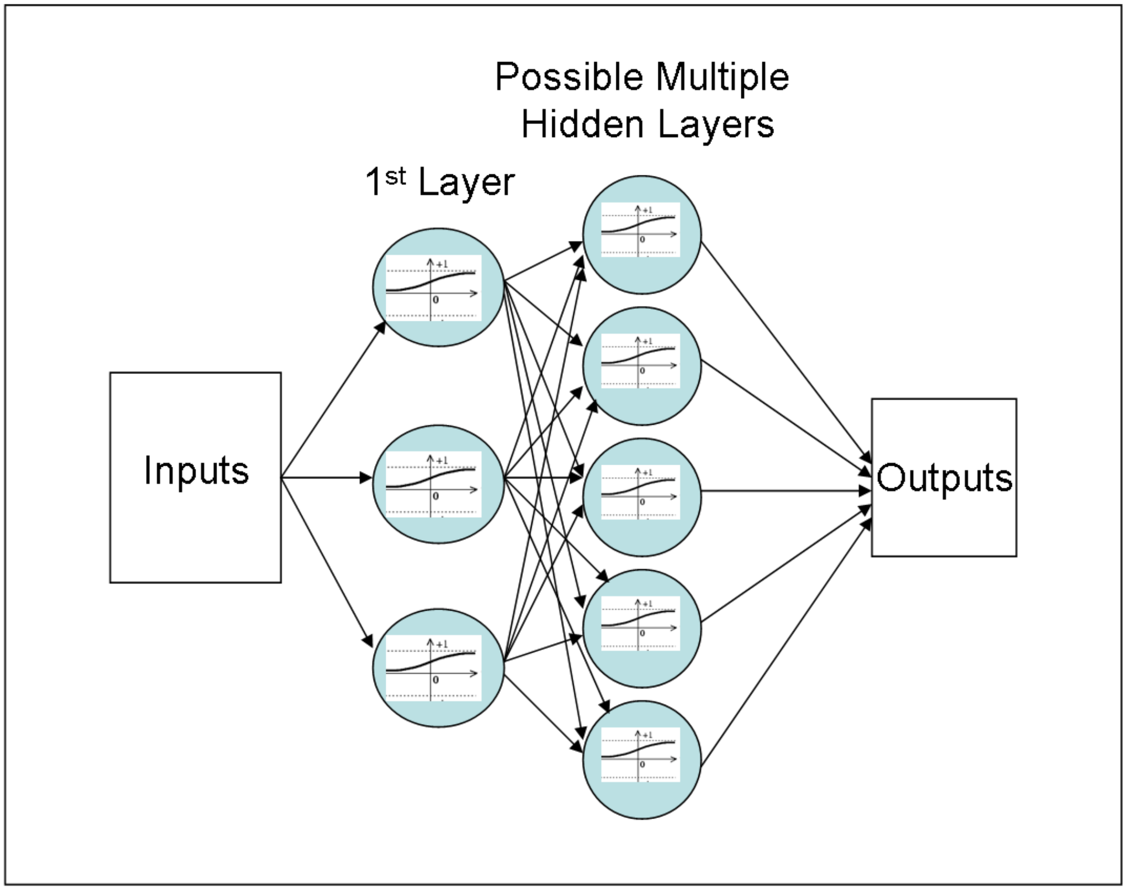


Figure 4-1: Conceptual Diagram of a Neural Network, a number of inputs are fed into the network, which then pass the data through a number of layers of differing numbers of nodes to form the final output

These networks vary in complexity from simple single-layer linear networks, which result in a linear classification algorithm, to very complex multi-layer, non-linear networks that can classify complex, non-linear systems. Additionally, these networks can be supervised or unsupervised, that is to say desired outputs can be used to train the network to give a desired response or the network can determine its own output based on statistical grouping of the outputs, for more information on neural networks see [8, 9, 10, 12]. The property that makes neural networks an attractive tool is the ability for the algorithm to learn from data. This means that a neural network has the ability to adjust its parameters in order to arrive at an optimal solution [9]. There are a number of training algorithms available which take advantage of this learning ability [89]. Each algorithm depends heavily on the type of features that are classified as well as the type of neural network being used. The most common training methods are evolutionary, simulated annealing, expectation-maximization and non-parametric methods.

4.1 Construction and Training of Traditional Static Neural Networks for Classification and Prediction

Of the various forms of neural networks mentioned above, the simplest kind of artificial neural network is a single-layer perceptron network. This type of network can be suited to a system by selecting the appropriate training algorithms and architecture. The basic perceptron consists of one layer of nodes, with each node corresponding to one input. These nodes make up the input layer of the network. The nodes each perform their specified function and are summed together. If this value is above a threshold then the output takes on activation value, if the value is below the threshold the output takes on the deactivation value. The typical perceptron network is created by setting the output values of the summed nodes between values of -1 and 1 with the threshold for the activation and deactivation states set to 0. However, these values can be set to any number of specified values. This type of algorithm is trained using the simplest training method. The individual weights used for the calculations at each of the nodes are adjusted according the difference between the network's output value and the desired output value. The difference between these values is then multiplied by the input and added to the current weight as shown in equation 4.1.

$$dw = (t - a)p^T \tag{4.1}$$

Where *dw* is the change in the value of the weight, t is the target output value, a is the actual output value and p is the input value. This is done recursively until the error falls below a certain threshold.

4.2 Fuzzy Neural Network Construction and Training

While traditional neural networks are adequate for a variety of applications, fuzzy neural networks provide an additional tool for data analysis. Fuzzy neural networks operate on fuzzy input values and provide fuzzy outputs. Fuzzy data is not sharply defined and has a membership function that determines its value.

4.2.1 Basics of Fuzzy Sets and Functions

"Fuzzy" sets are defined as a set whose boundary is not sharp. A fuzzy set, *A*, is characterized by a membership function $\mu_A(x)$ defined on **X**. **X** is a conventional set defined as $\mathbf{X} = \{x\}$. A set of ordered pairs $A = \{x, \mu_A(x)\}$, $x \in \mathbf{X}$, where $\mu_A(x)$ is the level of membership of x in *A* defined as

$$\mu_A : \mathbf{X} \rightarrow [0,1] \tag{4.2}$$

The fuzzy membership functions (FMFs) of fuzzy sets determine the membership level based on **X.** These can be of any form. The most common are triangular, trapezoidal, sigmoid and Gaussian. Graphic examples of these membership functions are given in Figure 4.3. An FMF contains a set of parameters that define the shape of the

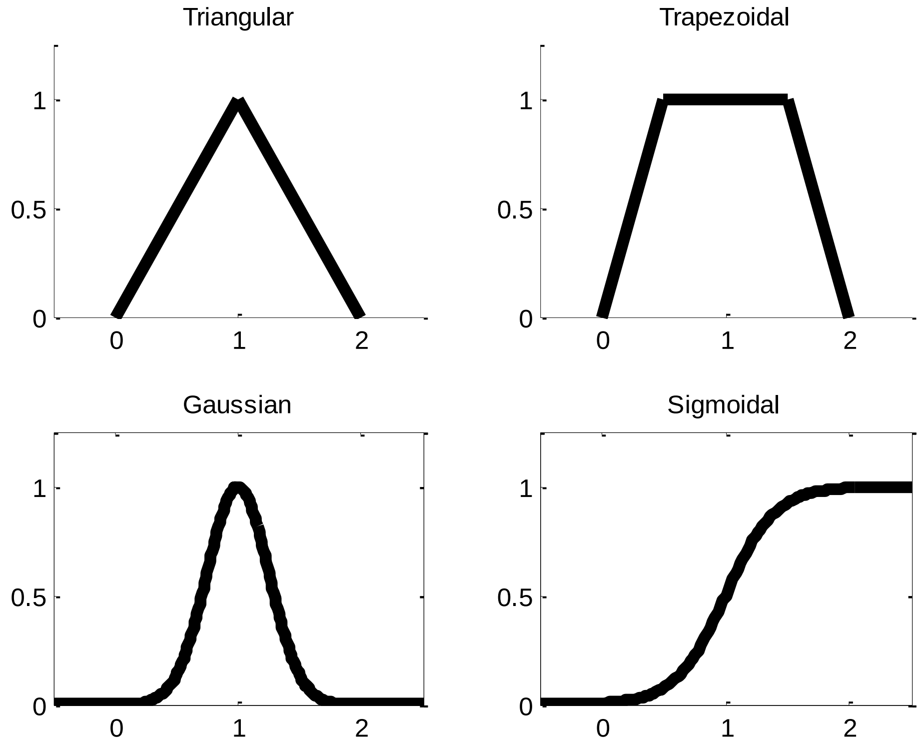


Figure 4-2: Examples of fuzzy membership functions, Axis labels

membership function. These parameters are often predetermined based on experience, knowledge or previously captured data. These values can be updated online to achieve optimal performance.

A fuzzy system will often have a basic configuration as shown in Figure 4.4. This structure contains 4 basic elements: a fuzzifier, fuzzy rule base, fuzzy inference engine and defuzzifier.

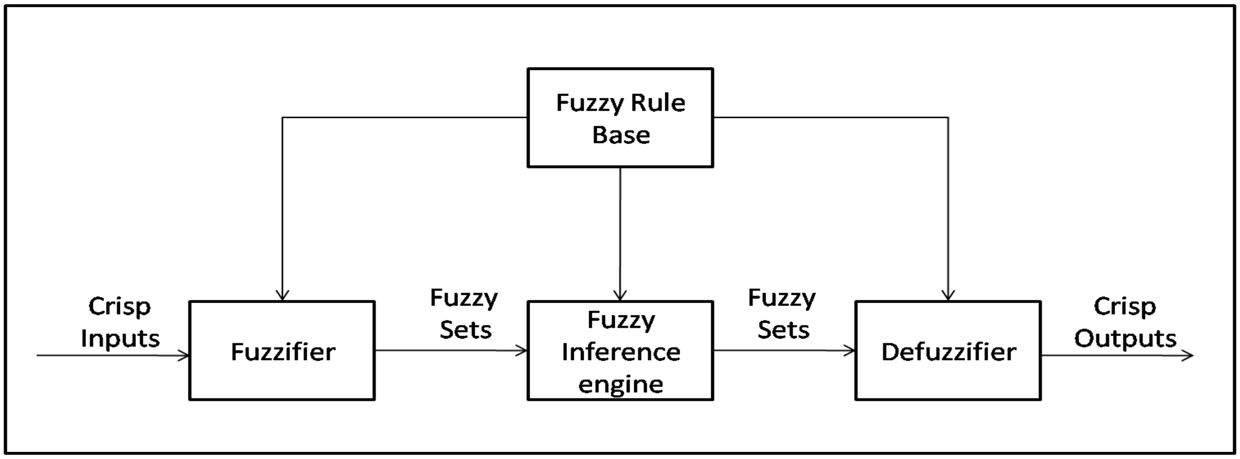


Figure 4-3: Schematic of a Fuzzy System

In this system, the fuzzifier maps the input space to the fuzzy set. The fuzzy set function is often labeled by variables such as "small", "medium", "large", etc. This results in crisp input data being converted to fuzzy values. The next element of a system, the fuzzy rule base, is a set of rules in the form of a If-Then statements. The collection of If-Then values determine the output value of the fuzzy system. A fuzzy rule base having M rules would have the following form,

$$R_j\ (j=1,2,\ldots M)\text{: } IF\ x_1\ is\ A^j_1\ ,\ AND\ x_2\ is\ A^j_2,\ AND\ ,\ldots,\ x_M\ is\ A^j_M$$

$$THEN\ y\ is\ B^j.$$

Where x_i (i=1,2,…m) are the inputs to the fuzzy system, y is the output variable and A^j_i, B^j are values determined by the fuzzy membership functions μ_A and μ_B, respectively. The fuzzy inference engine uses fuzzy rules provided by the fuzzy rule base to map the fuzzy inputs to fuzzy sets in the output space. The final element is the defuzzifier. This serves to map the fuzzy outputs to crisp values.

4.2.2 Formulation of Fuzzy Neurons and Fuzzy Neural Networks

A fuzzy neural network uses the basics of fuzzy logic and operations in order to form fuzzy neurons (FN). These neurons are used to form fuzzy neural networks trained using traditional neural network methodology. These neurons have fuzzy neural inputs and fuzzy synaptic weights found using fuzzy arithmetic. Using fuzzy arithmetic equation 4.3 gives the relationship between the fuzzy neural inputs, outputs and weights.

$$y = \sigma\left((+)_i^n w_i (\cdot) x_i\right), x_0 = 1 \tag{4.3}$$

Where (+) and $(\cdot)$ are fuzzy addition and multiplication operations respectively. Fuzzy neural inputs and synaptic weights are defined in an n-dimensional hypercube in terms of their membership functions, x_i and w_i, as shown in equation 4.4.

$$\begin{aligned} x &= [x_0\ x_1\ x_2 \dots x_n] \in [0,1]^{(n+1)}, x_0 = 1 \\ w &= [w_0\ w_1\ w_2 \dots w_n] \in [0,1]^{(n+1)} \end{aligned} \tag{4.4}$$

The weights are updated using fuzzy back propagation (FBP) learning. This procedure is similar to the back propagation scheme used in traditional static neural networks. Given a set of desired input/output pairs, ($\mathbf{x}$(k),$\mathbf{y}^d$(k)) (k=1,2,…N), the weights are optimized based on the output error. The error function is commonly defined as a summation of squared errors as shown in equation 4.5.

$$E(k) = \frac{1}{2}\sum_{j=1}^{m}\left[y_j^d(k) - y_j(k)\right]^2 = \frac{1}{2}\sum_{j=1}^{m} e_j^2(k) \tag{4.5}$$

Where the output error e_j is the error between the jth desired output and the jth neural network output and is defined in equation 4.6.

$$e_j = y_j^d - y_j \tag{4.6}$$

The weights are updated using a gradient technique similar to that of traditional back propagation learning. The updating formulations are shown in equations 4.7-4.10 below.

$$w_{il}^{(1)}(k+1) = sat\left(w_{il}^{(1)}(k) + \Delta w_{il}^{(1)}(k)\right), i = 0,1,2,\dots,p; l = 1,2,\dots,n \tag{4.7}$$

$$w_{jq}^{(2)}(k+1) = sat\left(w_{jq}^{(2)}(k) + \Delta w_{jq}^{(2)}(k)\right), j = 0,1,2,\dots,m; q = 1,2,\dots,p \tag{4.8}$$

Where

$$\Delta w_{il}^{(1)} = -\eta_1 \frac{\partial E}{\partial w_{il}^{(1)}} \tag{4.9}$$

And

$$\Delta w_{jq}^{(2)} = -\eta_2 \frac{\partial E}{\partial w_{jq}^{(2)}} \tag{4.10}$$

η_1 and η_2 are the learning rates for the weights in layer 1 and layer 2 respectively. The partial derivatives in equations 9 and 10 can be calculated using equations 4.11 and 4.12.

$$\frac{\partial u}{\partial x_i} = \begin{cases} w_n - w_n u_{n-1}, if\ i = n \\ (1 - w_n x_n)\frac{\partial u_{n-1}}{\partial x_i}, otherwise\ where\ u = u_n \end{cases} \tag{4.11}$$

$$\frac{\partial u}{\partial w_i} = \begin{cases} x_n - x_n u_{n-1}, if\ i = n \\ (1 - w_n x_n)\frac{\partial u_{n-1}}{\partial x_i}, otherwise \end{cases} \tag{4.12}$$

4.2.3 Hybrid Fuzzy Neural Network Structure and Training

A hybrid fuzzy neural network utilizes a traditional fuzzy neural network structure in conjunction with traditional neural layers [127]. For this system, statistical distributions generated by the boundary point training sets are used are the fuzzifier. The fuzzy input set contains part-based likelihoods for each of seven different body segments. The seven body segments are selected for their kinematic importance. These segments are torso, front thigh, back thigh, front shank, back shank, front foot and back foot. For each boundary point on the silhouette, the likelihood is calculated for each of the seven different kinematic families. The Gaussian likelihood function used as the fuzzifier is shown in equation 4.13.

$$p(x_i \mid L) = \frac{1}{\sqrt{2\pi}|\Sigma|^{-1/2}} \exp\left((X-\mu)^T \Sigma^{-1} (X-\mu)\right) \qquad (4.13)$$

Where L is the part family, X_i is the edge point and ($\mu_L, \Sigma_L{}^2$) are the parameters of the distribution for part family L. The parameters for each kinematic family are learned from manually labeled image sequences. A hybrid fuzzy network architecture is shown in Figure 4.5.

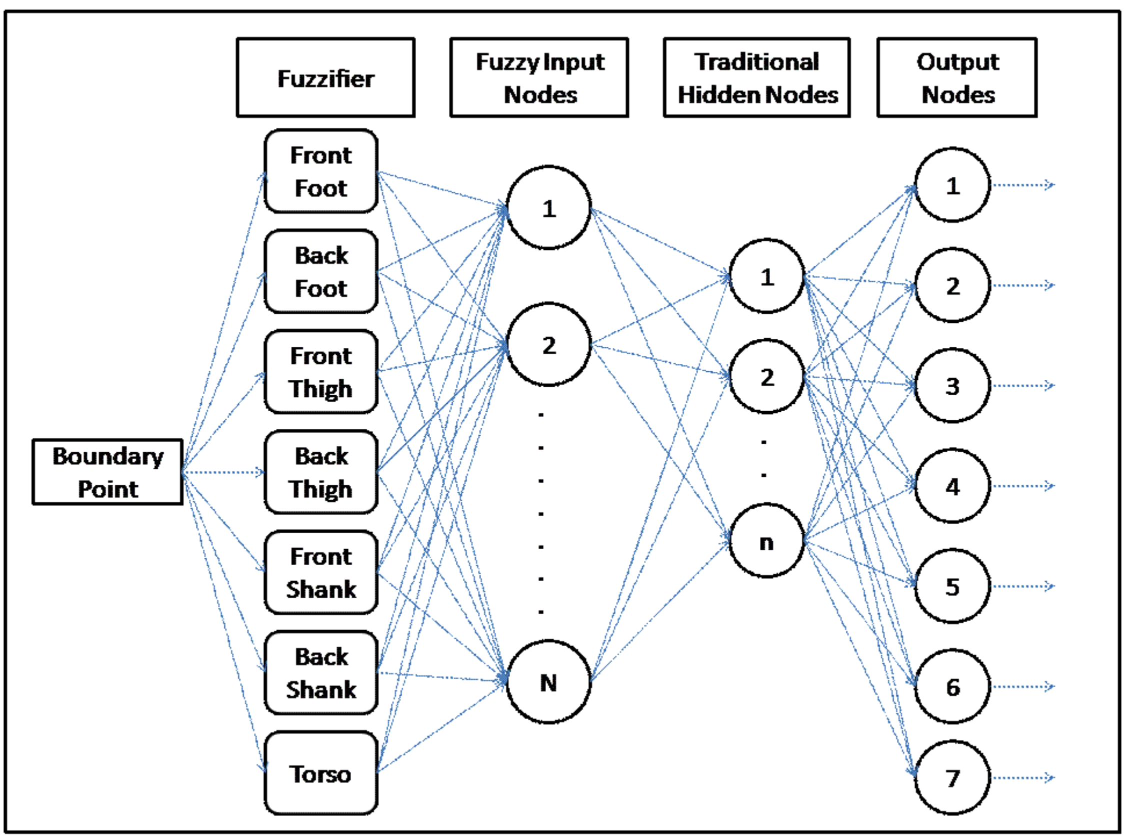


Figure 4-4: Hybrid fuzzy neural network architecture

Chapter 5: Description of Vision Operations Used in Human Visual Analysis Systems

Due to the large amount of information that is provided by image data, many methods have been developed for visual analysis. The large dimensionality of image data requires the use of many sophisticated methods. Many of these are based on features, such as edges, motion fields, such as optical flow, and change detection, such as background subtraction.

5.1 Image Sequence Motion Field Calculation

Moving objects exist almost everywhere within our world. The nature of the motion that is occurring in a scene provides a vast amount of information. A measure of the motion found within an image sequence is known as optical flow [2]. Optical flow is the measure of three-dimensional motion found within a scene projected to the two-dimensional image plane. Optical flow is calculated using consecutive frames in a video [3]. Specifically, this technique yields a velocity field that measures motion taking place in a camera's two-dimensional image plane. This is used in a wide variety of tasks such as video compression, scene analysis, navigation, and three-dimensional scene reconstruction [4]. This has been a highly researched area in computer vision for more than two decades [49]. The earliest research was done by Horn and Schunk [2]. This paper was the most influential research on optical flow computation. It provided the first real definition of what optical flow was and gave direction on its calculation. The paper defined optical flow as a "velocity field in an image which would provide a transformation for one image that would yield the next image in an image sequence" [2]. This definition provided the ground work for all optical flow research that would take place from that point on. Many different methods have been developed that attempt to estimate the optical flow in an image and continues to be a highly researched topic. Accurate optical flow computation provides a rich source of information regarding the motion taking place in a scene. This has direct application in the analysis of human gait.

As described above, there have been many different methods proposed to calculate optical flow. All of these techniques take into account a number of constraints imposed on the image so that the image motion can be found within a sequence. The first of these assumptions is that the intensity values found within an image change only due to the motion contained within the scene [2]. This embodies itself in the equation of optical flow found in [2] and shown in equation 5.1.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + 1) \tag{5.1}$$

Where I(x,y,t) is the image intensity found at point (x,y) at time t, and $I(x+\delta x, y+\delta y, t+1)$ is the intensity at time t+1 found at point $(x+\delta x, y+\delta y)$, with δx is the displacement in x and δy is the displacement in y of the pixel. As can be seen in the above equation, a change in intensity at any point in the image would only be due to the movement of that pixel from one location in the scene to another location occurring in the time step of 1. This assumption is often violated in real situations due to lighting changes, variations due to noise, etc. Other assumptions that are often applied are related to temporal and spatial changes. Specifically that the motion found within a scene varies smoothly and gradually. This means that if an object is moving within a scene neither will it move erratically nor will it move rapidly from one side of the image to another [3]. The two predominant optical flow calculation methods either use differential or area matching techniques. Differential techniques use the spatiotemporal derivatives of image intensity to calculate the motion taking place in the vertical and horizontal directions. Area based techniques match small areas of an image from one frame to another to esitmate the motion taking place in the video. Block matching is the most popular area based motion estimation technique.

Block matching algorithms attempt to estimate optical flow by breaking the image into many small sections, or blocks. Each of these blocks is used to find where in the previous frame this image block was located [4]. This location information is used to estimate the motion for that block. This technique allows the motion estimation of the sequence to be robust to violations of some common optical flow assumptions.

Specifically, they allow for the violation of spatiotemporal consistency and constant intensity. Since block matching techniques do not retain motion information beyond the current image frame, the temporal consistency assumption is not imposed. The spatial consistency requirement can be tightened or relaxed by defining the search area of the block searching algorithm. The constant intensity assumption may be violated because a block of pixels is being used to search from one image to the next in the sequence. This means that an overall correlation measure can be used to match the blocks of pixels instead of searching the image for a specific intensity value. This makes block matching algorithms more accurate for real image data. These properties make block matching algorithms an ideal search technique that accounts for the errors associated with gait analysis.

Since the early development of block matching algorithms, many techniques have been proposed to increase the accuracy and the efficiency of the system. Since the middle of the 1980's block matching algorithms have been developed to estimate the motion found within a scene. The first step in the development of these algorithms is defining a cost function that measures the match of a current block to a block found in the search pattern. Cost functions vary in computational complexity and accuracy. The two most popular cost functions [3] are the mean absolute difference as given in equation 5.2.

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left| C_{ij} - R_{ij} \right| \tag{5.2}$$

And the mean squared error as shown in equation 5.3.

$$MSE = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(C_{ij} - R_{ij}\right)^2 \qquad (5.3)$$

Where N is the length of one side of the block, C_{ij} and R_{ij} are the pixels being compared within the current and reference block. An error based cost function is more robust then a statistical matching scheme due to the fact that a block can contain different parts and orientations of an object from one image frame to the next. An example of this would be the leg movement of a subject while walking, a block containing a portion of the leg and background in one frame, will most certainly contain a different part of the leg in the next frame when the leg rotates around the hip. This is illustrated in Figure 5.1. This will cause errors at edge boundaries of objects that are deformable or rotating. These errors can result in statistical block matching errors.

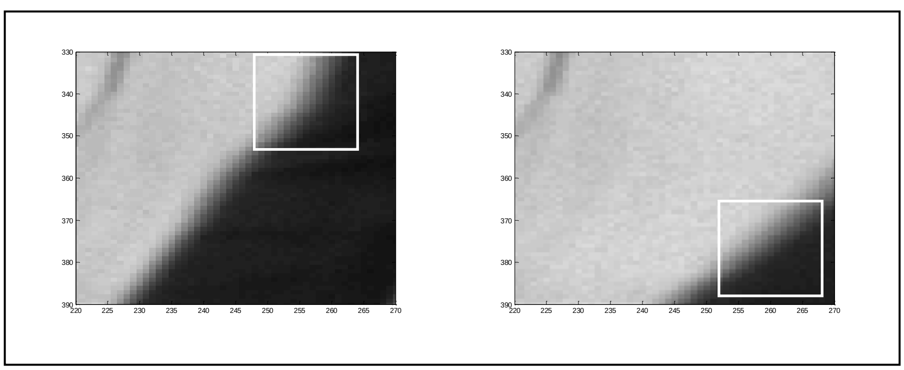


Figure 5-1: Example of block matching errors at the boundary of a rotating object

A statistical match of the block from one frame to the next would not be able to account for this rotational movement because the statistical matching criteria accounts for noise in the image and not errors due rotation and deformation. Because of this, an error term cost function is found to be more effective when used on scenes containing With a cost function defined, the next step in the development of a block matching motion estimation algorithm is the definition of the block size and search criteria.

The block size is typically tailored to the specific application for the motion estimation scheme. Larger blocks yield more accurate motion estimations for large objects moving within the scene, but have poor motion resolution. Smaller blocks yield denser motion estimation fields, but can be inaccurate if the size of the block is too small. For most applications, a block size of 16x16 is used [4] to provide a large enough sample of pixels to reliably match blocks and provide a sufficiently dense motion field. However, the block size must be tailored to the specific application, hardware and algorithm. In addition to block size, a search window size must be defined. This is application and algorithm dependent. A larger search window increases the computational cost of the algorithm while allowing for larger motions, but can lead to violations of the spatial consistence assumption. A smaller search window size reduces the computation time. However, motions larger than the search window can not be estimated. The search window parameter must be selected so that it is large enough to calculate desired motions, but not so large as to provide unnecessary computational burden. The next aspect of block matching algorithms is the block searching technique.

The first and most basic of these algorithms is the Exhaustive Search algorithm. As the name implies this is the simplest and most accurate of the block matching algorithms. This technique involves calculating the cost of a block of the reference image at all points in the search window of the current image (next image in the sequence). This yields the best possible match that a block in the reference image can obtain using the given cost function. Beyond the exhaustive search method, many different algorithms have been developed that search specific points within a search area instead of all possible points. These algorithms are more efficient due to the reduced number of search points, but can introduce a large amount of error in the final motion estimation. The most popular search algorithms developed during the last two decades are Three Step Search, New Three Step Search, Simple and Efficient Search, Four Step Search, Diamond Search, Adaptive Rood Pattern Search, the Cross Diamond Search [4]. Each of these search methods attempt to maximize computational efficiency while minimizing error. The only algorithm discussed here in detail is the Adaptive Rood Pattern Search. Details on the various additional search methods can be found in [4] and [5]. The Adaptive Rood Pattern Search takes advantage of the fact that the majority of motion taking place within an image sequence is usually coherent from block to block. That is to say that if one block moves to a location it is highly likely that a block located close to it moves the same way. This search algorithm uses the motion of the block located to its left to predict the motion of the current block. This predicted motion location is the first point searched by the algorithm, after searching this point, a rood pattern of distributed points is searched [4] as shown below in Figure 5.2.

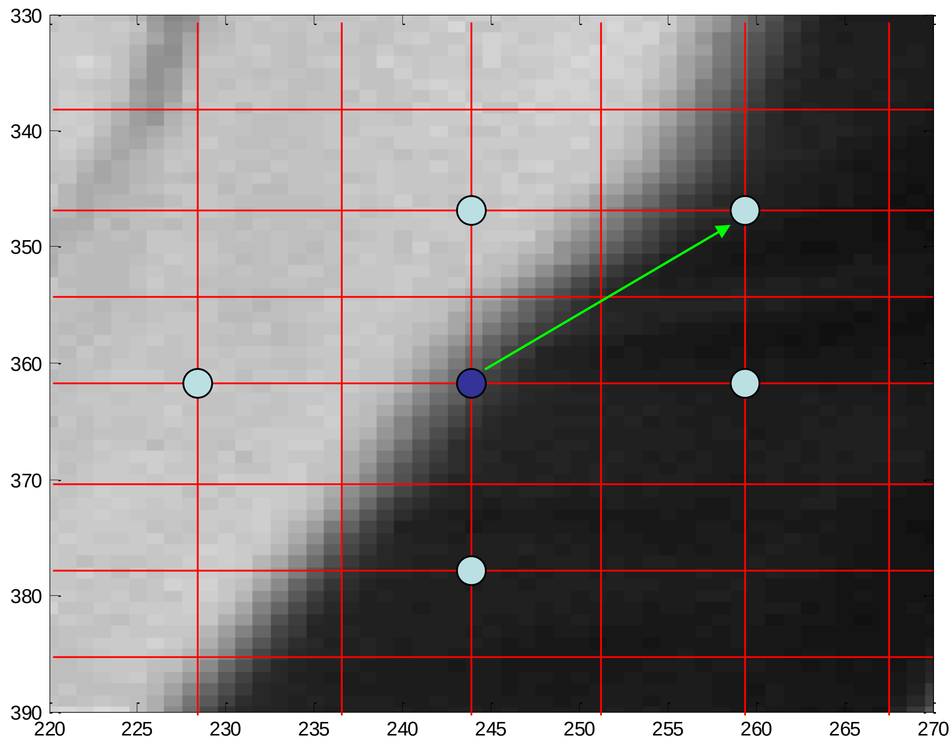


Figure 5-2: The search pattern for the Adaptive Rood Pattern Search for a block size of 8x8 pixels (block boundaries shown as red lines), where the green arrow is the previous block's motion vector, the step size is the maximum value of the motion vector in any one direction (16 pixels in this case), the blue point is the current block location in the previous frame and the light blue points are search pattern points

The step size of this search is determined using equation 5.4.

$$S = Max(|X|,|Y|) \tag{5.4}$$

Where X is the x-component of the predicted motion and Y is the y-component of the motion vector. Upon searching each of the five points defined above, the point with the best match is set as the origin for the subsequent search operations [4]. The next operation is a small diamond search around the point chosen above. This is not done if the point chosen above is found to be the current location of the block. The small diamond search pattern is shown in Figure 5.3.

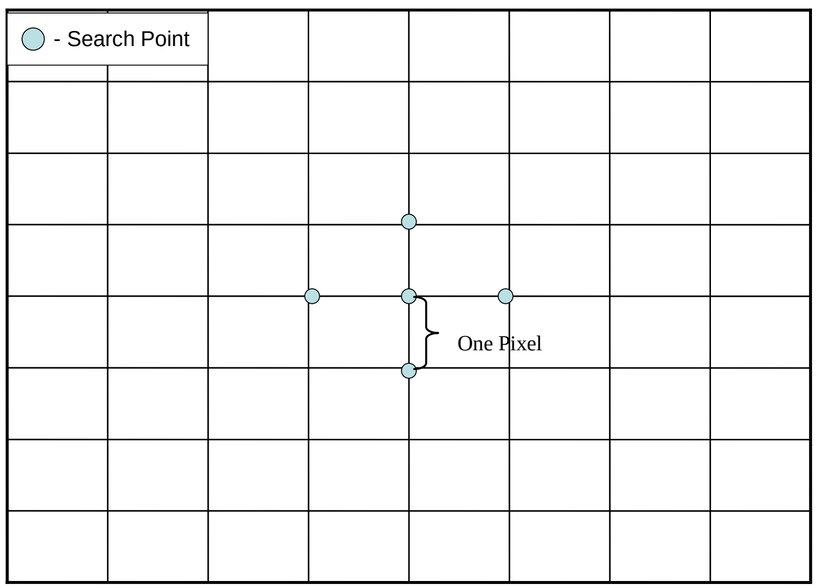


Figure 5-3: The Small Diamond Search Pattern

This search pattern finds the point that yields the lowest value of the chosen cost function and sets this point as the origin for the next small diamond search. This process is repeated until the point yielding the lowest cost is found to be at the center of the small diamond search pattern. This is the most efficient of the search algorithms because it takes advantages of information found in the adjacent block motion in order to calculate the motion of the currently searched block.

5.2 Image Background Subtraction and Subject Silhouette Generation

In addition to temporal data, spatial information provides a large amount of information for human gait analysis. Spatial analysis comprises a vast majority of the computer vision field and includes image filtering, edge detection, background determination, and shape analysis [3]. Spatial image data is used in many different gait analysis applications [6]. For many years spatial data has been used to develop biometric gait identification systems [6]. The main types are background extraction and shape analysis of binary silhouettes. To find a silhouette of the subject, background extraction is done to separate the complex background from foreground objects.

Background extraction consists of removing any objects not of interest in the scene from an image. In human gait analysis, the foreground objects of interest are human subjects and the background is defined as anything which is not part of or attached to the subject. Background can consist of any number of stationary or moving objects which all have a wide variety of intensity values. An example of a complex scene can be found below in Figure 5.4.

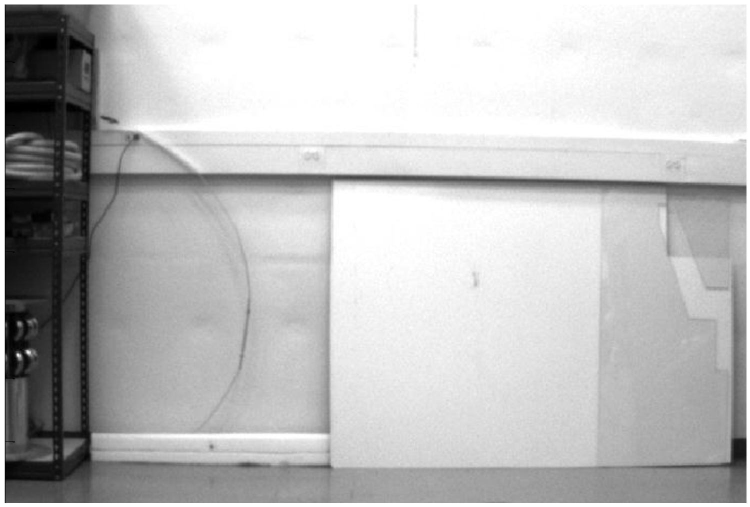


Figure 5-4: A background containing different objects

Background determination is typically accomplished by comparing an image to a model of the background. This is done in a number of ways. The simplest of these is to use a single frame where only the background is present and subtract this frame from the current image. This leads to a subtraction that does not adapt to background changes or noise in the image. This results in a high number of false positives. An alternative approach is the use of statistical background models. An average of background frames is used to create a model of the background that accounts for image noise. While this model provides robustness to noise, it still does not allow for background changes and results in a high number of false positives. Another type of statistical background model is one that constantly updates the background using incoming images. While these methods allow for global image changes, such as those due to lighting, they are not robust to small movements in the background, such as a tree branch blowing in the wind. A background extraction method that uses non-parametric statistics can create models that account for all of the above errors [13]. These models allow for adaptation to small background movements, such as those found in an outdoor scene, and provide a more accurate method of comparison between a given image and a background model.

The majority of testing in this work used a static background. However, a background model that can account for changes, such as lighting and small background movements, is very important when utilizing a system in uncontrolled environments outside of a laboratory. Robustness in a variety of applications and situations is why statistical background determination schemes are so desirable. The probability that a given pixel is part of the background can be accurately calculated by these models. This probability is then compared to a threshold value. If the calculated probability is below the threshold, then the pixel is determined to be part of the foreground. If the value is above the threshold, the pixel is part of the background.

5.3 Spatial Models for Human Activity Analysis

Spatial model based systems rely on a graphical representation of a human subject that is fitted to observed images. Graphical models used in these systems vary from simple "cardboard" models that represent each body part as a simple shape to more complex systems, which model the shape of different body parts and allow for deformation throughout the course of locomotion. Once created, these models are fit to an observed image. An image is searched for each pose of the model and the pose that results in the minimum error is the configuration of the observed subject. There are two main features that are used to fit a model to an observed image. These features are edges and image texture. In the case of edge features, the model consists of an array of edges which represent the human body as shown in Figure 5.5.

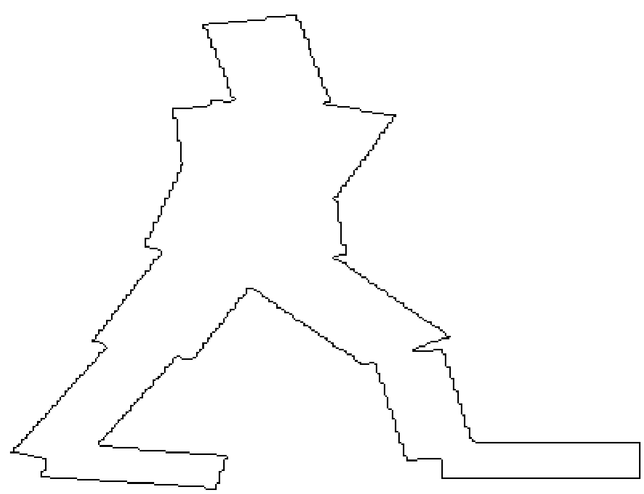


Figure 5-5: Edge based human model

Models that use image texture are more complex, they rely on colors and patterns of the image to fit the model. These can be based on previous images in the sequence, a texture normalized image or the by an appearance model. The most commonly used example of texture based models is those that match regions of exposed skin to locations that are assumed to be uncovered (ie. a subject's face). An advantage of texture model based systems is that background extraction is rarely performed on the input images. Because of this, these systems are often used in order to indentify human subjects located within a scene as well as the pose of a subjects. More information on human detection in images can be found in [6, 27, 92].

Many systems use frame to frame information to further increase the accuracy of model fitting. There area many types of temporal models. The majority of these models focus on a sequence of poses that take place during gait motion. The sequence is learned from exemplar or model data and used to create a system that predicts the next pose given the current pose. Many temporal models use Hidden Markov Models (HMMs) or joint trajectories to predict the changes in a subject's pose during locomotion.

When combined, spatial models used to find a subject in the frame and temporal models used to predict the pose in future frames is an efficient way of estimating a subject's joint kinematics. Many of these systems use a spatial model to find a "key" frame in an image sequence. This usually corresponds to an easily recognizable part of the gait sequence, such as the point where the feet are furthest apart. Once this key frame is identified a temporal model is used to generate the pose and of the subject in the next frame. The spatial model can then be further fit the image. These types of systems attempt to capitalize on the strengths of both methods to maximize the accuracy and robustness of the system while minimizing the computational time required.

5.3.1 Description of the Major Types of Spatial Models Used in Human Visual Analysis

As mentioned in the previous section, spatial models attempt to fit previously defined models of human appearance to observed image data. This can be accomplished using a variety of methods. The main types of models used in current research are cardboard, silhouette and part based models. Each of these models attempt to create a simplified representation of the human body to easily and accurately find the location of key body parts in an image. Of the three main types of models, the simplest is the cardboard model. It is referred to by this name because it represents each body part as a simple shape such as a rectangle or trapezoid. An example of a cardboard model is shown in Figure 5.6.

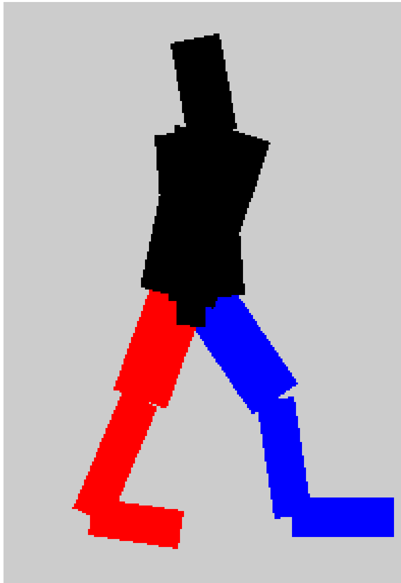


Figure 5-6: Cardboard human model

A slightly more complex model is a silhouette model. This model is built using either exemplar silhouettes or from synthetic data generated by motion capture. These models provide a better fit to image data at the cost of more complex model generation. These models need a large number of exemplars to accurately determine a silhouette's change in shape throughout an image sequence. The increase in data required to accurately model an observed subject is much greater than that of a simple cardboard model. An example of a silhouette model is shown in Figure 5.6.

Figure 5-7: Sequence of a silhouette model

Part based models vary from a simple collection of cardboard model parts to extremely complex deformable models that are governed by physical constraints of human physiology. These models are fit to specific areas in an image that correspond to a specific body part. These can either be fit based on image location or the location of previously fit portions of the model.

5.3.2 Methods Used to Fit Spatial Models to Image Data

Once the model type has been chosen, the model must be fit to observed images. Many different model fitting techniques have been used. The most commonly used techniques are based on edges or texture regions. The types of fitting systems are determined by the type of model used and the desired search space. Typically, the model is placed at a region in the image and the degree of match for that region is calculated. This is repeated for each region of the image and the region that yields the best match is the area in the image that contains the model. A popular method used for edge model matching is the chamfer distance transform (CDT) [264]. In this technique the edge image is transformed to an edge energy image. This is an image in which each pixel location takes a value that is its distance to the closest edge. An example of an edge energy image is given in Figure 5.8.

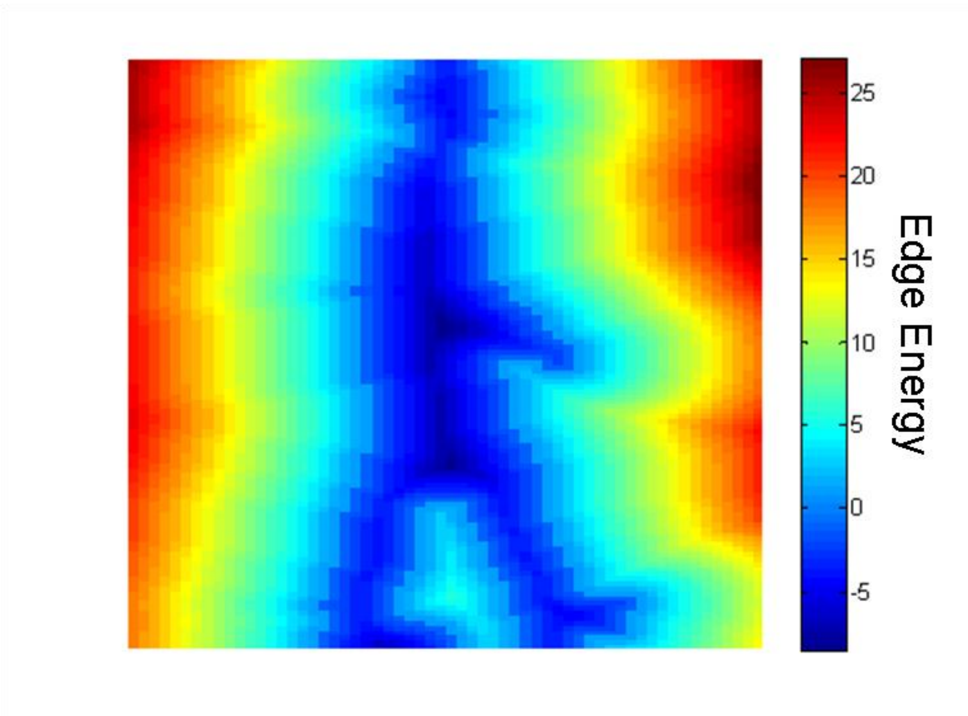


Figure 5-8: Edge energy image

Next, the model is placed on the edge energy image and the sum of the energy at each model point is used to determine the level of fit. This is repeated for each position in the image. Another method that is used to fit a model to an image is texture fitting. This is used with color or monochrome cameras. These systems use a known appearance model to find its location in an image. While this can result in better accuracy than edge based methods, the appearance of individuals outside of a controlled setting is often unpredictable and difficult to model. These systems can be used in conjunction with an edge based system to reduce the error of the model fitting.

5.3.3 Temporal Based Methods for Fitting Human Spatial Models to Image Data

Currently many model based method use temporal fitting methods in addition to spatial model fitting. These are used for tracking as well as improving the accuracy of the pose estimation through image noise and subject occlusion. The most popular techniques used for incorporating temporal restraints onto spatial model fitting methods are Hidden Markov Models (HMM) [78] and statistical models [164]. Each of these methods attempts to fit the spatial model to consecutive images or fit the spatial model to the entire gait sequence at one time.

Hidden Markov Models can be used to fit a spatial human model to consecutive frames by assuming that in each frame the observed configuration of the subject is a "state" in the HMM. These "states" are determined *a priori* based on the human gait cycle. For each frame, the previous configuration is used as the previous state and the pose in the current frame is then estimated used the state transition probabilities of the HMM. A visual example of this process is shown in Figure 5.8.

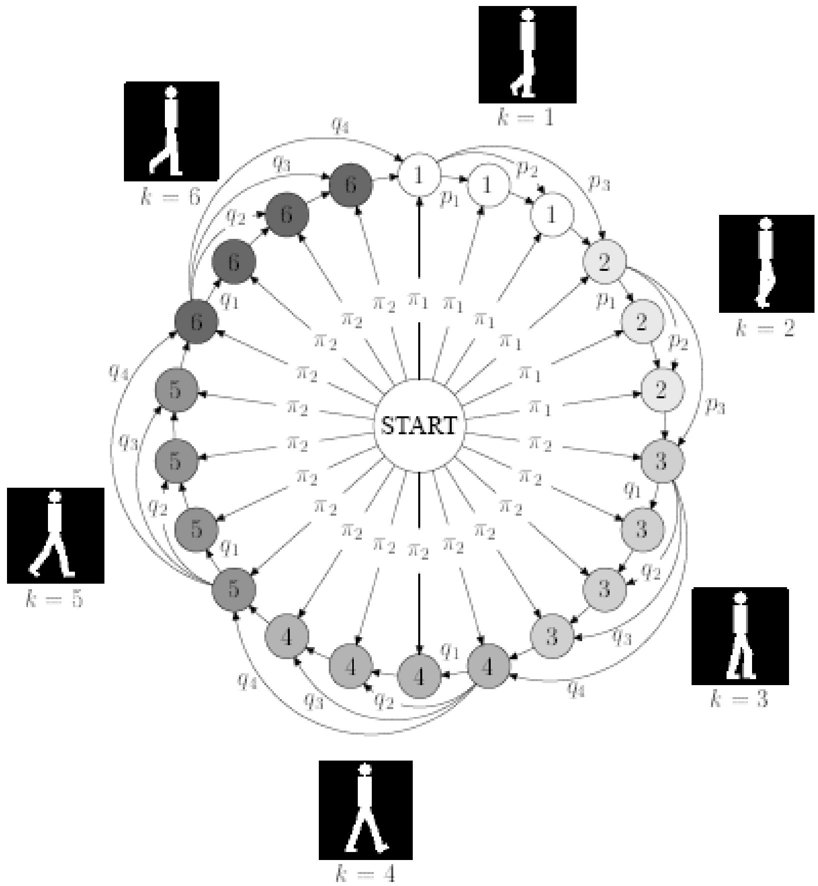


Figure 5-9: Hidden Markov Model for gait sequence, figure taken [78] Z. Zhou et al. 2006

As opposed to the HMM approach of estimating the pose for the next frame based on the current frame, a statistical approach attempts to fit the model to an entire section of the image sequence simultaneously. This involves estimating the model positions at each frame in the sequence by maximizing the likelihood, or posterior, of the statistical model over the entire section of the sequence. This is accomplished by finding which group of model terms yields the highest probability according to the chosen statistical model. An example of such a statistical model is given in equation 5.5. The

$$P(X,Y|I) \propto P(I|X,Y)P(X|Y)P(Y) \tag{5.5}$$

terms of the model are often learned from exemplar data. Additionally, a Bayesian approach can incorporate a prior distribution, based on previous experience or expert opinion that can further improve the accuracy of the model.

5.3.4 The Limitations of Spatial Model Based Human Activity Analysis Systems

Spatial model based systems are the most popular type of system used for human visual analysis but they have some serious limitations. The first of these is that spatial model matching is often inaccurate due to the presence of edge noise, missing parts of a contour and/or a high dimensional search space. These errors are often enough to make pose estimations inaccurate. To compensate for these errors, temporal based modeling systems are used in conjunction with traditional spatial modeling systems. Although these algorithms greatly improve performance, they are limited in that they can only analyze normal human movement and often times force all motion to fit into a "normal" walking model. These models limit the capability of such systems to distinguish between speed differences, motion differences and stride to stride gait deviations.

5.4 Temporally Based Systems for Human Activity Analysis

While spatial model based systems might rely partially on temporal information, temporally based human pose estimation systems rely solely on time information to estimate the pose of an observed subject. These methods fit the entire motion to a known manifold or trajectory model.

5.4.1 Manifold Based Methods for Activity Recognition, Viewpoint and Pose Estimation

Manifold methods rely on dimensionality reduction techniques to break down an image sequence of human movement into a low-dimensional representation. These methods operate under the assumption that human gait lies on a one-dimensional manifold that is embedded in a three dimensional volume of high dimensionality [79]. These systems are used for view determination [235], view synthesis [79, 235] and human identification [79]. These systems utilize a number of non-linear dimensionality reduction techniques to obtain a manifold. These techniques take an image sequence, $I \in \{n \times m \times F\}$, and reduce it to a manifold of size $M\{d \times F\}$, where F is the number of images in a sequence, n x m is the size of each individual image and d is the desired size of the manifold space. These methods attempt to optimally determine a function that best maps the information contained in the image sequence to the desired size of the manifold space. The motion manifold is then used to classify observed motions. The primary methods of dimensionality reduction are locally linear embedding (LLE) [267], Probabilistic Principle Component Analysis (PPCA) [269] and Gaussian Process Latent Variable Models (GPLVM) [268].

5.4.2 Kinematic Trajectory Methods for Human Pose Estimation and Tracking

Trajectory methods attempt to fit motion information obtained from an image sequence to trajectories that are created by certain parts (joints, arms, etc.) of an observed subject's body. The trajectory information is obtained either through exemplar data or through a kinematic model. The trajectory information for the image sequence is used to find the pose of the observed subject in each individual image frame. These types of systems require that the image motion field, or optical flow, be calculated prior to analysis.

5.4.3 Problems Associated with Temporally Based Human Analysis Systems

Many motion based methods estimate a subject's pose in limited situations to an acceptable degree. However, like many model based human analysis methods the observed motion must be the one that is modeled by the system. This means that if a subject is running and the system does not have a model for running motions, the system will fit the observed motion to a known model. This limits the ability to analyze unpredictable motion. Additionally, the calculation of image motion is often time consuming and prone to errors. Another issue with temporal based methods is that they must be initialized before analysis. This means that all motion must be analyzed from a specific starting and end point. An example of this for gait would be left heel contact to left heel contact. A problem unique to manifold methods is that they do not generalize to novel image sequences. This makes them incredibly sensitive to image changes such as lighting, shifts, rotation and scale.

Chapter 6: Abnormal Gait Detection and Classification Systems

A system that can determine whether an individual is walking normally or abnormally is a valuable first step in many high level gait recognition and classification systems. This information can be used to quickly and accurately determine whether a subject is walking normally and can further be analyzed by the system. Such a preliminary analysis system consists of five stages:

1) Acquisition phase
2) Background Processing
3) Foreground Operations
4) Motion Determination
5) Analysis

Each of these stages is discussed in detail in the following sections.

6.1 Image Acquisition

The acquisition phase consists of taking images of the scene and processing them so that they may be analyzed. For this system, an Imaging Source DMK21AF04

CCD camera is configured to collect video at 30 frames per second with a resolution of 640x480 pixels per frame. A manually adjusted lens in used for focusing and zooming of the camera to maintain versatility. A manually adjusted lens is chosen instead of an automatic focus lens to preserve frame to frame consistency. This is important in any video analysis system as automatic zooming and focusing can change aspect ratios and cause large differences in measurement. To connect the camera to the video acquisition system an IEEE-1394, Firewire, interface is used. The video acquisition system used is a standard desktop PC, with Matlab's video acquisition toolbox used to interface the computer with the camera. Matlab is used to acquire and process video for all levels of this system.

6.2 Image Foreground Determination and Subject Silhouette Generation

Foreground determination is the first step in the processing of video data for this gait classification system. To determine which parts of the video are foreground and which are background, a preliminary background determination step is completed. This is done using a non-parametric background extraction scheme similar to that discussed above in section 2 of chapter 5. The background is modeled using a selection of background frames that do not contain subjects. This is done because a limited number of frames are taken, roughly 100 per subject. This is an insufficient number of frames for online background extraction. Due to this limitation, the background extraction method used accounts only for the noise present due to the video acquisition system and environmental changes. Slight modification of the algorithm can be used to adapt the system to larger outdoor scenes where such limitations are not practical. The first step in this background extraction algorithm is determining which frames should be used to build the background model. For this system the first 30 frames of every video sequence are used because they do not contain a subject. The standard deviation of each individual pixel is found using the training frames. The standard deviation for the overall image is found by using equation 6.1.

$$\sigma_{i,j}{}^{2} = \left(\frac{1}{N-1} \sum_{f=1}^{N} x_f(i,j) - \bar{x}(i,j) \right)^2 \qquad (6.1)$$

Where

$\sigma_{i,j}^2$ - variance of pixel (i,j)

N – total number of frames used for standard deviation calculation

f – the current frame number

$\bar{x}$ - mean pixel value over N frames at pixel (i,j)

The pixel variance is then used to find the overall background variance.

$$\sigma = \sqrt{\frac{1}{c*r}\sum_{i=1}^{c}\sum_{j=1}^{r}\sigma_{i,j}^{2}} \tag{6.2}$$

Where

σ - pixel wise average image standard deviation

c – number of pixels per column in the image

r – number of pixels per row in the image

With the pixel wise standard deviation for the video known, a test to accurately determine the portions of an image that are statistically different from the background is performed. The test that is used in this implementation is taken from [6] and is shown in equation 6.3.

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x_t - x_i)^2}{\sigma^2}} \tag{6.3}$$

Where

$\Pr(x_t)$ - probability that pixel t is part of the background

N – number of frames used for the background extraction

σ^2 - pixel wise image variance

x_t - intensity value of pixel t

x_i - intensity value of pixel in background frame i

The pixel is determined to be foreground if the probability value is below the threshold value, *th*. This threshold value is adjusted depending on the video data being processed, for this system a threshold value of 5 is chosen based on experiment. The effect of different thresholds on the number of false positives is shown in Figure 6.1.

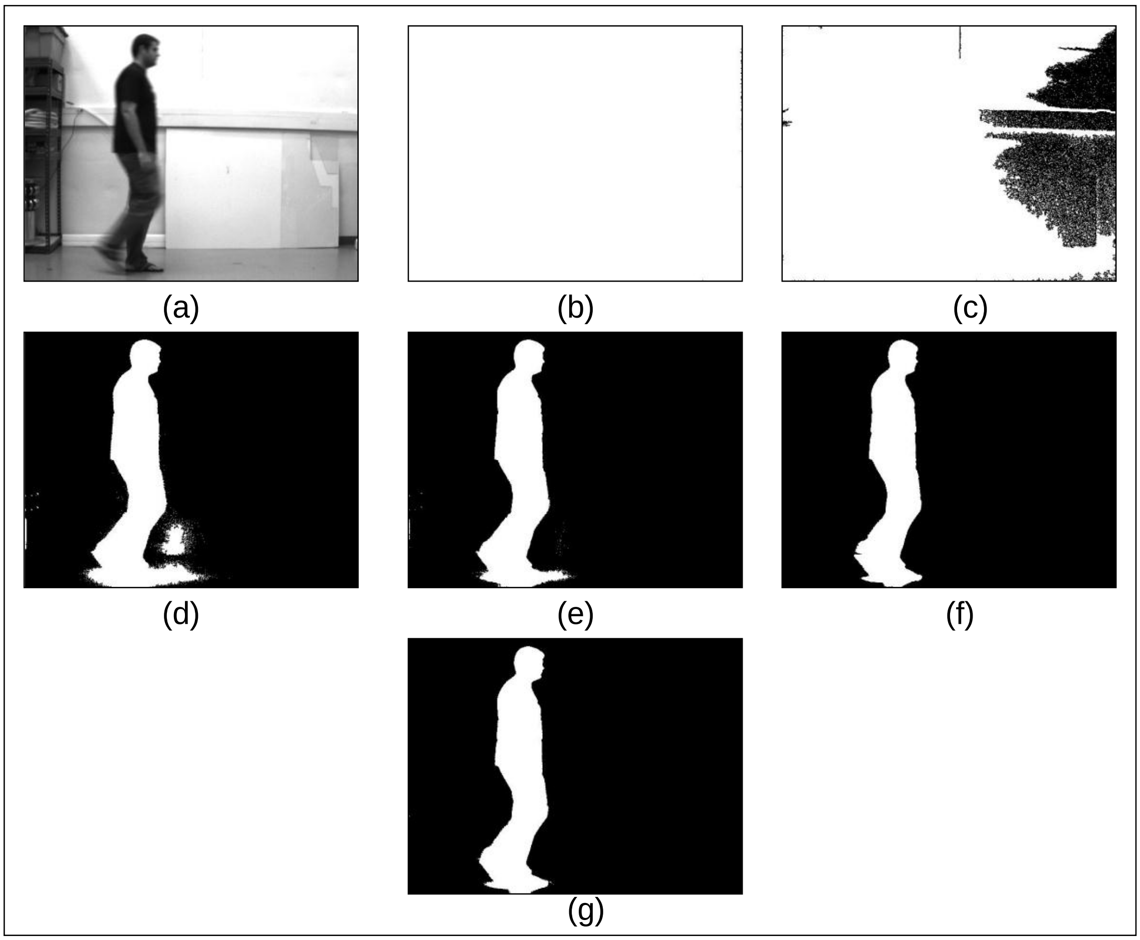


Figure 6-1: Results of the non-parametric background extraction algorithm, (a) shows the original image while the remaining images are results of the background extraction

algorithm with the threshold set at (b) 0.05, (c) 0.10, (d) 0.50, (e) 1.00, (f) 2.50, and (g) 5

There is a distinct difference in the output of the algorithm based on the threshold value. This threshold value does not relate to the traditional values of statistical probability which rely on an assumption of the normal distribution. Because of this and the figures shown above, the threshold of 5 iss chosen to minimize the amount of false positives and increase the accuracy of the foreground detection algorithm.

Once the background has successfully been extracted from the input frame, additional steps are needed to determine the foreground regions. While this step may seem unnecessary, often binary information given from background extraction can contain false positives and false negatives. These errors are overcome using the following steps. First, the pixels within the foreground are set to a value of 1 and all others are set to a value of 0. This creates the initial binary silhouette. Next, binary dilation and fill operations are performed to compensate for false negatives. Dilation is the process of gradual enlargement of an image that serves to fill small gaps in a silhouette. This is accomplished by placing a structuring element over each of the pixels in the image. If any foreground pixels in the image correspond to those of the structuring element then that pixel is set to a value of 1. The structuring element used for this operation is shown in figure 6.2.

1
1
1
1
1

Figure 6-2: The 5x1 structuring element used for dilation operations

More information regarding dilation operations on binary images can be found in [9]. The next operation is the binary filling operation. This process involves “filling” in all regions that are completely surrounded by the silhouette. This serves to make the silhouette more coherent. An example of the fill and dilation operation being performed is shown in figure 6.3.

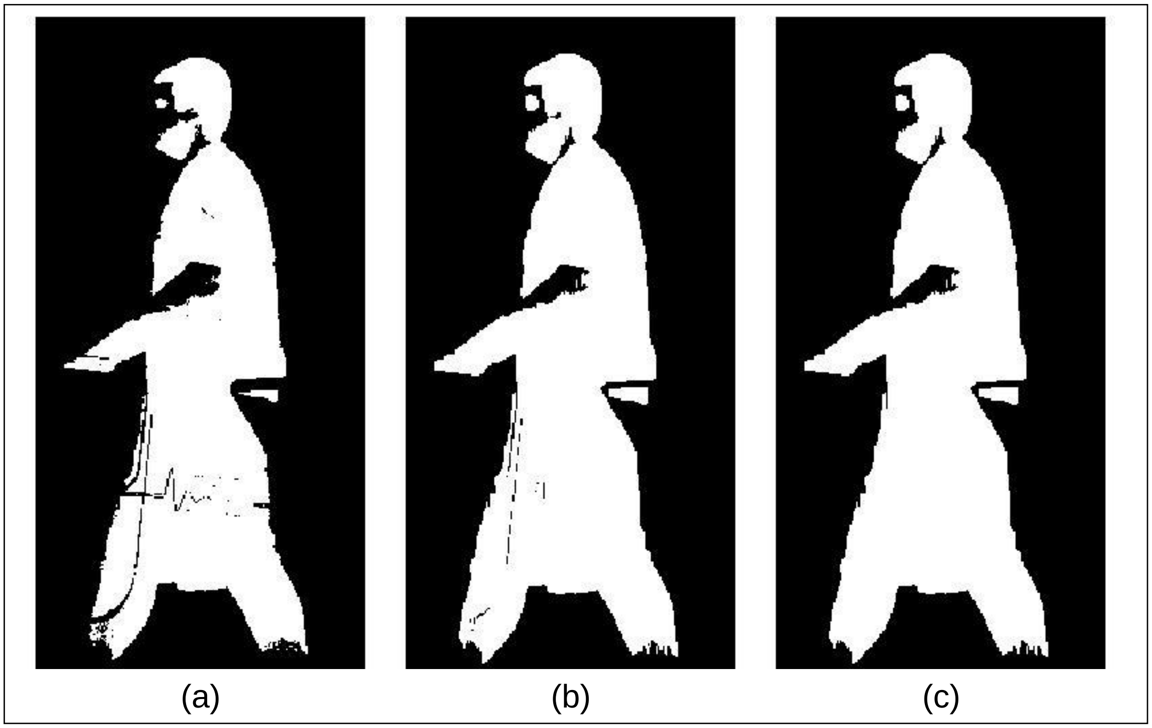


Figure 6-3: An example of a binary silhouette generate for use with the gait classification system, (a) is the binary silhouette after background extraction and application of a binary threshold, (b) shows the same silhouette after performing a dilation operation (c) is after performing a fill and dilation operation

6.3 Joint Angle Estimations from Silhouette

With a reliable silhouette of the subject established, the next stage of the system processes the binary data to generate features. The first is the centroid of the binary silhouette. This provides information on how the overall subject is moving in time as well as an approximate location of the hip. This is used to determine where other parts of the body lie in the binary image. Traditionally the centroid of an object is found using the following equation.

$$\frac{\int xf(x)dx}{\int f(x)dx} \tag{6.4}$$

Approximations of the centroid are used in this system to reduce the computational complexity of the algorithm. The centroid of the silhouette is approximated using two different methods, one uses the average of all the boundary locations and the other calculates the centroid using the entire area encompassed by the silhouette. While both methods yield similar results, the method using area is more consistent with the location of the hip. The method utilizing boundary locations is done by finding the boundary of the silhouette in the binary image. This is done by using Matlab's *bwboundaries* function which traces all objects found within a binary image and stores their boundary locations in a unique cell. The next step is to remove all boundary locations of objects which are smaller than a specified threshold. This accounts for the presence of false positives in the background extraction process. The boundary locations are averaged in the horizontal (x) and vertical (y) direction to yield the centroid location. This method is more robust to the presence of false positives, but can lead to a less accurate estimate. A flow chart of the above method is found in figure 6.4.

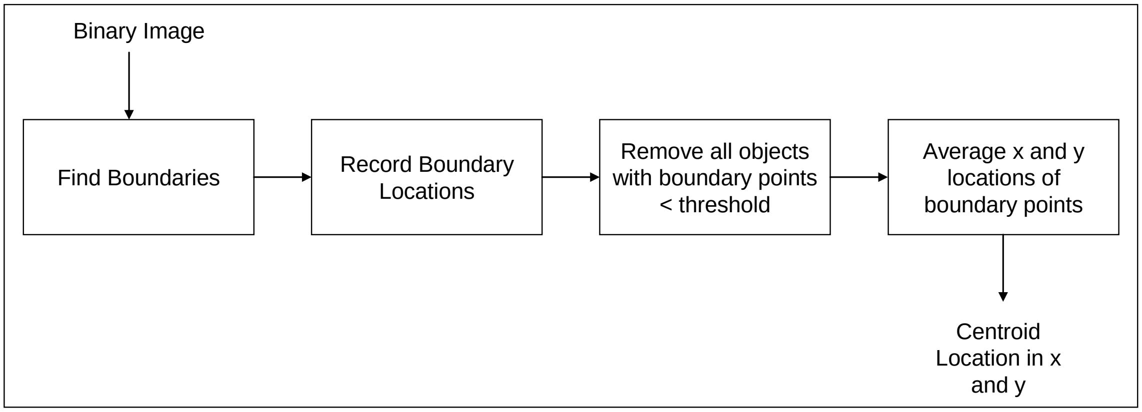


Figure 6-4: Flowchart of operations to approximate centroid location using boundary positions

The second method uses the entire area of the silhouette found in the background extraction process to calculate the centroid. This is done by searching the entire binary image and recording the horizontal, x, and vertical, y, location of foreground objects (1's in this case). The locations in x and y are then averaged respectively to yield the overall centroid location. This method is robust to the presence of false negatives, as described above, because all foreground objects are taken into account. The method is somewhat robust to false positives because the size of the subject's silhouette is very large when compared to false positives generated by the background extraction process. Figure 6.5 below shows the centroid location of a subject using both of the above described methods.

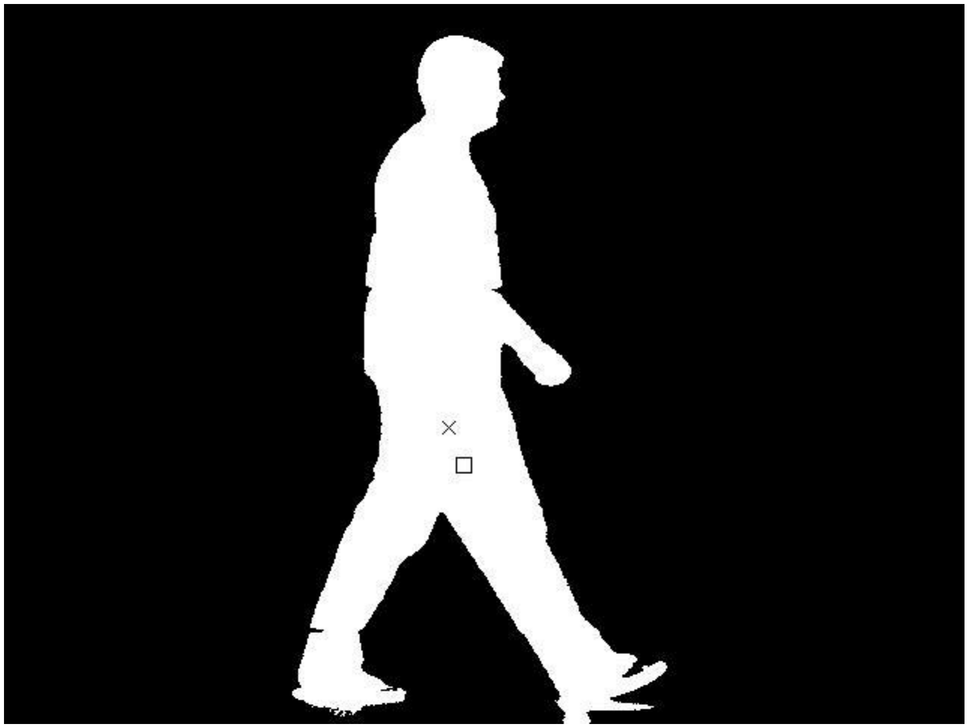


Figure 6-5: A binary silhouette with a relatively low noise level, the ‘X’ is the centroid found using the entire area of the silhouette and the square is the centroid found using the boundary method

Figure 6-6: A binary silhouette with noise and false positives, the ‘X’ is the centroid found using area and the square is the centroid found using boundary position. Notice how little difference the false positives make in the location of the centroid

As can be seen in Figure 6.5, the centroid calculated using the entire area yields a result that is much closer to the hip. The next step in the process is finding the location of the feet. The centroid of a foot will provide an approximate location of the ankle. This position is then used to estimate the hip angle. This location is found in much the same way as the entire silhouette’s centroid except the area used for the calculation is limited. The search area chosen is based on the limitation of movement during gait. The search area for the rear foot is located behind the body’s centroid 150 pixels and up 60 pixels from the floor. The search area for the front foot is located 150 pixels ahead of the centroid and up 60 pixels from the floor. The search areas are shown in the superimposed boxes in Figure 6.7.

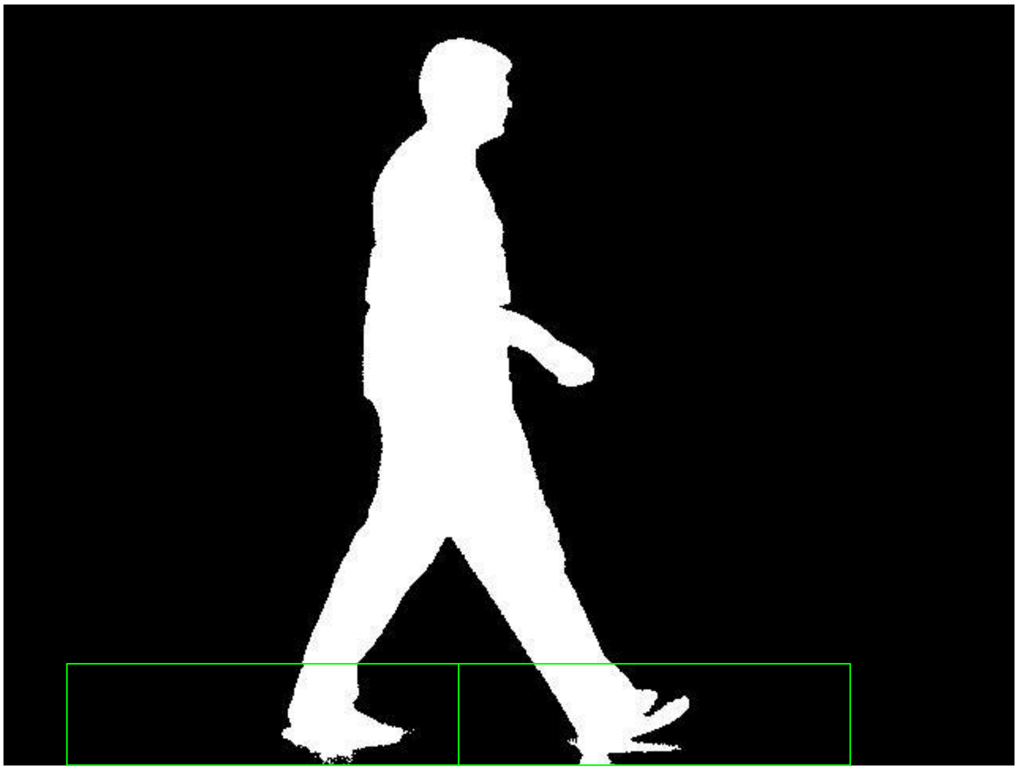


Figure 6-7: Search area for foot centroids superimposed on a binary silhouette

As can be seen in the figure the search area encompasses the feet, but does not include a large amount of any other part of the subject's body. The centroids of the feet are shown as 'X's in Figure 6.8.

Figure 6-8: Foot centroids shown as an ‘x’ superimposed on a binary silhouette

The resulting foot centroids give a good approximation of the ankle location of the leading and trailing foot. The foot centroid locations are used in conjunction with the body centroid to approximate the hip angle. This is done by utilizing the rule of sines as shown in equation 6.5.

$$\Theta = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \tag{6.5}$$

Where

Θ - hip angle

a,b,c – vertices of the triangle

While this does not provide an exact measure of the hip angle, it does provide a good approximation to the actual hip angle. The centroid locations are connected and superimposed on the binary image of the subject's silhouette in the figure below.

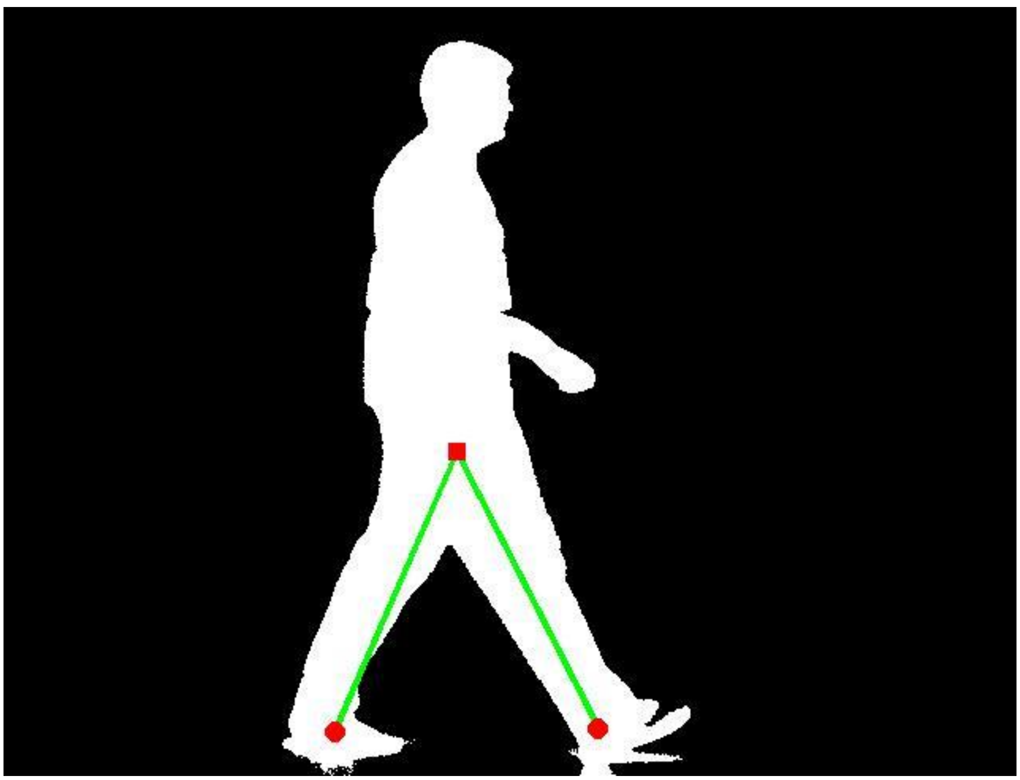


Figure 6-9: The binary silhouette with foot and body centroids superimposed as red circles and a red square respectively. The line segments used to estimate the hip angle are shown as green lines.

The time history of the hip angle measurement found for a normally walking subject is shown in Figure 6.10. Although the hip angle is not exact, it does provide a repeatable measure of how the feet move underneath the center of the body. A similar procedure is used to provide a measure of the torso angle of the subject. This is done by calculating the angle between the highest point on the silhouette and the body's centroid. The points used to calculate the angle are shown in Figure 6.11

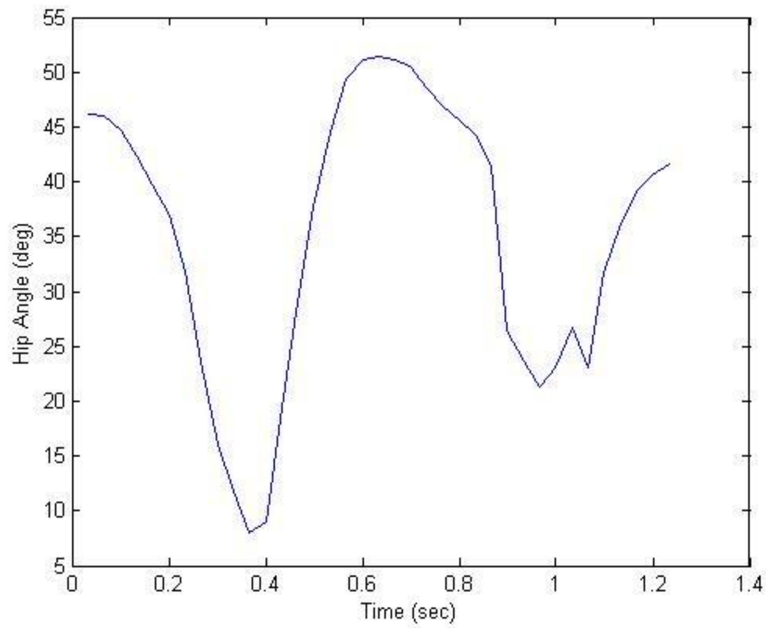


Figure 6-10: Hip Angle in degrees plotted versus time for a normally walking subject taken from video

Figure 6-11: The centroid for the body and the top of the head are shown above as a red square and red circle respectively, the line segment shows the segment used to find

The final values calculated using the silhouette are the height of the subject, the stride length, and max stride length. The height of the subject is approximated by taking the highest point found in the silhouette at a given time. The stride length is found by taking the difference between the point on the silhouette with the highest and lowest value in the horizontal direction. The max stride length is the maximum value of the stride length over the gait cycle. The points used to find the height and stride length are shown in figure 6.12.

Figure 6-12: The point that is used for approximate height is shown as a red square and the two points used to calculate stride length are shown as red squares, both are superimposed on the binary silhouette which was used to calculate the values

6.4 Subject Motion Estimation from Image Sequences

In addition to spatial, information, motion provides another layer of descriptors that can be used to classify gait. Motion information is calculated from a video sequence using a block matching motion estimation scheme. The type of motion estimation scheme used allows for quick estimation of all motion within the scene. An example of the layout that is used for testing of the system is shown in figure 6.13, where the optimal plane for proper motion estimation is labeled in relation to the camera's position.

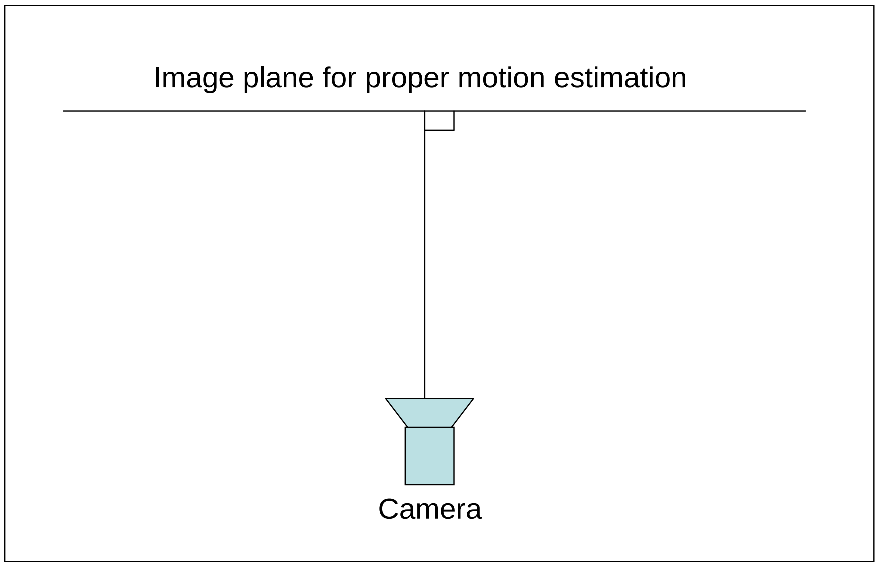


Figure 6-13: Overhead view of how the camera is setup for image capture

If motion does not take place in this plane then the motion calculated is no longer accurate, but is a projection onto the image plane. This does not directly equal the motion calculated for the proper plane. However, the motion could be transformed to the proper plane if the plane of the observed motion is known.

With motion known to be taking place in the proper plane within the scene, a block matching scheme is then utilized to estimate motion. As mentioned in Section 2 of Chapter 5 there are many different features that can be chosen in a block matching scheme. The first of these features is the cost function utilized by the search scheme. In Chapter 5 two of the most popular cost functions were described; Mean Absolute Difference and the Mean Square Error. Both of these functions are computationally efficient, yet still provide a significant measure of how well two blocks correlate to each other. For use in this algorithm, the Mean Absolute Difference function is chosen due to its simplicity and computational efficiency as compared to the various other cost functions. Figure 6.14 shows two quiver plots of the motion calculated using two different cost functions.

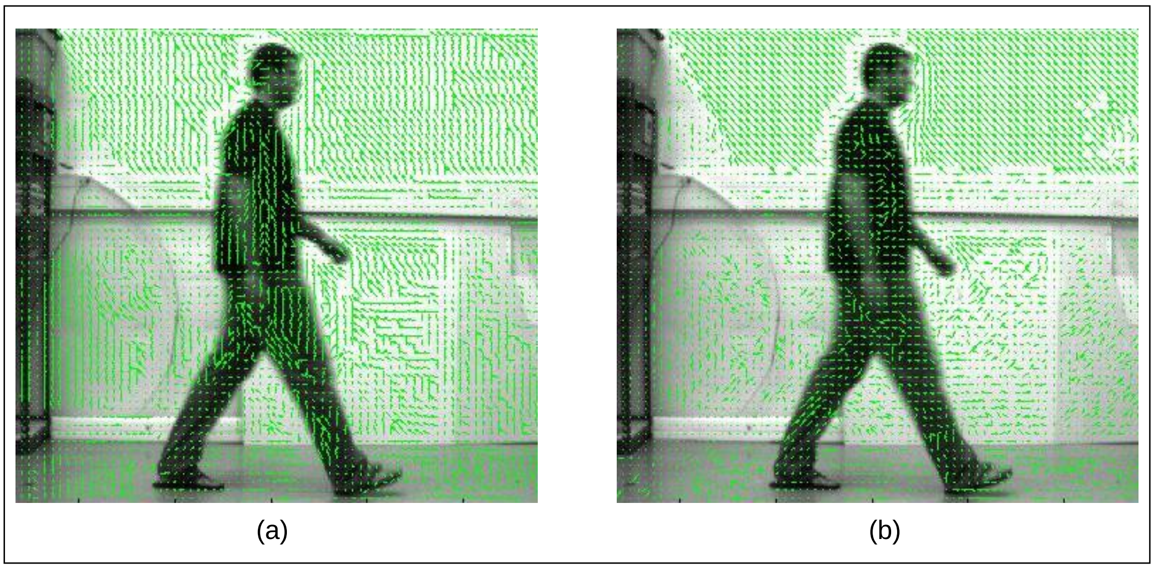


Figure 6-14: Quiver plots of the motion information overlaid on the images from the video sequence, (a) shows the motion calculated using a Mean Absolute Difference and (b) is calculated using a Mean Square Error

With a cost function defined, the next feature to be chosen is the search algorithm. In Chapter 5 the most popular searching schemes were briefly discussed. To choose an effective algorithm for use in this system, various search algorithms were tested. Conceptually, the motion taking place in the subject's sagittal plane during normal walking is that of translational motion through the frame, this largely takes place in the upper body, and the rotational motion of the legs pivoted at the hip. Taking this knowledge into account, it would seem that search schemes which utilize a diamond type search pattern would be the most effective. Two different search schemes are test, the Diamond Search and Adaptive Rood Pattern Search. Additionally, the exhaustive search algorithm is used to provide a baseline of performance. The exhaustive search algorithm provides a "best case" for motion estimation when using block searching [1]. While accuracy is of paramount importance, computational efficiency is also a concern. With this in mind the number of computations, as measured by number of iterations needed for each frame, is also computed for the three search algorithms. The motion found using the three algorithms is shown in the quiver plots found below. The motion vectors are overlaid on the original frame. The motion is found using four consecutive frames taken from a video of a subject walking normally. The video is taken using the camera described in section 1.

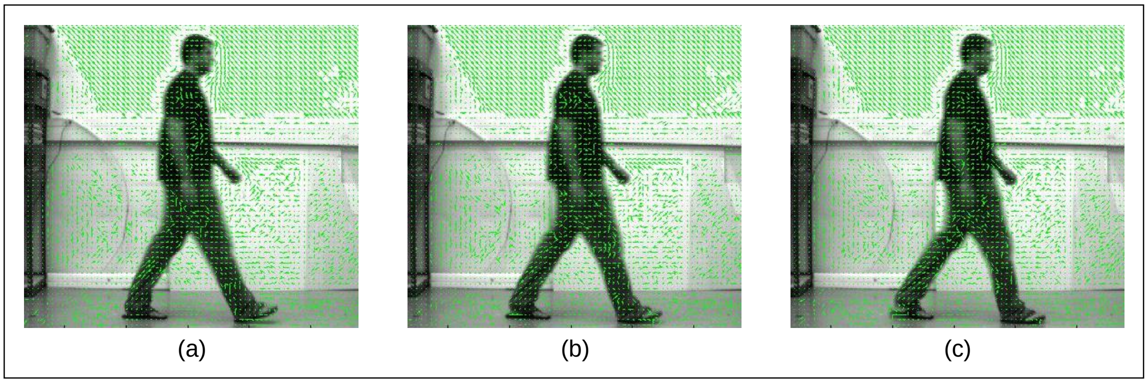


Figure 6-15: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Exhaustive Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video, notice the large amount of errors

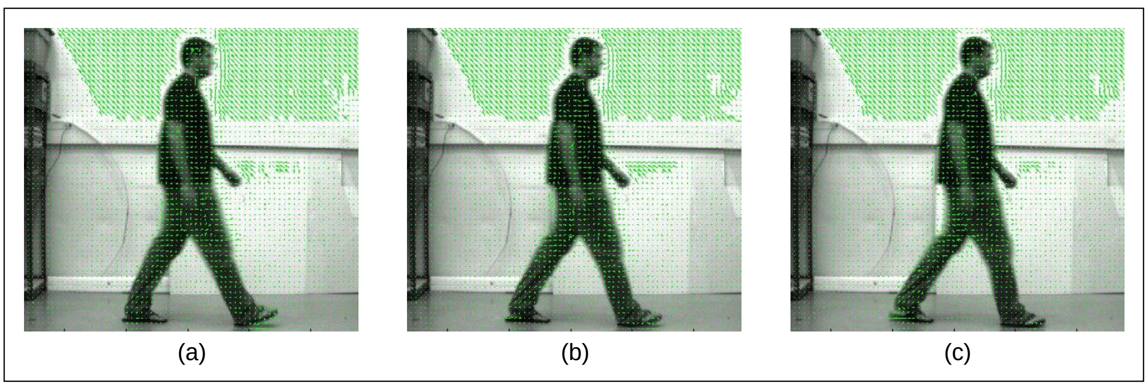


Figure 6-16: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Diamond Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video, notice the large amount of errors

Figure 6-17: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Adaptive Rood Pattern Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video

The above motion estimations are very close to one another, however, a more quantitative measure of the accuracy of the search schemes is desired. To accomplish this, the peak-signal-to-noise ratio is calculated for the motion compensated image as described in [1]. The motion compensated image uses the motion information to construct the next frame in the video sequence. The error between the actual and motion compensated image is then calculated. This is done by using equation 6.6.

$$PSNR = 10\log_{10}\left[\frac{(p_N)^2}{MSE}\right] \tag{6.6}$$

where the MSE is calculated as

$$MSE = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(I(i, j) - M(i, j) \right)^2 \qquad (6.7)$$

and

p_N - peak pixel intensity in the original image

I – original image

M – motion compensated image

m – number of pixels in x direction

n – number of pixels in y direction

The peak-signal-to-noise ratio found for the image sequence used above and the three search methods chosen for analysis is given below.

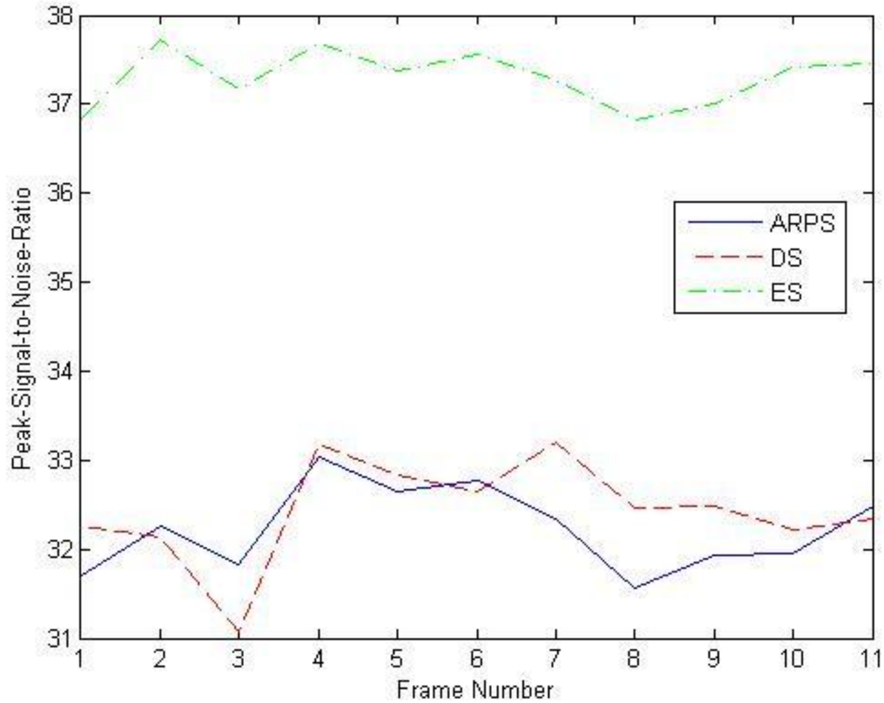


Figure 6-18: Plot of the peak signal to noise ratio for the three search methods, Adaptive Rood Pattern Search (ARPS), Diamond Search (DS) and Exhaustive Search (ES)

A comparison of the computational efficiency for each algorithm is shown in figure 6.19 and figure 6.20.

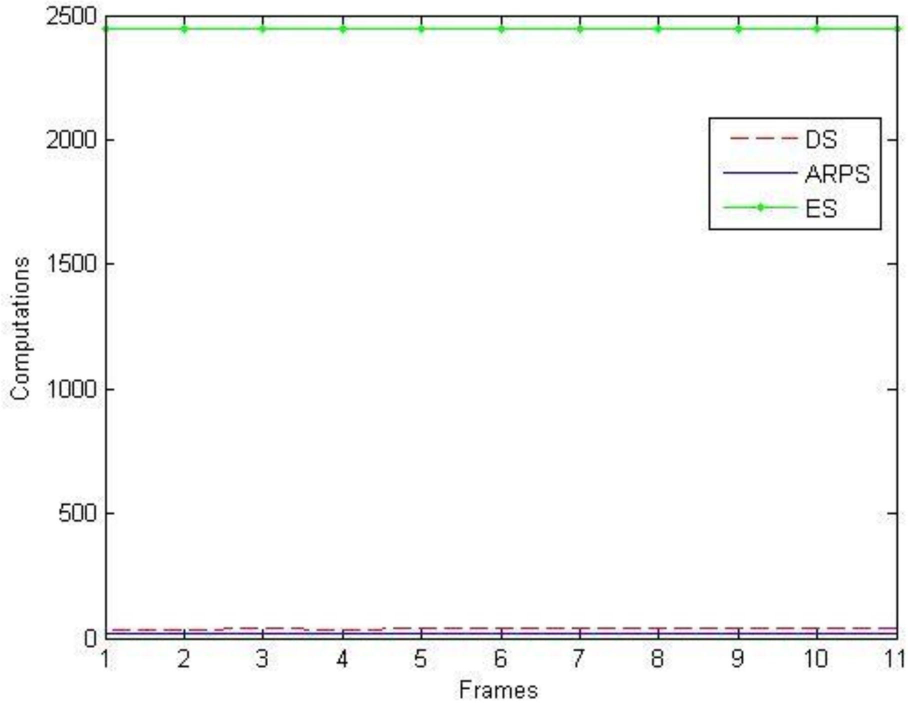


Figure 6-19: Plot of the required number of computations for each search method to arrive at its respective motion estimation for each frame

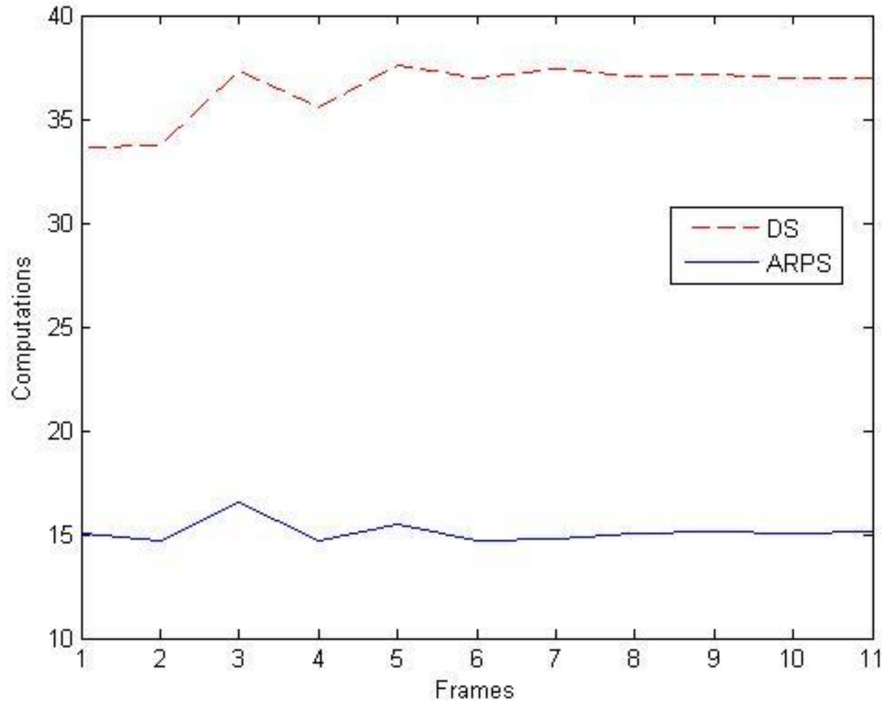


Figure 6-20: Plot of the required number of computations for ARPS and DS search methods to arrive at their respective motion estimations for each frame

As can be seen in the above figures, there is a large difference in the computationally efficiency of the three algorithms. The Adaptive Rood Pattern Search algorithm is by far the most efficient of the search schemes, requiring roughly 20 times less iterations than the exhaustive search method. Additionally, the peak-signal-to-noise ratio of the Adaptive Rood Pattern Search algorithm is only marginally lower than that calculated for the exhaustive search method.

The next step is to select the block size as well as the size of the search pattern. The search pattern must be selected so that it is large enough to capture all motions taking place in the video. However, a larger search pattern also results in a computationally expensive algorithm, so care must be taken to not set the search pattern to large. From observation of normal human gait, a search pattern of 25 pixels is chosen. The block size determines how accurate the motion estimation operation is and determines the motion resolution. A larger block size will provide more information for the block matching algorithm, but objects and areas moving within the scene that are smaller than a block cause errors in the motion estimation. Additionally, a smaller block size causes the algorithm to perform more searches over the image space which makes the algorithm computationally inefficient. Taking these factors into account and researching current motion estimation systems used for MPEG-4 encoding [9] a block size of 8 is chosen. This provides a sufficient amount of motion resolution while still providing a reasonably accurate and efficient block matching algorithm.

6.5 Fusion of Motion Estimation and Silhouette Generation Algorithms

With all information gathered by the system thus far, an efficient fusion of all the above techniques, background extraction, foreground determination and motion estimation is performed. Each method should not operate individually when information could be passed from one algorithm to another to create a more efficient system. Foreground areas determined in the previous steps of the system are used to limit the search area for the motion estimation algorithm. This reduces computational time and also reduces errors associated with edge regions typically found when using block matching schemes.

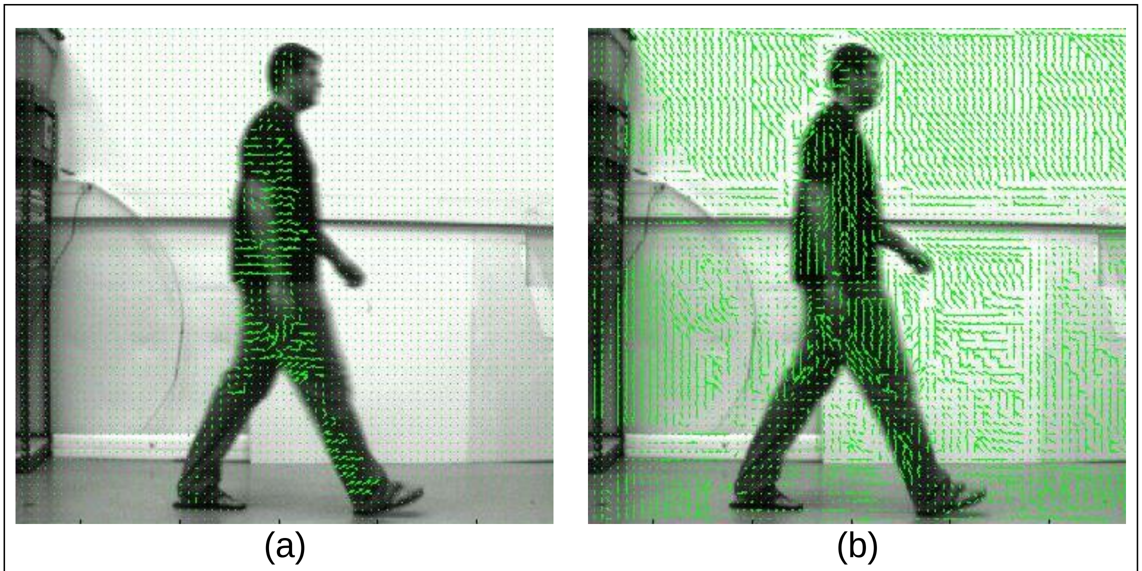


Figure 6-21: The result of algorithm fusion for this system, both images have the associated motion information superimposed, (a) is the algorithm used in the system and (b) is the traditional ARPS algorithm

In the implementation of this system, the foreground location is passed to the motion estimation algorithm. This information is used to define a search area of the image. This eliminates search operations where no foreground object exists. Due to these improvements, the motion information calculated from the video sequence is much more accurate.

An additional operation carried out on the motion data is windowing. This centers the subject in the image for each frame. This way all motion of the subject is calculated relative to the upper body. This allows motion of a specific region or part of the body to be analyzed solely on its relative motion and eliminates the effect of translational body motions. A number of variables are used to accomplish this, including stride length and centroid information. The window is created using the entire height of the frame, but is limited in the horizontal direction by maximum stride. This allows all motion taking place in the scene for a given subject to be wholly contained by the window. Next, the location of each of the motion vectors so that they will only move spatially in relation to the upper body in the windowing frame is done. Using the centroid location of the subject for each frame an offset is calculated to shift the entire subject's body into the space of the window. This is done by finding the difference between the subject's centroid and the center of the window then setting the value of the offset to this difference. This offset value is then added to the location information. Resulting from the operation is a video sequence consisting of motion vectors that move in relation to the centroid. A block diagram of the system operations up to this point in the processing is shown in figure 6.22.

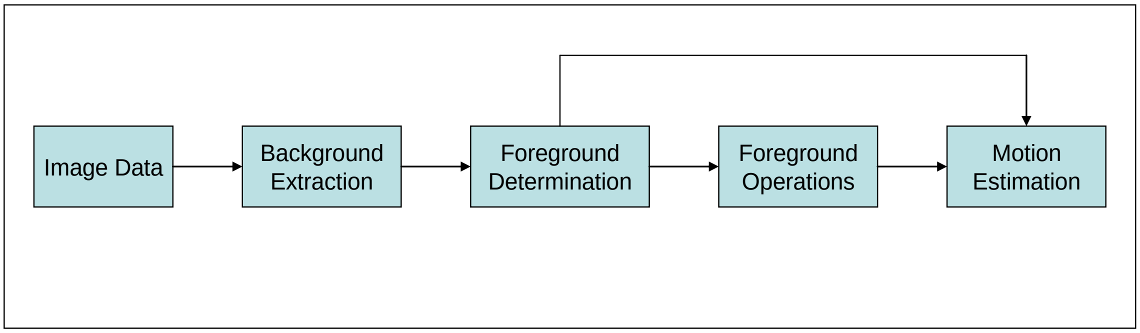


Figure 6-22: Flow chart of the operations performed by the system to form descriptors of human gait

With all of the above information complied, the next step of the system is to analyze the processed data so that classification of the gait as normal or abnormal can be done.

6.6 Classification System for Normal Gait Determination using Image Sequences

Once all the descriptors of the subject have been calculated and complied by the system, the next step is to accurately determine if the subject is walking normally. There are a variety of steps taken to analyze this data. The recognition scheme consists of feature generation and selection as well as classification scheme. Each of these determines the effectiveness of the overall pattern recognition system.

The first step to select the classifier. Of the many types available, the perceptron neural network algorithm is chosen. This is chosen based on its simplicity, ease of implementation and ability to be trained using actual data. The perceptron algorithm is described in detail in Chapter 4. The next step in the design of this algorithm is feature generation. Since the data gathered from the system is almost entirely vectors, the data must be dimensionally reduced so that it may be used in the pattern recognition scheme. A comparison between the observed data and normal gait is needed. To provide an accurate measure of how well an incoming signal correlates to a normal walking signal, the incoming signal is compared to normal walking signals. The similarity values generated for each of the normal signals is averaged to form the similarity measure. Three types of signal quantification are tested for use with this system, cross-correlation at zero lag, correlation coefficient and error between the two signals. The error term used is shown below.

$$\frac{\int_1^N (X_n - N_n)^2 \, dn}{\int_1^N N_n^2 \, dn} \tag{6.8}$$

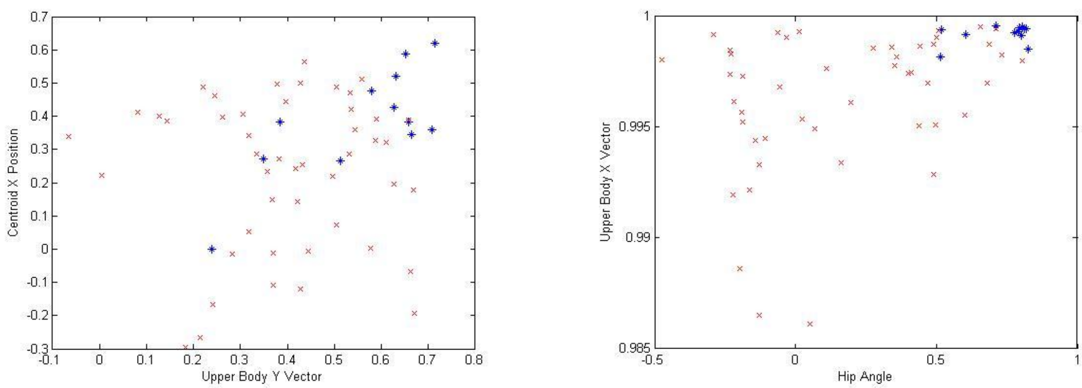
where

X_n - n^th point of the incoming signal

N_n - n^th point of normal signal

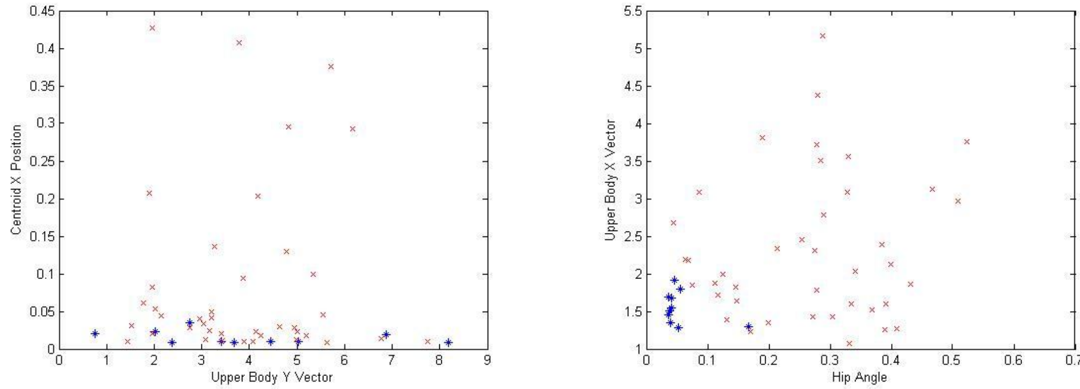
N – total number of points in the signal

Each of these methods is tested using a set number of features in the perceptron architecture. This suboptimal search method is used because it is difficult to estimate the effectiveness of the final classification system using seperatability criteria when a large number of features are used [8]. An example of scatter plots is shown for two feature values generated using the correlation coefficient.



(a) (b)

Figure 6-23: Scatter plots of feature values for use in the classification system, (a) shows a plot of Upper Body Y Vector vs Centroid X Position and (b) shows Hip Angle vs Upper Body X Vector. The blue stars represent normal walking data and the red x's represent abnormal walking data. The most effective way to compare data sets is two at a time since the final perceptron will create a hyperplane between that best separates the number of features chosen



(a) (b)

Figure 6-24: Scatter plots of feature values found using the error term, (a) shows a plot of Upper Body Y Vector vs Centroid X Position and (b) shows Hip Angle vs Upper Body X Vector. The blue stars represent normal walking data and the red x's represent abnormal walking data.

Figure 6.25 shows the scatter plots using the error term. As can be seen in the above figures, the method that provides the best linear separation boundary (since a one layer perceptron is being used) is difficult to determine. A set number of features are used in conjunction with the perceptron to determine the effectiveness of the feature generation methods. The method that yields the most accurate classification is the feature with the highest classification effectiveness. The testing is done using the Hip Angle and Centroid X position features in conjunction with a one layer perceptron. The features are generated as described above. The perceptron is then trained using 5 normal and 5 abnormal walking data sets. The remaining data sets are used for validation. The results from the test in table 6.1.

Feature Generation Method	**Accuracy of final system**
Cross Correlation	6.79%
Correlation Coefficient	84.75%
Error	77.97%

Table 6.1: Results of Feature Generation Testing

As can be seen in Table 6.1, when the correlation coefficient is used the results of the final system are much better than the other two methods tested. Because of this, the correlation coefficient is the chosen as the feature generation method for this system.

The next step is to decide which combination of features yields the most accurate classification system. This is again done using a suboptimal search. A Forward Selection technique is used to select features for use in the system by forming all possible combinations of two feature vectors and the classification accuracy for each combination is calculated. The combination that yields the best classification accuracy is then adopted and taken to the next stage. Next, each of the remaining features is paired with the selected combination and the most accurate system is taken to the next stage. This process is repeated until the desired accuracy is obtained or the desired number of features is added into the system. The Hip Angle, Centroid Y Position, and Upper Body Y Motion are selected for use in the system as they yield the highest classification accuracy using the smallest number of feature vectors. The results of the intermediate steps in the suboptimal Forward Selection technique can be found in Appendix A.

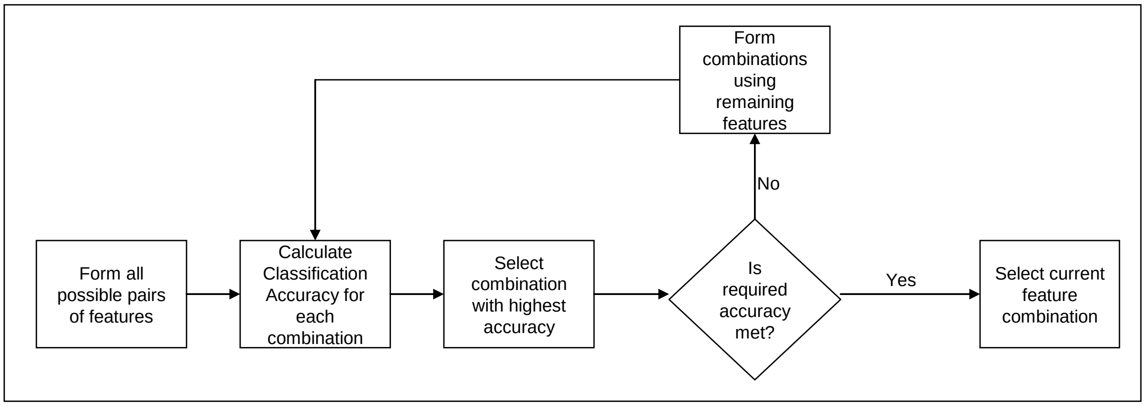


Figure 6-25: Flowchart of suboptimal search method technique used for feature selection

6.7 Testing of Normal Human Gait Classification System

The system must now be verified and tested to determine the effectiveness of gait classification. This is done by taking videos of a number of subjects walking normally as well as abnormally to create a baseline set of data with which to train the system and to test the effectiveness of the system.

6.7.1 Experimental Setup for Data Collection

The system is trained and tested using video of people walking in front of a camera. This is done because the complexity of the human body does not allow for easy modeling. Additionally, abnormal walking provides a scenario in which individuals will react differently. All data collected for this system was done in a controlled indoor environment. A walking path was created in an open area over a flat, clean floor to reduce the effects of the floor on the subject's gait. The path created for walking consisted of two marked points located directly across from each. The camera used for image acquisition was placed so that it was perpendicular to the walking path. The area behind the walking path was kept static during the duration of data. One subject's data was collected in front of a moving background to determine the robustness of the system to noise. The setup for this data collection processes is shown in figure 6.26.

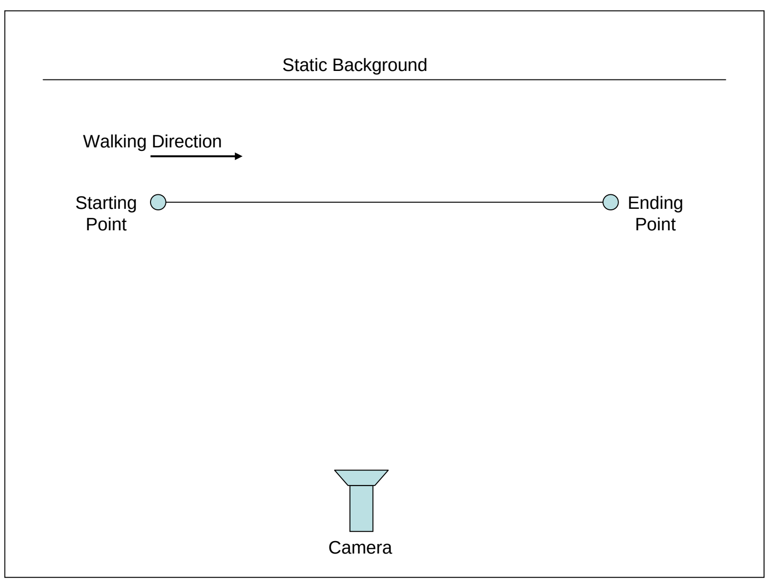


Figure 6-26: The experimental setup, an overhead view

Each subject was instructed to walk in one of five different ways:

- Normally
- Limping
- Shuffling
- With One Leg Straight
- Jogging

The five walking styles were completed twice by each subject. This resulted in ten data sets for each individual subject. 6 subjects were chosen for use in this project. To reduce the effect of gender and other various physiological effects, the subjects were all males between the ages of 22 and 25 of various heights and weights. The age, height and weight data for each subject can be found in table 6.3.

Subject	**Age**	**Height (in)**	**Weight (lbs)**
1	24	73	170
2	26	68	152
3	22	74	185
4	23	73	172
5	24	77	255
6	24	71	163
Average	**23.83**	**72.67**	**182.83**

Table 6.2: Subject Data for testing, all subjects are male

All data collection and processing was done by the system described above. The data is then fed into the system where background extraction, foreground determination, spatial processing, and motion estimation processes and classification are done. The classifier is trained using 5 of the normal data sets and 5 of the abnormal data sets from random test subjects. The remaining data from the test is then processed and the final classification decision for each of the subject's walking trials is computed. This output is then compared to the known gait classifications so that the effectiveness of the system can be determined. The accuracy of the final system classification is given by a percentage of correct classifications.

6.7.2 Results of Normal Gait Classification System

The final system accuracy obtained for the testing described above is 93.22%. This means that the system will classify an observed gait correctly over 93% of the time. The results for each individual subject can be found in table 6.3, the subject walking with a moving background is subject number 6 in the table.

Subject	Correct Classifications	Total Trials	Accuracy
1	11	12	91.67%
2	12	12	100%
3	12	12	100%
4	12	12	100%
5	11	12	91.67%
6	9	11*	81.82%
Total	**55**	**59**	**93.22%**

Table 6.3: Classification results for each subject, Note: * denotes one data set did not process due to gross background errors

As can be seen above, much better results are obtained by the system when using a purely static background. In fact 50% of the error in the classification is due to Subject 6. A more accurate classification would result for Subject 6 if the background extraction algorithm was modified to model the movement taking place in the scene. Despite the gross error encountered by Subject six's data set, the system still manages to accurately classify an acceptable number of the observed gaits.

6.7.3 Conclusions of Gait Classification Experiment

The system provides an efficient and accurate scheme for determining if a human is walking normally or abnormally. Even though the abnormalities used for the evaluation of this system are not “true” gait abnormalities they provide subtle difference in the human gait pattern that can be distinguished from normal gait with a properly trained system. Based on the results from this test it can be inferred that true abnormalities would be even easier to detect using this system, as the gait resulting from someone who nominally walks abnormally will provide motion which is much more consistent over the course of the gait cycle. Greater repeatability would provide more stable data to the classification system and theoretically better results from the system as a whole.

6.8 Classification of the Type of Observed Human Motion

To provide an additional level of analysis, further development of the system described above is done to classify the type of abnormal gait observed. For this system, the hip angle estimates generated as described in section 6.4 are used. A comparison of the hip angle calculated using 3-D motion capture data and the method described above is shown in Figure 6.28.

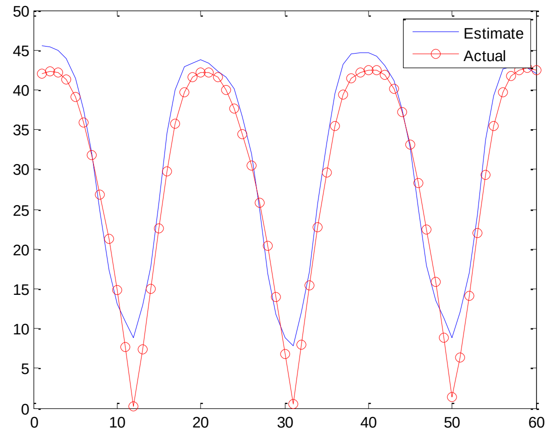


Figure 6-27: Estimated versus actual hip angle

The final measure taken from the visual input data is a hip angle normality measure. This is calculated by finding the correlation coefficient between a representative normal series of hip angle measurements and the input series of hip angle measurements. This assists in accurate determination of the observed walking style by the neural network classifier. This measure as well as the raw hip angle estimate form the input set for the neural network classifier.

6.8.1 Neural Network Approach to Human Activity Classification

There have been a wide variety of neural networks developed in the past 20 years aimed at solving a number of problems [82, 86, 89, 90]. Many of these neural networks have been applied to a wide variety of pattern recognition applications with a few being used in the field of visual gait and activity analysis [97, 98]. This system utilizes a multilayer neural network to form a powerful classifier to determine walking style given gait data obtained through a computer vision system.

To classify the hip angle input, the neural network was constructed with 3 layers and a maximum operator. The size of the input layer for this system is 46 neurons. This allows for 45 hip angle points and a hip angle normality factor to be input to the neural network. The size of the second layer is constructed to minimize the number of neurons required and still provide an acceptable final classification performance. This results in a second layer size of 60 neurons. Each of the first two layers is constructed using a log-sigmoid activation function. The output layer consists of 5 neurons, one for each walking style to be classified, each with pure linear activation functions. The maximum of the 5 output neurons is the final classification decision. The architecture used in this system can be found in Figure 6.29.

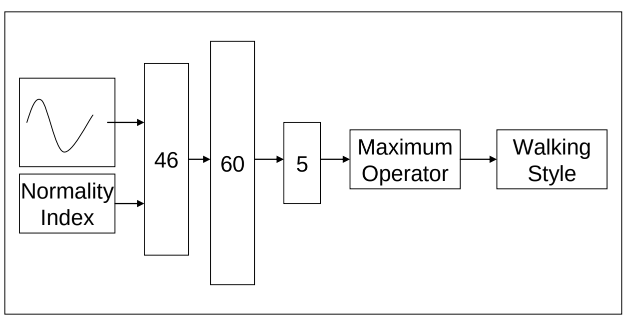


Figure 6-28: Neural Network Architecture

The network used in this system is a supervised classification scheme and therefore a training data set is used. 20% of the available data is used to train the network. This allows for good generalization of the classification results and not just those observed in the experiment. The training algorithm chosen is a quasi-newton method known as the Marquardt backpropagation algorithm [99]. This uses Levenburg-Marquardt optimization to modify the traditional Gauss-Newton method and results in a training algorithm that is more efficient than traditional backpropagation (steepest decent) as well as conjugate gradient algorithms [127]. The error of the network was calculated using mean square error (MSE). The goal for the training error is an MSE of 1e-20, with a maximum of 1000 epochs.

6.8.2 Testing and Results for Human Activity Classifier

For this testing 10 subjects were observed performing various styles of walking. The subjects were all males between the ages of 22 and 32 of various heights and weights. The age, height and weight data for each subject can be found in Table 6.4.

Subject	**Age**	**Height (in)**	**Weight (lbs)**
1	24	73	170
2	26	68	152
3	22	74	185
4	23	73	172
5	24	77	255
6	24	71	163
7	23	67	147
8	32	70	182
9	22	71	178
10	24	73	182
Average	**24.4**	**71.7**	**178.6**

Table 6.4: Subject Data for testing, all subjects are male

A narrow band of the population was chosen in order to focus on the differences in gait due to the abnormality exhibited and not due to the subjects' gender, race or some other outside factor.

The 10 individuals walked along a path situated so that they would be walking perpendicular to the camera's primary axis. The path was over a flat surface free from obstructions. This was done to view the saggital plane of the subject. Each subject was instructed to walk from one point to the other using a variety of different walking styles for each trial. This allowed for the collection of one complete gait cycle.

Each subject was instructed to walk in one of four different ways:

- Normally
- Limping
- Shuffling
- Jogging

Each of the four walking styles was repeated once by each subject. This results in a total of eight separate data sets for each subject.

The video of each subjected was collected at 30 frames per second with a resolution of 640x480 pixels per frame. The neural network classifier was trained using 40 of the data sets selected at random from the 80 available data sets. The remaining data from the test was then processed using the system and the final classification decision for each of the subject’s walking trails was computed. This output was then compared to the known gait classifications and the effectiveness of the system was determined based on the accuracy of the final system classification, as given by a percentage of correct classifications.

The testing described above yielded a correct classification decision for 34 of the 40 test data sets this is an 85% classification rate. Examples of the image sequences provided to the system are shown in Figure 6.29.

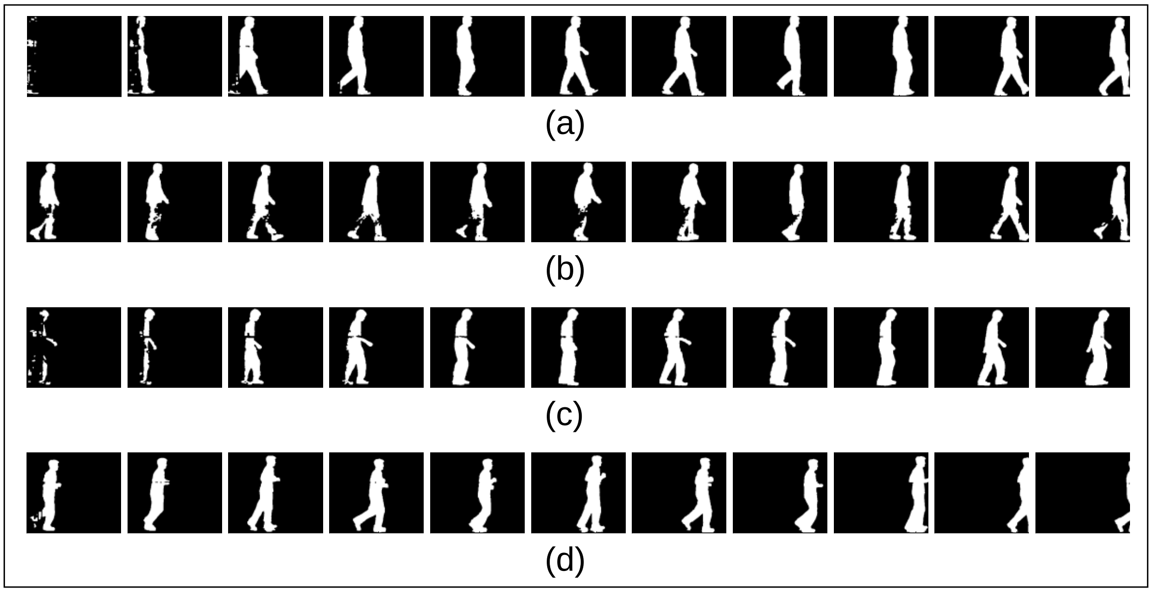


Figure 6-29: Image data provided to the system, (a) normal walking, (b) limping, (c) shuffling and (d) jogging

The final classification decision is compared to the actual observed class in Table 6.5.

Trial	Classification	Actual Class
60	3	3
13	1	1
2	1	1
17	**3**	**1**
36	2	2
59	3	3
72	4	4
70	4	4
71	4	4
65	4	4
54	3	3
64	4	4
66	4	4
16	1	1
29	2	2
45	**2**	**3**
3	1	1
24	**1**	**2**
34	2	2
44	**2**	**3**
63	4	4
19	1	1
35	2	2
48	3	3
51	3	3
28	2	2
50	3	3
12	1	1
30	2	2
80	4	4
77	**3**	**4**
8	1	1
76	4	4
14	1	1
20	1	1
15	**2**	**1**
68	4	4
26	2	2
1	1	1
22	2	2

Table 6.5: Classification Results

The classification failures are highlighted above. The figures 6.30-6.35 show the hip angle estimates for the 6 classification failures compared to examples from that walking class.

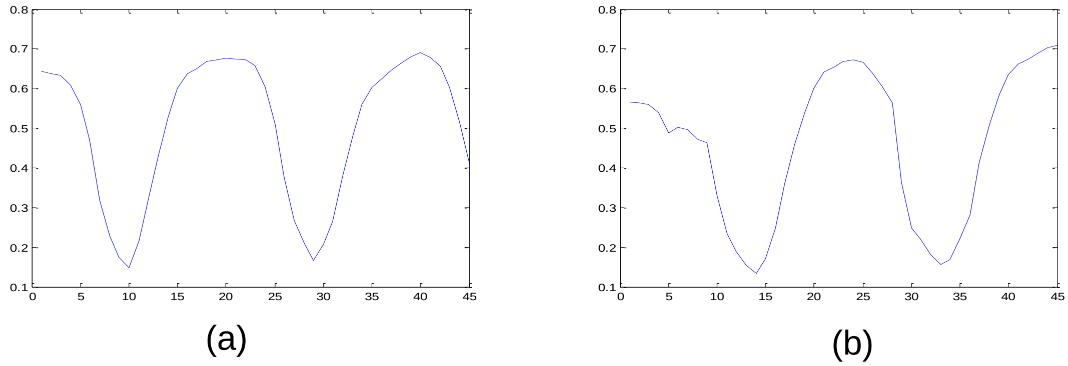


Figure 6-30: (a) Correctly classified normal walking (dataset 1), (b) incorrectly classified normal walking (dataset 17)

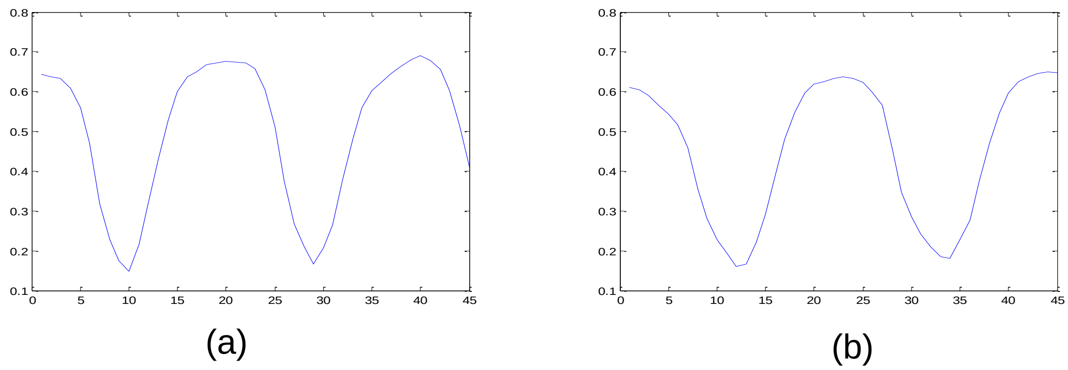


Figure 6-31: (a) Correctly classified normal walking (dataset 1), (b) incorrectly classified normal walking (dataset 15)

Figure 6.32: (a) Correctly classified normal walking (dataset 1), (b) incorrectly classified normal walking (dataset 15)

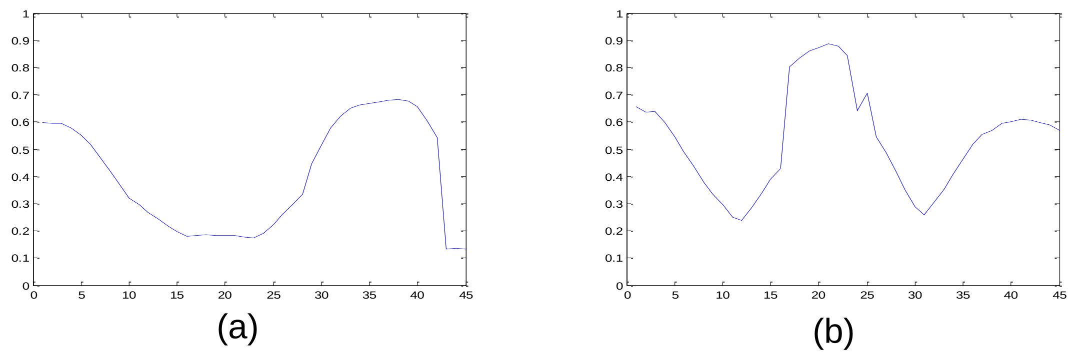


Figure 6-32: (a) Correctly classified limping (dataset 35), (b) incorrectly classified limping (dataset 24)

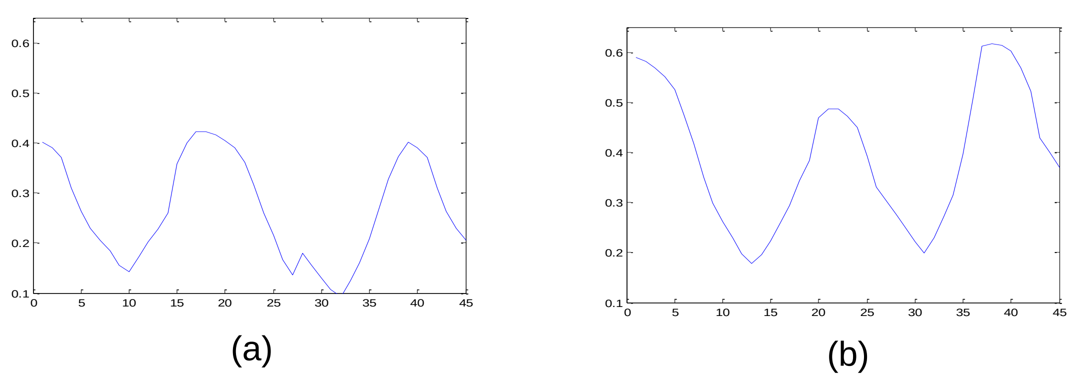


Figure 6-33: (a) Correctly classified shuffling (dataset 50), (b) incorrectly classified shuffling (dataset 44)

Figure 6.34: (a) Correctly classified shuffling (dataset 50), (b) incorrectly classified shuffling (dataset 44)

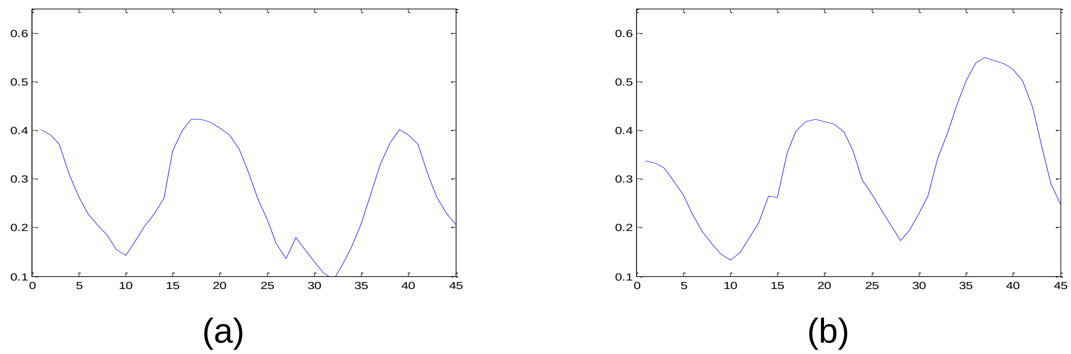


Figure 6-34: (a) Correctly classified shuffling (dataset 50), (b) incorrectly classified shuffling (dataset 45)

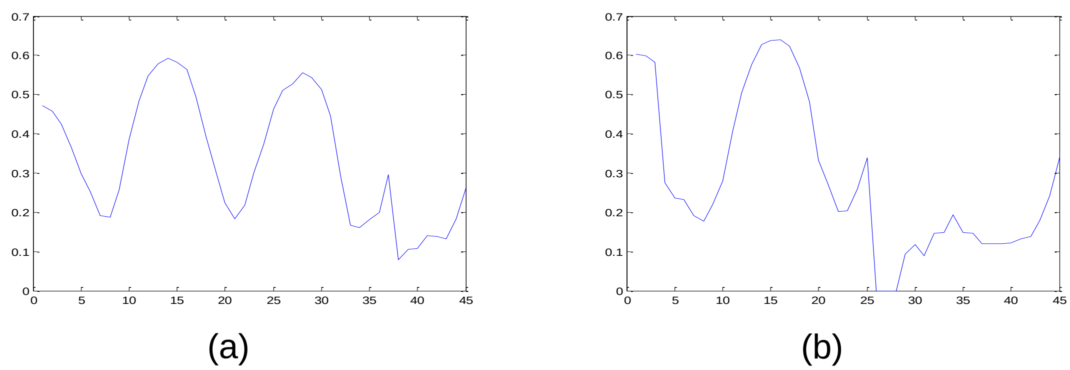


Figure 6-35: (a) Correctly classified jogging (dataset 70), (b) incorrectly classified jogging (dataset 77)

The classification failures are due to differences between the hip angle estimates calculated for each image sequence. Some of these failures can be attributed to noise in the silhouettes caused by background extraction errors (datasets 17, 77 and 24) while others are caused by differences in the actual walking motion performed by the subjects (datasets 44 and 45).

6.8.3 Conclusions for Human Activity Classification System

Neural networks are utilized to create a powerful, efficient and robust classifier which can correctly classify an observed walking style using only simply obtained image metrics. This classification is done using a three layer neural network in conjunction with a maximum operator. The image metrics provided to the classification network are simple to obtain and robust to image errors and self occlusions. Once the network is fully trained, the system can run quickly and efficiently to classify walking motions observed in an area by a single monocular camera. The system is shown to provide an 85% correct classification rate when classifying four different styles of walking.

The system described is currently limited to a single view. However, further development can allow for the system to be generalized to multiple views. This generalization remains simple because the only modification that must be made for the system is the ability to obtain the hip angle estimate from additional view points. Because this estimate is based on simple image metrics it should not be difficult to find from multiple views, especially when using a stereo imaging system.

Chapter 7: Image Based Detection of Hidden Load Conveyance by Human Subjects

For this system, statistical distributions generated by the boundary point training sets are used in a hybrid fuzzy neural network to classify silhouette boundary points into kinematic families. The fuzzy input set contains part-based likelihoods for each of seven different body segments. The seven body segments are selected for their kinematic importance. These segments are torso, front thigh, back thigh, front shank, back shank, front foot and back foot. For each boundary point on the silhouette, the likelihood is calculated for each of the seven different kinematic families. The Gaussian likelihood function used as the fuzzifier is shown in equation 7.1.

$$p(x_i \mid L) = \frac{1}{\sqrt{2\pi}|\Sigma|^{-1/2}} \exp\left((X-\mu)^T \Sigma^{-1} (X-\mu)\right) \qquad (7.1)$$

Where L is the part family, X_i is the edge point and (μ_L, Σ_L^2) are the parameters of the distribution for part family L. The parameters for each kinematic family are learned from manually labeled image sequences. Hybrid fuzzy network architecture is shown in Figure 4.5.

The outputs of the neural network consist of scores for each of the 7 kinematic families. Each output node corresponds to a kinematic family and the resulting point classification is determined by the node with the maximum value. To train the neural network, manually labeled edge points were generated for 8 subjects. 4 subjects were then used to train the network and 4 subjects were used for network verification. The network achieves a 95.8% correct classification rate. The majority of the incorrect classifications exist in the shank regions due to complexities of front/back classifications and self occlusions. Examples of the neural network classifications for previously unseen subjects are shown in Figures 7.1-7.5. The figures below show a typical image sequence (every fifth frame) for each of the load conditions (loaded/unloaded) and both types of subjects (male/female). As can be seen, the normalization makes each subject approximately the same size in the image. Additionally, there is no visual evidence of the presence of an external load.

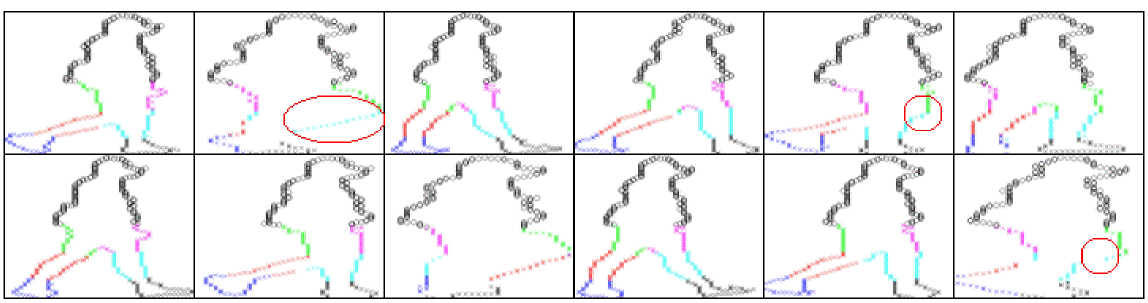


Figure 7-1: Example of Kinematic Family Classifier for Loaded Male Subject

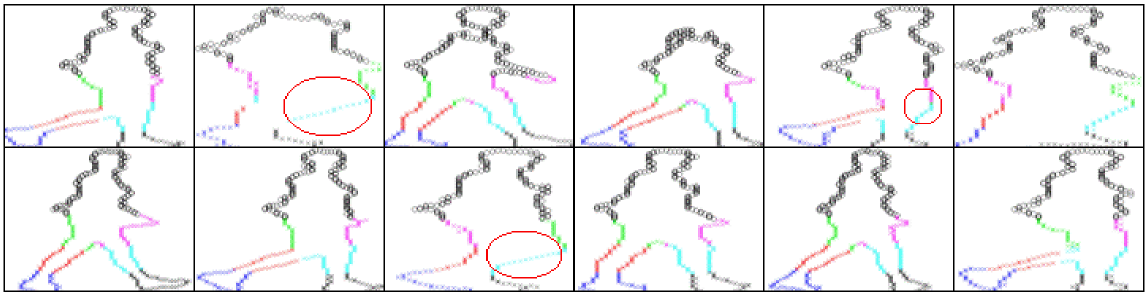


Figure 7-2: Example of Kinematic Family Classifier for Loaded Male Subject

Figures 7.1 and 7.2 demonstrate a typical image sequence for a loaded male subject. As can be seen in the figures, the majority of the point misclassifications occur in the shank region during the single stance phase of the gait cycle. These errors are highlighted by the red circles in Figures 7.1-7.5. These errors occur because the front/back leg classification is determined based on foot location. As can be seen above, the majority of the back leg can be located forward of the front leg even though the back foot remains behind the front foot. This causes the front/back misclassifications seen in the shank regions of both subjects.

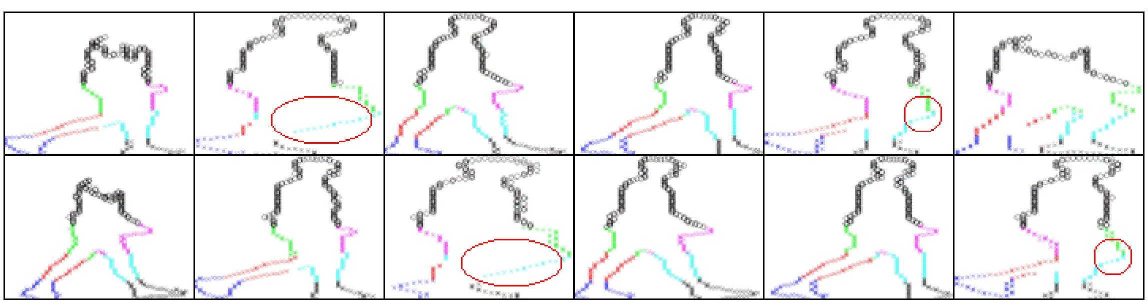


Figure 7-3: Example of Kinematic Family Classifier for Unloaded Male Subject

Figures 7.1-7.3 give examples of loaded and unloaded subjects. As can be seen in these figures, there is no visual difference between loaded and unloaded male subjects. Figure 7.3 also shows that the front/back shank misclassifications in the single stance phase of the gait cycle occur as in the loaded case (as shown in Figures 7.1 and 7.2).

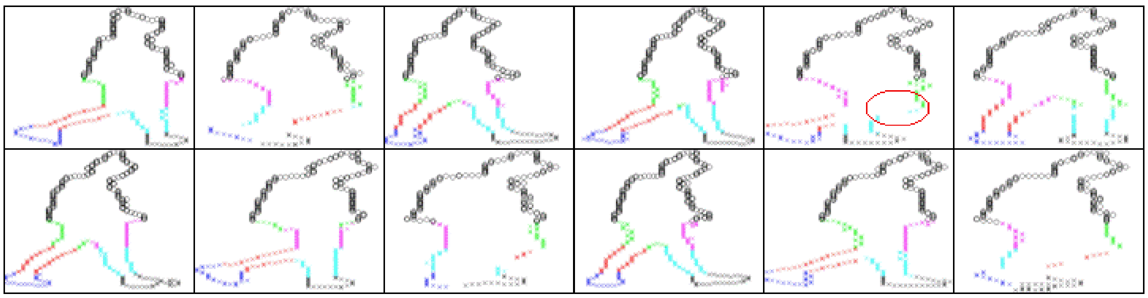


Figure 7-4: Example of Kinematic Family Classifier for Loaded Female Subject

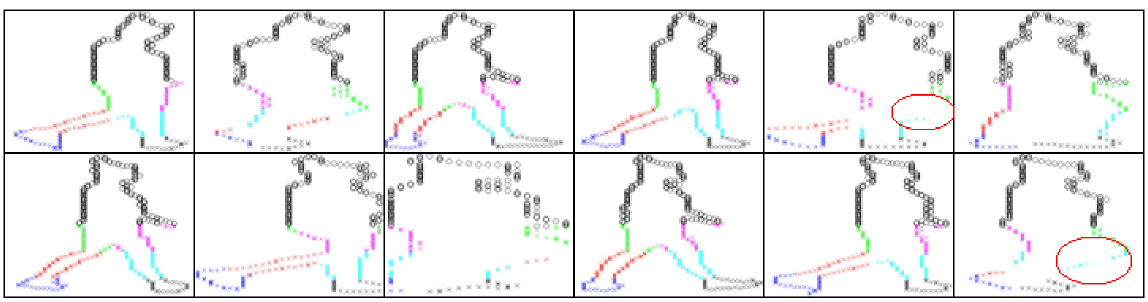


Figure 7-5: Example of Kinematic Family Classifier for Unloaded Female Subject

Figures 7.4 and 7.5 demonstrate image sequences for loaded and unloaded female subjects. As can be seen the female and male subjects remain visually the same after image normalization. The boundary point classifications are also similar to those for the male subjects. As can be seen in Figures 7.1-7.5, the front/back shank misclassifications exist for the female subjects as well as the male subjects.

7.1 Joint Angle Estimations for Use in Hidden Load Classification System

The joint locations of the observed subject are estimated in each frame by the labeled edge point data. The first step in this estimation is to determine landmark points in the labeled edge sets that can be used for joint position estimation. The boundary points between kinematic families were chosen as landmark points due to their kinematic significance. These points best describe the movement of the lower body limbs. The offsets from these landmark points to the actual joint location were then learned from training data. An iterative technique was adopted to determine the value of each landmark offset. Initial estimates for each offset were made based on visual observation and the ground truth angle data. Each of the remaining landmark offsets were then estimated so that the error between the estimated and actual joint angle was minimized for each training image. This process was repeated for each joint until the angle error between the visual and motion capture data was minimized for the training set of 4 subjects. Joint angle measurements found in the validation set are shown in Figures 7.6-7.10.

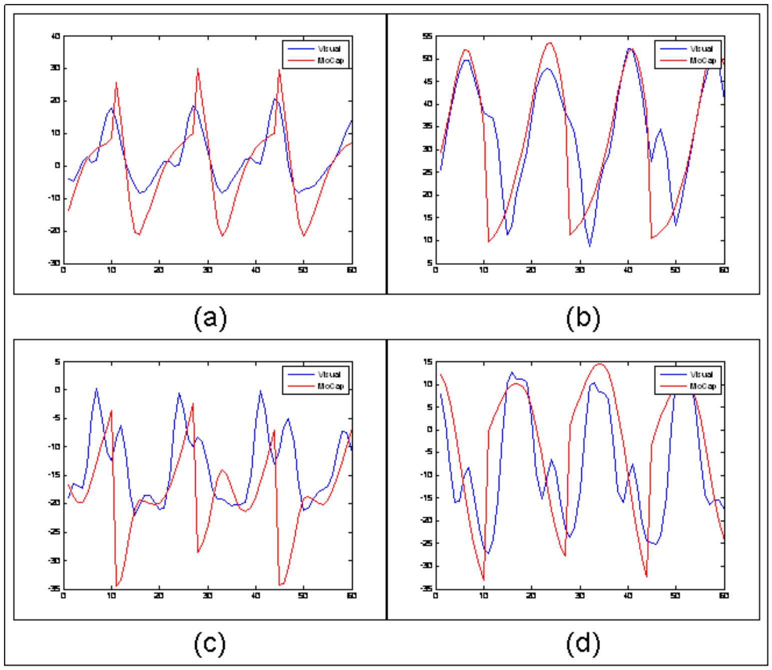


Figure 7-6: Example of Angle Estimations for (a) Front Knee, (b) Back Knee, (c) Front Hip, (d) Back Hip of an Loaded Male Subject

Figure 7.6 shows typical angle measurements for a loaded male subject. As can be seen in the above figure the visual estimates and motion capture angles are not exactly the same. However, the phases of the joint angles are similar. As can be seen in the Figure 7.6, the back joint angles appear to be more accurate than the front joint angles. The front/back transitions for each angle do not occur at the same in the visual and motion capture systems. This is another source of error in the estimation. This is evident in each figure when large step transitions occur at different points. Typically the visual system remains slightly behind that of the motion capture system. This is due to the fact that the front/back transition is determined by the ankle marker (furthest back point in the leg) in the motion capture system and by the furthest back heel point in the visual system.

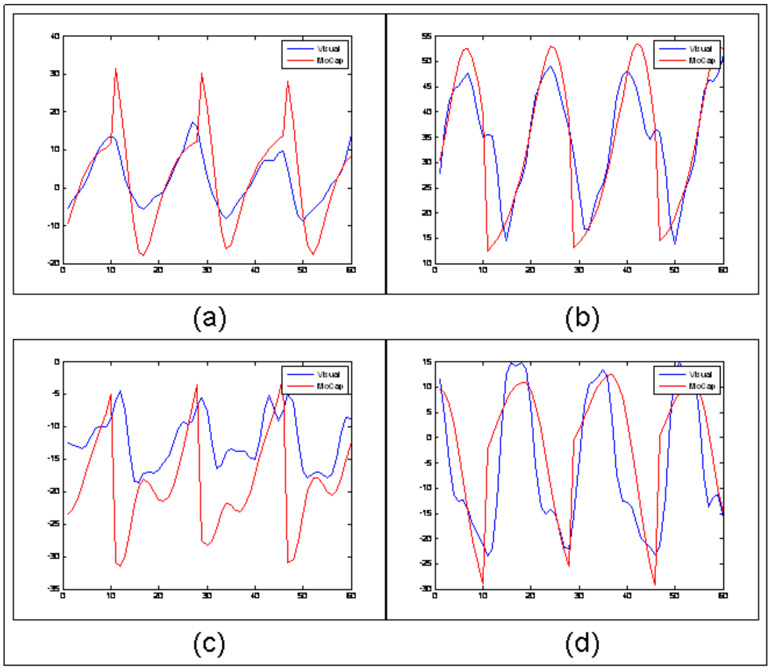


Figure 7-7: Example of Angle Estimations for (a) Front Knee, (b) Back Knee, (c) Front Hip, (d) Back Hip of an Unloaded Male Subject

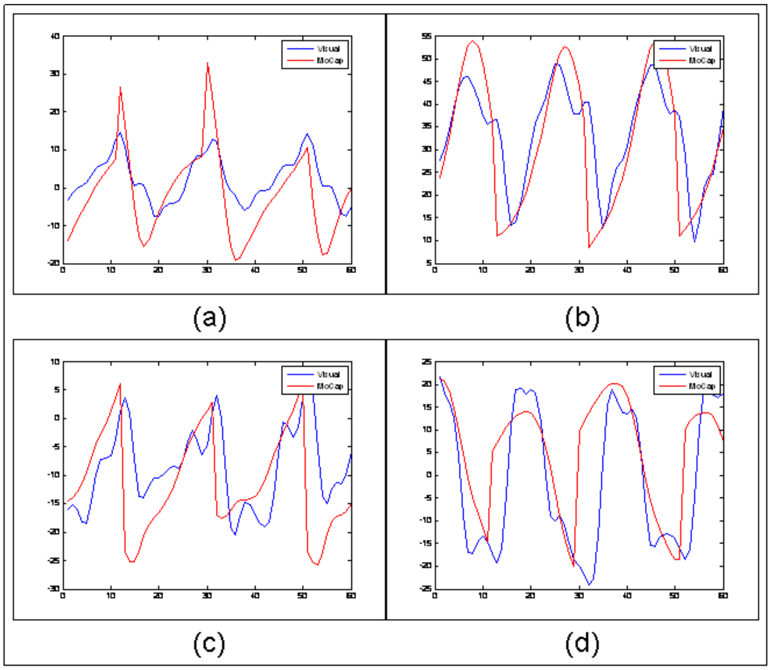


Figure 7-8: Example of Angle Estimations for (a) Front Knee, (b) Back Knee, (c) Front Hip, (d) Back Hip of an Unloaded Male Subject

Figures 7.6-7.8 demonstrate the typical angle estimates for a loaded (Figure 7.6) and unloaded (Figures 7.7 and 7.8) male subject. As can be seen in the figures, the visual angle estimations are similar for each load condition. The phase differences and errors between the visual and motion captures systems are also consistent between load and unloaded data.

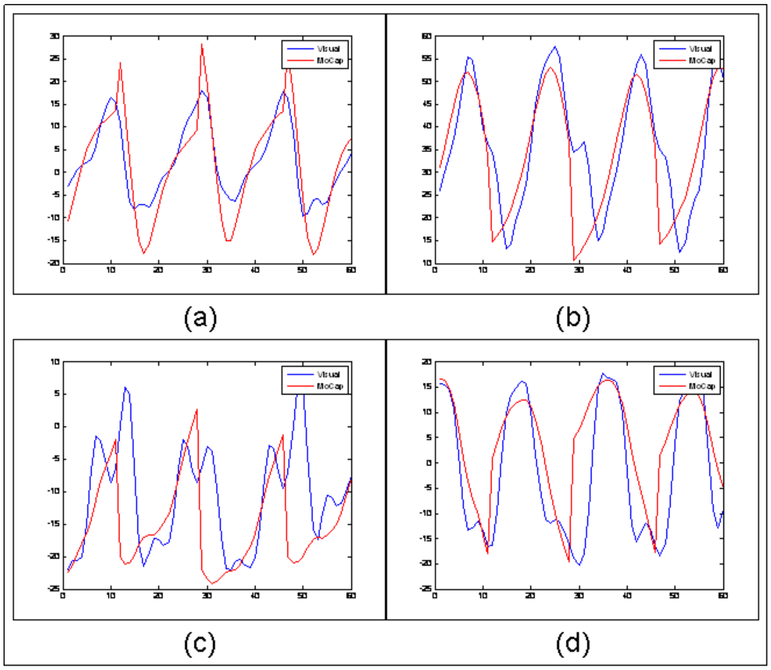


Figure 7-9: Example of Angle Estimations for (a) Front Knee, (b) Back Knee, (c) Front Hip, (d) Back Hip of an Unloaded Female Subject

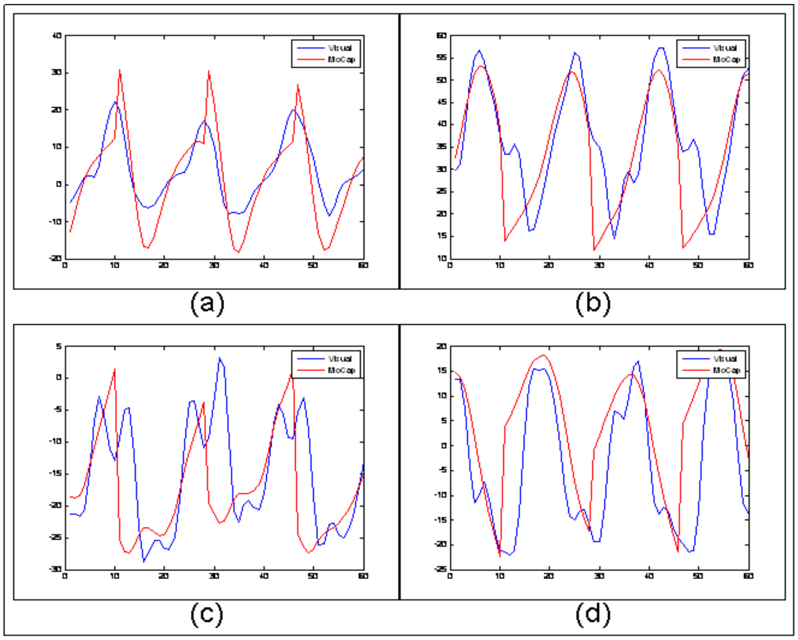


Figure 7-10: Example of Angle Estimations for (a) Front Knee, (b) Back Knee, (c) Front Hip, (d) Back Hip of an Loaded Female Subject

7.2 Classification of Hidden Load Conveyance Using Joint Angle Estimations Calculated from Image Data

A hidden load detection system based solely on joint angle analysis has been developed using motion capture data as inputs to the system [107, 108]. To accomplish this analysis a complete stride of an observed subject is used. To segment a complete stride from the image sequence, the stride length of the subject was calculated in each image. The first and third peak values were used to determine the beginning and end of a complete stride. The lower body joint angles, as calculated by the above described system, were then segmented by stride and filtered using a 7 Hz low pass, forth-order Butterworth filter [108]. The segmented and filtered angle data were used to generate phase portraits for each sequence. Phase portraits are the plot of angular position vs. angular velocity. The angular velocities were estimated in each frame by calculating the change in angle over time between measurements (1/30 sec.). The path length of each joint was then calculated by equation 7.2.

$$PL = \sum_{1}^{n} \sqrt{\left(\theta(i+1) - \theta(i)\right)^2 + \left(\dot{\theta}(i+1) - \dot{\theta}(i)\right)^2} \qquad (7.2)$$

The phase portraits generated for various load conditions and subjects are shown in Figures 7.11 and 7.12.

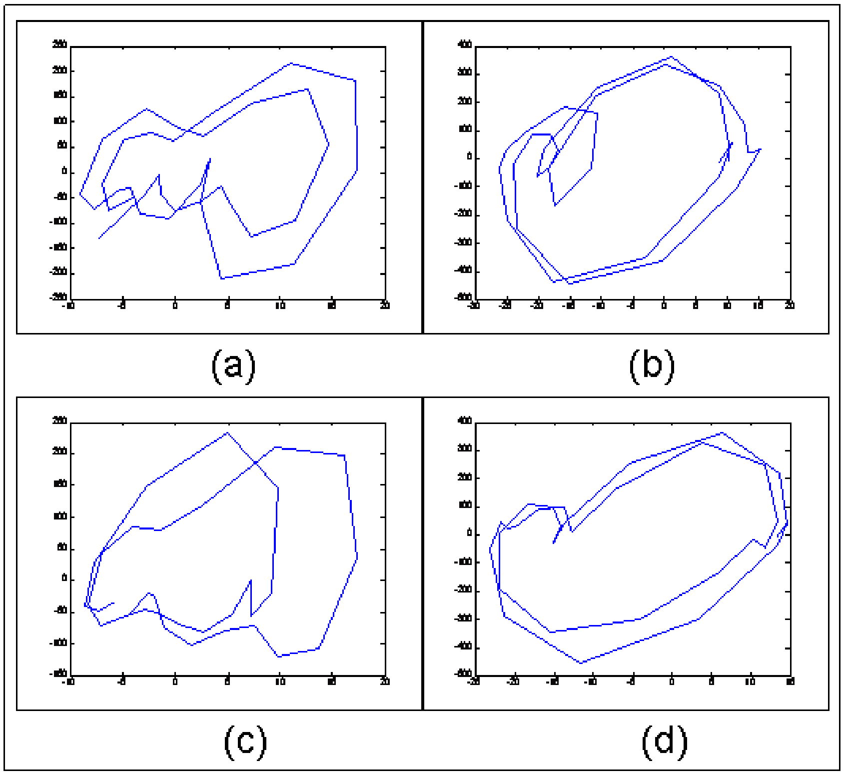


Figure 7-11: Examples of Unloaded (a) Front Knee, (b) Back Hip and Loaded (c) Front Knee, (d) Back Hip phase portraits

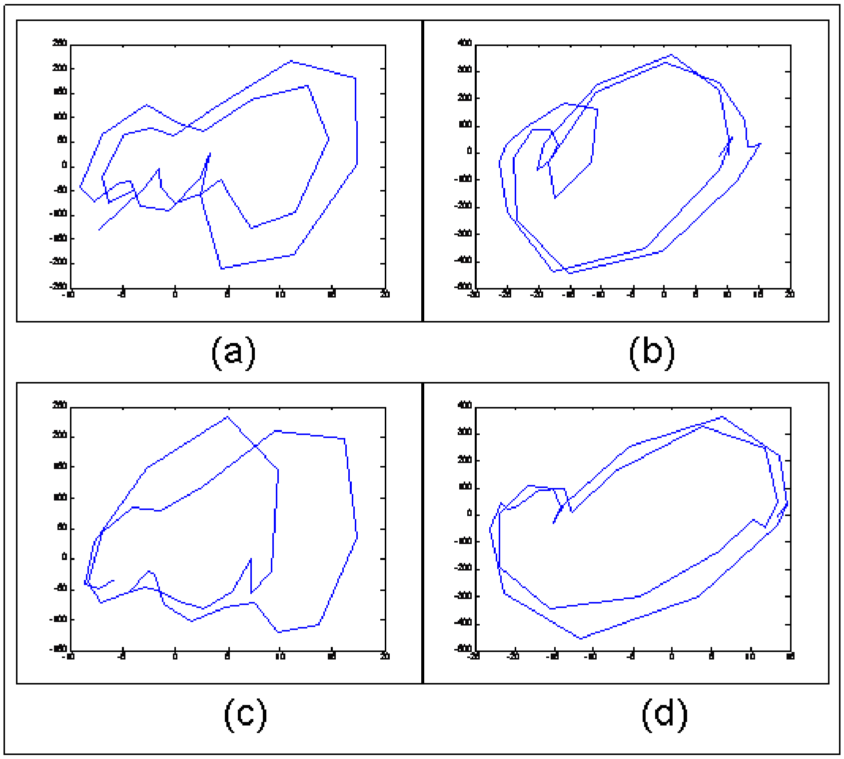


Figure 7-12: Examples of Unloaded (a) Front Knee, (b) Back Hip and Loaded (c) Front Knee, (d) Back Hip phase portraits

7.3 Experiments and Results of Hidden Load Classification System

In order to quantify the performance of the system, a hidden load detection experiment was conducted similar to that in [107]. The system was used to classify loaded and unloaded walking using 8 subjects (5 male, 3 female) of varying heights, weights and ages [107]. All the subjects used in this experiment were not used in the training and validation of the edge point classification network. These 8 subjects generated a total of 50 datasets consisting of 26 loaded and 24 unloaded image sequences. Lower body joint angles were calculated by the system for each of the 50 image sequences. Each image sequence resulted in a set of front knee, back knee, front hip and back hip angle estimates. The path lengths for each of these angles were calculated for one complete gait cycle.

To train the hidden load classifier a leave-one-subject out methodology was adopted. 7 subjects were used for training the classifier while 1 subject was used for validation. This was repeated for each subject so that the training set never contained a subject in the validation set. The final classification result found using this system is 88%. This is comparable to the results found when using a motion capture system and a similar classification scheme [107] of 92.5%. A plot of the path lengths for each dataset is shown in Figure 7.13.

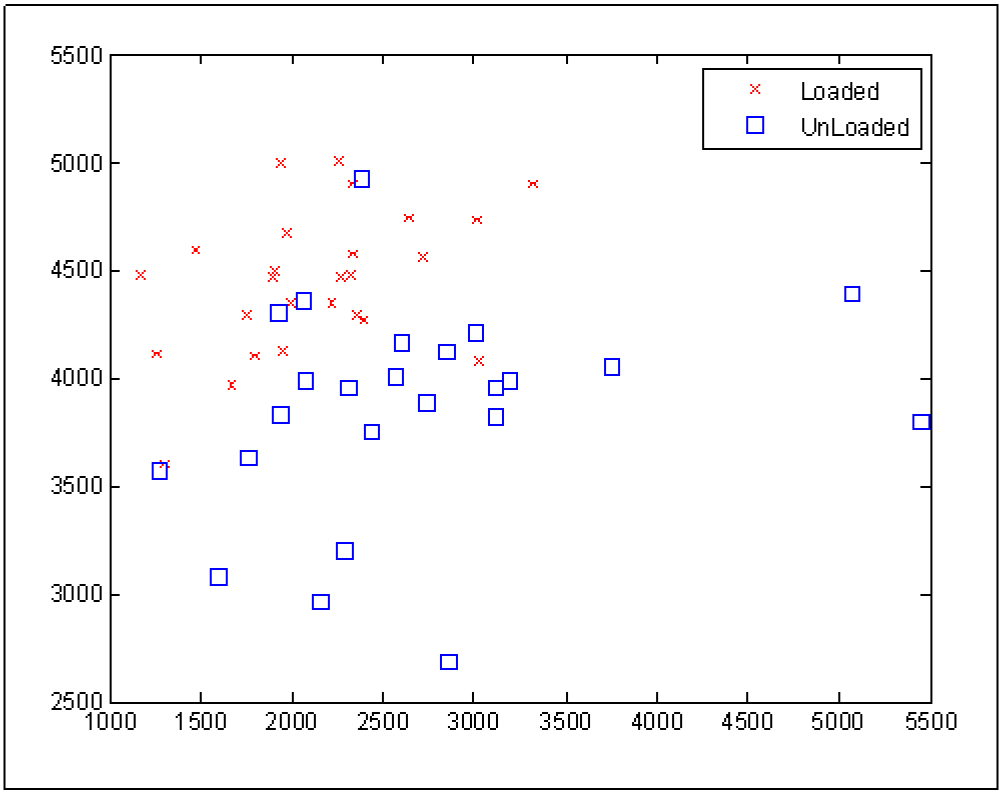


Figure 7-13: Path lengths of each dataset

Examples of correctly and misclassified phase portraits for various subjects are shown in figures 7.14-7.15.

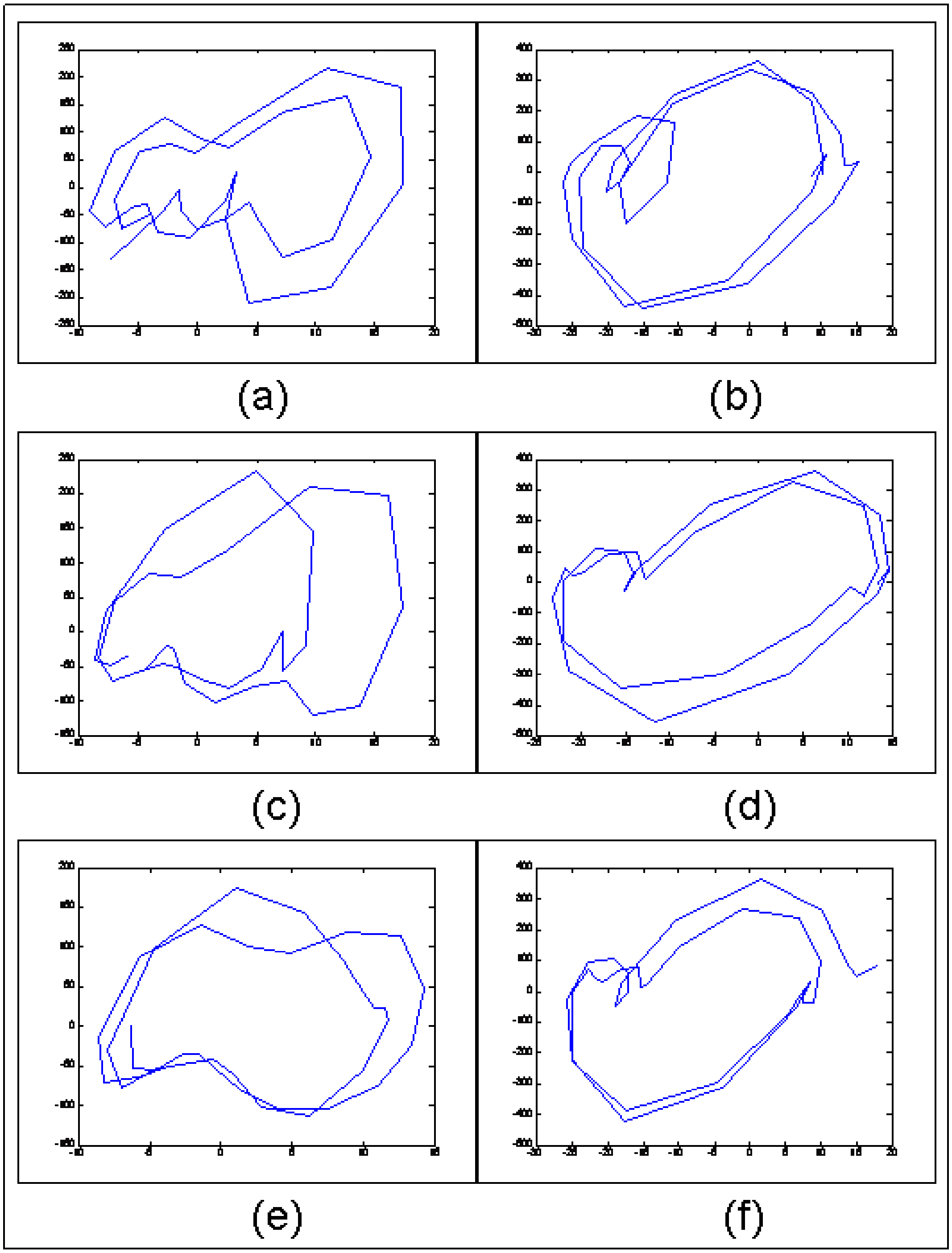


Figure 7-14: Correctly classified (a, b) Loaded and (c, d) Unloaded datasets and incorrectly classified Loaded (e, f)

Figure 7.14e and 7.14f shows an example of an incorrectly classified loaded data set. As can be seen the phase portrait of the incorrectly classified loaded set is very similar to that of an unloaded subject and is classified as unloaded. The typical loop found in the loaded phase portrait is not present as it is in Figure 7.14b.

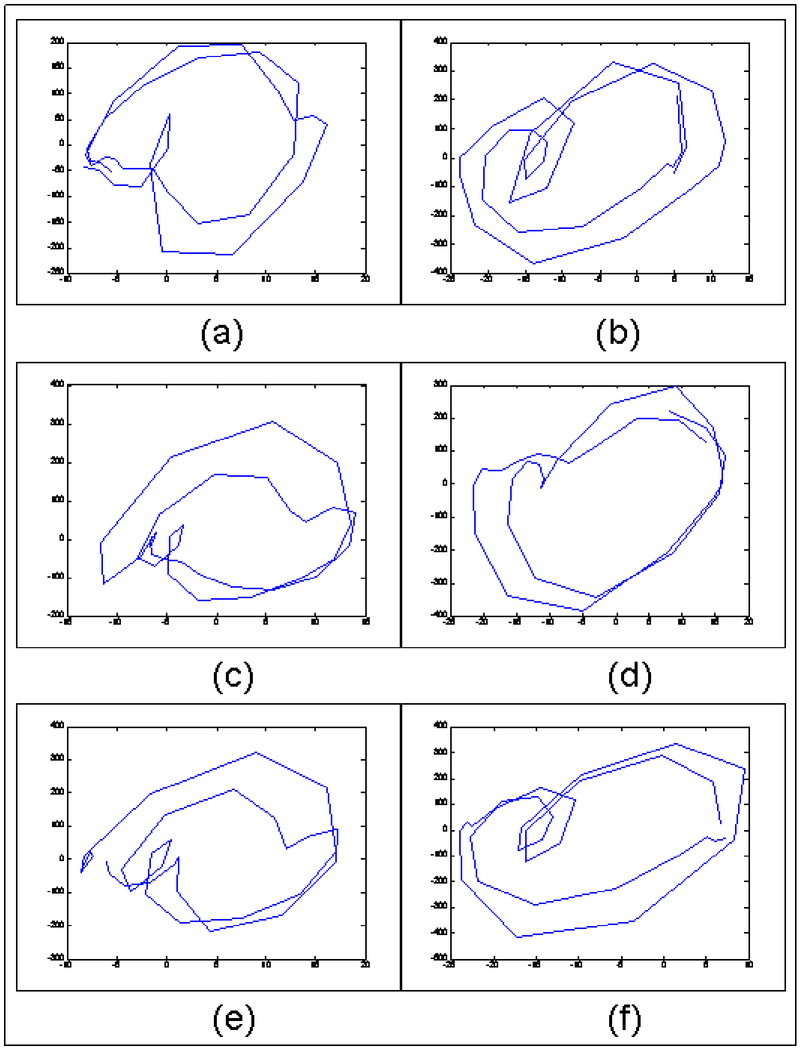


Figure 7-15: Correctly classified (a, b) Loaded and (c, d) Unloaded datasets and incorrectly classified Unloaded (e, f)

Figure 7.15e and 7.15f shows an example of an incorrectly classified unloaded data set. As can be seen the phase portrait of the incorrectly classified unloaded set is very similar to that of a loaded subject and is classified as loaded. The typical loop found in the loaded phase portrait is present in this figure as it is in Figure 7.15b.

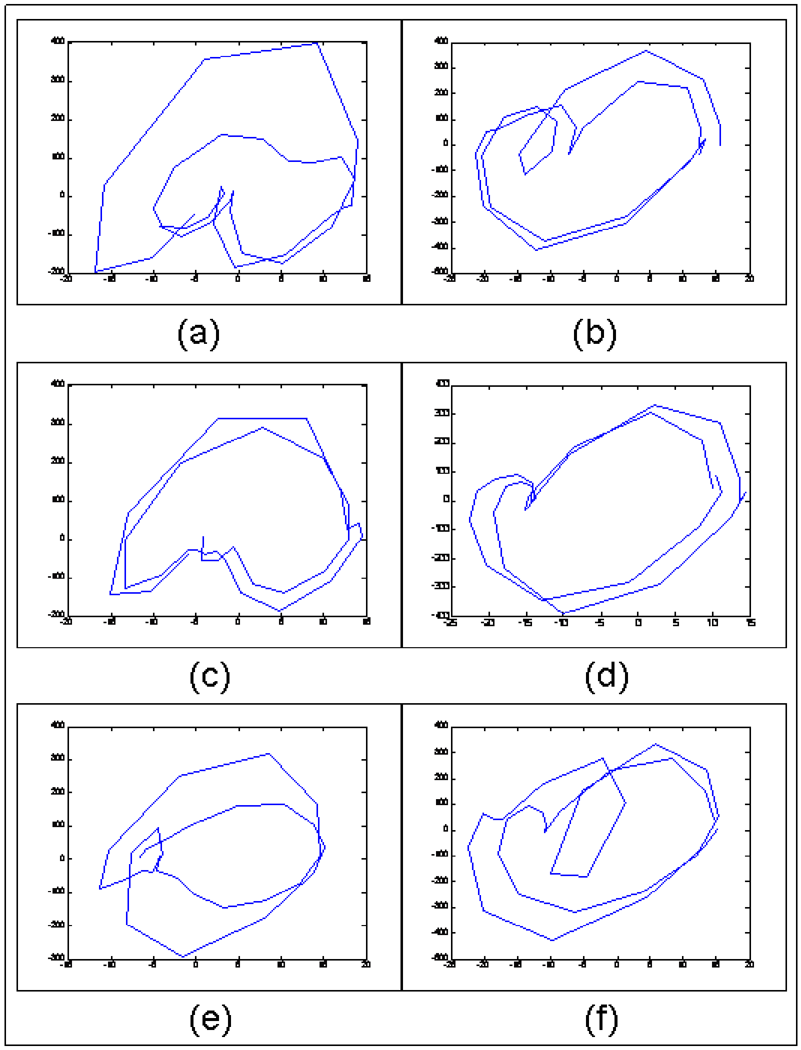


Figure 7-16: Correctly classified (a, b) Loaded and (c, d) Unloaded datasets and incorrectly classified Unloaded (e, f)

Figure 7.16 shows another example of incorrectly classified unloaded path lengths. Again, the back hip phase portrait shows the loop typically found in loaded data sets. This causes a longer path length and the subsequence misclassification of the data set.

Chapter 8: Conclusion

In this work a system which can be used to determine if human gait observed by a video camera at a remote distance is normal or abnormal is presented. The system utilizes a wide variety of different traditional research areas, including computer vision, biomechanics, and harmonic system analysis as well as pattern recognition to determine the final system output. The system described utilizes the video input to generate various gait descriptors by performing background extraction, foreground determination, and motion estimation. Each of these levels of analysis interacts with one another to produce an effective and efficient algorithm that generates a wide variety of descriptors. The data generated from the video is classified. This uses all available gait descriptors to accurately arrive at a classification of the observed gait.

Neural networks are used to create a powerful, efficient and robust classifier which can correctly classify an observed walking style using only simple image metrics. This classification was done using a three layer neural network in conjunction with a maximum operator. The image metrics provided to the classification network are simple to obtain and robust to image errors and self occlusions. With a fully trained network, the system can run quickly and efficiently to classify walking motions observed in an area by a single monocular camera. The system was shown to provide a 100% correct classification rate when classifying 5 different styles of walking.

A new vision based gait analysis system was presented which produced results comparable to those found using a motion capture system in the analysis of evenly distributed hidden loads. Silhouettes of human subjects observed in the front-to-parallel plane were extracted from image sequences. The boundary points of these silhouettes were then used as inputs to a hybrid fuzzy neural network classifier that grouped the points into kinematic part families. These part families were used to generate landmark points in order to estimate joint locations. These estimated joint locations were used to estimate lower body joint angles of observed subjects. These angles, calculated without using an obtrusive system, were shown to be close to those calculated using a high resolution, marker-based motion capture system. The performance of the system for use in gait analysis is then quantified using the limiting case of evenly distributed hidden load detection. The system was shown to yield results comparable to those found using a traditional motion capture system.

This approach to gait analysis is promising in that it does not rely on an explicit spatial model, eliminating the need for computationally expensive model fitting. Additionally, the elimination of any temporal based gait model allows for analysis of subtly different gait motions across individuals. This shows that the system has potential for use in many different forms of gait analysis including clinical and surveillance applications.

8.1 Future Work

Further application of the gait classification system to other areas of visual human analysis will be done including biometrics, abnormal activity analysis and surveillance applications. Ongoing work will focus on the generalization of such a system to multiple view points. Training the hybrid fuzzy neural network from multiple views should allow for the estimation of joint angles from multiple viewpoints. Additional viewpoint modeling could also be used in the creation of a view invariant joint angle estimation system. An expansion of the hybrid fuzzy classification network will also be done using edge points not limited to the silhouette boundary. This should allow for the estimation of upper body joint angles as well as the lower body angles.

In addition to multiple-viewpoints development of this system for use without the need for a background subtraction scheme is in development. Currently work has begun in conjunction with Micha Andriluka and Dr. Bernt Schiele at Max Planck Institute to expand the hidden load detection system for use in a pedestrian detection and tracking system. This will allow the hidden load detection system to be used without the assumption of a single subject in the scene and eliminate the need for a background subtraction step. Refinement of the kinematic classification and joint position estimations will also be done to further improve on the joint estimations when compared to ground truth motion capture data.

Appendix A: Feature Selection Values for Normal Gait Classification System

The following table shows the values for each selection of features found when using the suboptimal Forward Selection method for classification feature selection. The feature numbers correspond to the following features:

1. Hip Angle
2. Horizontal (X) Centroid Position
3. Vertical (Y) Centroid Position
4. Upper Body Vertical (Y) Motion
5. Maximum Stride Length
6. Upper Body Horizontal (X) Motion
7. Torso Angle

The table shows the Accuracy of the classification for the final system when using each of the selected variables. The first stage consists of performing classification using every combination of two variables that is possible with the given features, for stage two the combination tested above which yields the most accurate classification is selected. Each of the remaining features are then added to this combination and the classification results of the system are again calculated, the combination which yields the most accurate combination is again selected for use in the next stage. This process is repeated until the desired accuracy of classification is achieved.

Forward Selection for Classifier Feature Value Selection		
Stage	**Features Included**	**Percent of Correct Classification**
1	1, 2	54.24%
1	1, 3	89.83%
1	1, 4	93.22%
1	1, 5	77.97%
1	1, 6	55.93%
1	1, 7	89.83%
1	2, 3	0.00%
1	2, 4	83.05%
1	2, 5	32.20%
1	2, 6	71.19%
1	2, 7	8.47%
1	3, 4	77.97%
1	3, 5	30.51%
1	3, 6	83.05%
1	3, 7	38.98%
1	4, 5	77.97%
1	4, 6	74.58%
1	4, 7	81.36%
1	5, 6	76.27%
1	5, 7	28.81%
1	6, 7	28.81%
2	1, 3, 2	74.58%
2	1, 3, 4	91.53%
2	1, 3, 5	83.05%
2	1, 3, 6	77.97%
2	1, 3, 7	89.83%
3	1, 3, 4, 2	81.36%
3	1, 3, 4, 5	84.75%
3	1, 3, 4, 6	77.97%
3	1, 3, 4, 7	86.44%

Table A.1: Intermediate Steps of Feature Selection

Bibliography

[1] J. Boyd. Synchronization of Oscillations for Machine Perception of Gaits. *Computer Vision and Image Understanding*, vol. 96, p. 1, pp. 35-59, 2004.

[2] B. Horn and B. Schunck. Determining optical flow. Artificial Intelligence, vol. 17, no. 1-3, pp. 185--203, 1981.

[3] E. Davies. *Machine Vision*, 3rd edition, New York: Morgan Kaufmann, 2005.

[4] A. Barjatya. Block Matching Algorithms For Motion Estimation. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=8761&objectType=File , Mathworks File Exchange, 2005.

[5] J. Lu and M. Liou. A Simple and Efficent Search Algorithm for Block-Matching Motion Estimation. *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 429-433, April 1997.

[6] D. Gavrila. The Visual Analysis of Human Movement: A Survey. *Computer Vision and Understanding*, vol. 73, no. 1, pp. 82-98, 1999.

[7] J. Rose, and J. Gamble. *Human Walking*, 3rd edition, New York: Lippencott Williams and Wilkins, 2006.

[8] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*, 3rd edition New York: Academic Press, 2006.

[9] O. Omidvar and J. Dayhoff, ed. *Neural Networks and Pattern Recognition*, New York: Academic Press, 1998.

[10] I. Nabney. *Netlab: Algorithms for Pattern Recognition*, New York: Springer, 2002.

[11] A. Patla, ed. *Adaptability of Human Gait: Implications for the Control of Locomotion*, New York: North-Holland, 1991.

[12] E. Micheli-Tzanakou. *Supervised and Unsupervised Pattern Recognition*, New York: CRC Press, 2000.

[13] A. Elgammal, D. Harwood and L. Davis. Non-Parametric Model for Background Subtraction. *6th European Conference on Computer Vision. Dublin, Ireland, June/July 2000*.

[14] L. Wang and T. Tan. Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, December 2003.

[15] R. Zhang, C. Vogler and D. Metaxas. Human Gait Recognition. *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, vol. 1, no. 1, 2004.

[16] P. Phillips et al. The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm. *16th International Conference on Pattern Recognition*, vol 1, 2002.

[17] S. Dockstader, K. Bergkessel and A. Tekalp. Feature Extraction for the Analysis of Gait and Human Motion. *16th International Conference on Pattern Recognition*, vol 1, 2002.

[18] J. Davis and S. Taylor. Analysis and Recognition of Walking Movements. *16th International Conference on Pattern Recognition*, vol 1, 2002.

[19] S. Dockstader and N. Imennov. Prediction for Human Motion Tracking Failures. *IEEE Transactions on Image Processing*, vol. 15, no. 2, February 2006.

[20] Z. Zhang and N. Troje. View-Independent Person Identification from Human Gait. *Neurocomputing*, vol. 69, pp. 250-256, 2005.

[21] J. Han and B. Bhanu. Individual Recognition Using Gait Energy Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, Februrary 2006.

[22] R. Urtasun and P. Fua. 3D Tracking for Gait Characterization and Recognition. *Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[23] M. Shah and R. Jain, ed. *Motion-Based Recognition*, Boston: Kluwer Academic Publishers, 1997.

[24] H. Su and F. Huang. Human Gait Recognition Based on Motion Analysis. *Proc. of the 4th International Conference on Machine Learning and Cybernetics*, 18-21, August 2005.

[25] J. Saboune and F. Charpillet. Markerless Human Motion Capture for Gait Analysis. *3rd European Medical and Biological Engineering Conference* October 2005.

[26] F. Asano et al. A Novel Gait Generation for Biped Walking Robots Based on Mechanical Energy Constraint. *Proceedings of the International Conference on Intelligent Robots and Systems*, October 2002.

[27] P. Viola, M. Jones and D. Snow. Detecting Pedestrians Using Patterns of Motion and Appearance. http://www.merl.com (Mitsubishi Electric Research Laboratories, 2003).

[28] Y. Song et al. Monocular Perception of Biological Motion in Johansson Displays. *Proceedings of International Conference of Computer Vision*. 2000.

[29] F. Remondino. Tracking of Human Movements in Image Space. *Internal technical report at IGP - ETH Zurich*, April 2001.

[30] A. Bobick and J. Davis. The Recognition of Human Movement Using Temporal Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, March 2001.

[31] S. Lee and M. Crawford. Unsupervised Multistage Image Classification Using Hierarchical Clustering With a Bayesian Similarity Measure. *IEEE Transactions on Image Processing*, vol. 14, no. 3, March 2005.

[32] J. Wei. Video Content Classification Based on 3-D Eigen Analysis. *IEEE Transactions on Image Processing*, vol. 14, no. 5, May 2005.

[33] A. Batur and M. Hayes. Adaptive Active Appearance Models. *IEEE Transactions on Image Processing*, vol. 14, no. 11, November 2005.

[34] V. Ha and J. Moura. Affine-Permutation Invariance of 2-D Shapes. *IEEE Transactions on Image Processing*, vol. 14, no. 11, November 2005.

[35] C. Barron and I. Kakadiaris. Estimating Anthropometry and Pose from a Single Image. *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2000.

[36] C. Barron and I. Kakadiaris. On the Improvement of Anthropometry and Pose Estimation from a Single Uncalibrated Image. *Proceedings, Workshop on Human Motion*, 2000.

[37] O. Williams, A. Blake and R. Cipolla. A Sparse Probabilistic Learning Algorithm for Real-Time Tracking. *Proceedings, 9th IEEE International Conference on Computer Vision*, 2003.

[38] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[39] M. Black and A. Jepson. EigenTracking: Robust Matching and Tracking of Articulated Objects using a View-Based Representation. *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63-84, 1998.

[40] S. Krinidis and I. Pitas. Fast Free-Vibration Modal Analysis of 2-D Physics-Based Deformable Objects. *IEEE Transactions on Image Processing*, vol 14, no 3, March 2005.

[41] M. Schlemmer et al. Fast Clifford Fourier Transformation for Unstructured Vector Field Data. *Proceedings International Conference on Numerical Grid Generation in Computational Field Simulations*, 2005.

[42] C. Bregler. Learning and Recognizing Human Dynamics in Video Sequences. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[43] S. Wachter and H. Nagel. Tracking of Persons in Monocular image Sequences. *Proceedings IEEE Workshop on Articulated and Non-Rigid Motion*, 1997.

[44] S. Singh and K. Waldron. Motion Estimation by Optical Flow and Inertial Measurements for Dynamic Legged Locomotion. *IEEE Conference on Robotics and Automation*, 2005.

[45] C. Bauckhage, J. Tsotsos and F. Bunn. Detecting Abnormal Gait. *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, 2005.

[46] M. Kohle, D. Merkl and J. Kastner. Clinical Gait Analysis by Neural Networks: Issues and Experiences. *Proceedings of the IEEE Symposium on Computer-Based Medical Systems*, pp. 138-143, 1997.

[47] A. Dalen, J. Nilsson and A. Thorstensson. Factors influencing a prolonged foot march. Stockholm Sweden: *Karolinska Institute*, FOA Report C50601-H6, 1978.

[48] J. Knapik et al. Injuries associated with strenuous road marching. *Military Medicne,* 157, 64-67, 1992.

[49] H. Kinoshita. Effects of different loads and carrying systems on selected biomechanical parameters describing walking gait. *Ergonomics, 28*(9), 1347-62, 1985.

[50] Y. Hong and G. Brueggemann. Changes in gait patterns in 10-year-old boys with increasing loads when walking on a treadmill. *Gait and Posture, 11*, 254-9, 2000.

[51] J. Knapik, E. Harman and K. Reynolds. Load carriage using packs: a review of physiological, biomechanical and medical aspects. *Appl Ergon, 27*(3), 207-16, 1996.

[52] M. Wittman et al. Visual analysis of the effects of load carriage on gait. *Proc SPIE, 5779*, 15-22, 2005.

[53] A. Must and R. Strauss. Risks and consequences of childhood and adolescent obesity. *Int J Obes Relat Metab Disord*, Vol. 23, S2-S11, 1999.

[54] B. McGraw et al. Gait and posture stability in obese and nonobese prepubertal boys. *Archives of Physical Medicine and Rehabilitation, Vol. 81*, 484-489, 2000.

[55] S. Messier. Osteoarthritis of the knee and associated factors of age and obesity: effects on gait. *Medicine and Science in Sports and Exercise, Vol. 26*, 1446-1452, 1994.

[56] S. Messier. Obesity: effects on gait in an osteoarthritic population. *Journal of applied Biomechanics, Vol. 12*, 161-172, 1996.

[57] A. Hills. Gait characteristics of obese children. *Arch Phys Med Rehabil*, 72, 403-407, 1991.

[58] D. Hall. What use is the BMI. *Archives of Disease in Childhood, 91*, 283-286, 2006.

[59] J. Haddad. Adaptations in interlimb and intralimb coordination to asymmetrical loading in human walking. *Gait and Posture, 23*, 429-434, 2006.

[60] T. Prieto. Measures of postural steadiness: differences between healthy young and elderly adults. *IEEE Trans Biomed Eng*, Vol. 43(9), 956-966, 1996.

[61] D. Chow et al. The effect of backpack load on the gait of normal adolescent girls. *Ergonomics*, Vol. 48(6), 642-656, 2005.

[62] S. Kim, T. Lockhart and H. Yoon. Relationship between age-related gait adaptations and required coefficient of friction. *Safety Science*, Vol. 43, 425-436, 2003.

[63] J. Mickelborough et al. Validity and reliability of a kinematic protocol for determining foot contact events. *Gait and Posture*, Vol. 11, 32-37, 2000.

[64] E. Harman et al. The effects of backpack weight on the biomechanics of load carriage. *US Army Research Institute of Environmental Medicine Military Performance Division Technical Report T-00-17*, Natick, MA, 2000.

[65] M. LaFiandra et al. How do load carriage and walking speed influence trunk coordination and stride parameters? *J Biomech* 36, 87-95, 2002.

[66] M. Kurz and N. Stergiou. Effect of normalization and phase angle calculations on continuous relative phase. *J Biomech*, Vol. 35, 369-374, 2002.

[67] T. Griffin, T. Roberts and R. Kram. Metabolic cost of generating muscular force in human walking: insights from load-carrying and speed experiments. *J Appl Physiol*, Vol. 95, 172-183, 2003.

[68] M. Whittle. Gait analysis: an introduction, 4th edition. Butterworth-Heinemann, 2007.

[69] R. Browning et al. Effects of obesity and sex on the energetic cost and preferred speed of walking. *J Appl Phsiol*, Vol. 100, 390-398, 2006.

[70] G. Foster et al. The energy cost of walking before and after significant weight loss. *Med Sci Sports Exerc*, Vol. 27, 888-894, 1995.

[71] S. Lazzer et al. Assessment of energy expenditure associated with physical activities in free-living obese and nonobese adolescents. *Am J Clin Nut*r, Vol. 78, 471-479, 2006.

[72] F. Berrigan et al. Influence of obesity on accurate and rapid arm movement performed from a standing posture. *Int J Obes* (Lond), Vol. 30, 1750-1757, 2006.

[73] P. DeVita and T. Hortobagyi. Obesity is not associated with increased knee joint torque and power during level walking. *J Biomech*, Vol. 36, 1355-1362, 2003.

[74] P. Spyropoulos et al. Biomechanical gait analysis in obese men. *Arch Phys Med Rehabil*, Vol. 72, 1065-70, 1990.

[75] K. Jones and K. Barker. Human movement explained. London: Woburn: Butterwoth Heinemann, 1996:1996.

[76] J. Scholz. Dynamic pattern theory – some implications for therapeutics. *Physical Therapy*, 7012, pp. 827-843.

[77] D. Gravrila and V. Philomin. Real-time object detection for smart vehicles, *IEEE International Conference on Computer Vision (ICCV)*, pp. 87-93, 1999.

[78] Z. Zhou et al. Bayesian Framework for Extracting Human Gait Using Strong Prior Knowledge. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1738-1752, 2006.

[79] A. Elgammal and C. Lee. Nonlinear Manifold Learning for Dynamic Shape and Dynamic Appearance. *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 31-46, 2007.

[80] I. Haritaoglu, D. Harwood and L. Davis. W4: Real-Time Surveillance of People and Their Activities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809-830, 2000.

[81] I. Jolliffe. *Principle Component Analysis*, Springer, Berlin, 1986.

[82] H. Wang, P. Li and T. Zhang. Histogram feature-based Fisher Linear Discriminant for Face Detection. *Neural Comput & Applic*, vol. 17, pp. 49-58, 2008.

[83] R. Fablet and M. Black. Automatic Detection and Tracking of Human Motion with a View-Based Representation. *Proc. European Conf. on Computer Vision*, vol. 2350, pp. 476-491, 2002.

[84] M. Brand. Shadow Puppetry. *Proc. Int'l Conf. on Computer Vision*, vol. 2, p. 1237, 1999.

[85] R. Rosales and R. Scaloroff. Specialized Mappings and the Estimation of Human Body Pose from a Single Image. *Workshop on Human Motion*, pp. 19-24, 2000.

[86] J. Deutscher, A. Blake and I. Reid. Articulated body motion capture by annealed particle filtering. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 126-133, 2000.

[87] D. Beymer and T. Poggio. Image Representations for Visual Learning. *Science*, vol. 272, p. 5250, 1996.

[88] V.F. Leavers, The dynamic generalized Hough transform: its relationship to the probabilistic Hough transform and an application to the concurrent detection of circles and ellipses, *Image Understanding*, vol. 56, pp. 381-398, 1992.

[89] A. Coolen, R. Kuhn and P. Sollich. *Theory of Neural Information Processing Systems*, New York, Oxford University Press, 2005.

[90] T. Horprasert, D. Harwood and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection, *ICCV Frame-Rate Workshop*, pp.1-19, 1999.

[91] D. Hogg. Model-Based Vision: A Program to See a Walking Person. *Image and Vision Computing*, vol. 1, no. 1, pp. 5-20, 1983.

[92] J. Aggarwal and Q. Cai. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 428-440, 1999.

[93] X. Lan, and D. Huttenlocher. A Unified Spatio-Temporal Articulated Model for Tracking. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 722-729, 2004.

[94] J. Zhang, R. Collins and Y. Liu. Representation and Matching of Articulated Shapes. *Proc. 2004 IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 342-349, 2004.

[95] E. Aguiar et al. Marker-less Deformable Mesh Tracking for Human Shape and Motion Capture. *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 1-8, 2007.

[96] B. Barrois and C. Wohler. 3D Pose Estimation Based on Multiple Monocular Cues. *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 1-8, 2007.

[97] Z. Zhang and N. Troje. View-independent Person Identification from Human Gait. *Neurocomputing*, vol. 69, pp. 250-256, 2005.

[98] H. Lakany. A Generic Kinematic Pattern for Human Walking. *Neurocomputing*, vol. 35, pp. 27-54, 2000.

[99] M. Hagan and M. Menhaj. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Trans. on Neural Networks*, vol. 5, no. 6, pp. 989-993, 1994.

[100] J.G. Barton, A. Lees, An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams, Gait and Posture 5 (1997) 28-33.

[102] R. Lafuente, J.M. Belda, J. Sanchez-Lacuesta, C. Soler, J. Prat, Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis, Clinical Biomechanics 13 (3) (1997) 216-229.

[103] A.P. Hills, A.W. Parker, Gait characteristics of obese children, Archives of Physical Medicine and Rehabilitation (72) (1991) 403-407

[104] S.P. Messier, W.H. Ettinger, T.E. Doyle, T. Morgan, K.K. James, M.L. O'toll, K. Burns, Obesity: effects on gait in an osteoarthritic population, Journal of Applied Biomechanics, 12 (1996) 161-172.

[105] D.C. Kerrigan, M.K. Todd, U.D. Croce, Gender differences in joint biomechanics during walking, American Journal of Physical Medicine and Rehabilitation, 77 (1998) 2-7.

[106] J.M. Haddad, R.E. van Emmerick, S.N. Whittlesey, J. Hamill, Adaptations in interlimb and intralimb coordination to asymmetrical loading in human walking, Gait and Posture 23 (2006) 429-434.

[107] M. Lee, M. Roan, B. Smith, T.E. Lockhart, Gait analysis to classify external load conditions using linear discriminant analysis, Human Movement Science, 28 (2009) 226-235.

[108] M. Lee, M. Roan, B. Smith, An application of principle component analysis for lower body kinematics between loaded and unloaded walking, Journal of Biomechanics (2009), Vol. 42 14, pp. 2226-2232.

[109] N.T. Nguyen, D.Q. Phung, S. Venkatesh, H. Bui, Learning and Detecting Activities from Movement Trajectories Using the Hierarchical Hidden Markov Model, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[110] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as Space-Time Shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (12) (2007) 2247-2253.

[111] C. Bauckhage, J.K. Tsotsos, F.E. Bunn, Automatic detection of abnormal gait, Image and Vision Computing, 27 (1-2) (2009) 108-115.

[112] R. Navaratnam, A. Thayananthan, P.H.S. Torr, R. Cipolla, Hierarchical part-based human body pose estimation, Proc. British Machine Vision Conference, 2005.

[113] V. Parameswaran, R. Chellappa, View independent human body pose estimation from a single perspective image, in: IEEE Conference on Computer Vision and Pattern Recognition, 2004.

[114] M.W. Lee, R. Nevatia, Body part detection for human pose estimation and tracking, in: IEEE Workshop on Motion and Video Computing, 2007.

[115] A. Mittal, L. Zhao, L.S. Davis, Human body pose estimation using silhouette shape analysis, in: IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.

[116] F. Jean, R. Bergevin, A.B. Albu, Trajectories normalization for viewpoint invariant gait recognition, in: Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, USA, 2008.

[117] F. Jean, A.B. Albu, R. Bergevin, Towards view-invariant gait modeling: Computing view-normalized body part trajectories, Patter Recognition 42 (2009) 2936-2949.

[118] J. Zhang, S. Gong, Action categorization with modified hidden conditional random field, 43 (2010) 197-203.

[119] C. McIntosh et al. Human limb delineation and joint position recovery using localized boundary models. *IEEE Workshop on Motion and Video Computing,* 2007.

[120] M.W. Lee, R. Nevatia, Human pose tracking in monocular sequence using multi-level structured models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (1) (2009) 27-38.

[121] D. Knossow, R. Ronfard, R. Horaud, Human motion tracking with kinematic parameterization of extremal contours, International Journal of Computer Vision, 79 (3) (2008) 247-269.

[122] G. Mori, J. Malik, Recovering 3D human body configurations using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (7) (2006) 1052-1062.

[123] B. Bhanu, J. Han, Individual recognition using gait energy image, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2) (2006) 316-322.

[124] X. Zou, B. Bhanu, Human activity classification based on gait energy image and coevolutionary genetic programming, Proceedings of the 18th International Conference on Pattern Recognition, 3 (2006) 556-559.

[125] C.S. Lee, A. Elgammal, Dynamic shape outlier detection for human locomotion, Computer Vision and Image Understading, 113 (2009) 332-344.

[127] M. Gupta, L. Jin, N. Homma, Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory, John Wiley & Sons, New Jersey, 2003.

[128] M. Lee and I. Cohen. Human Upper Body Pose Estimation in Static Images, *European Conference on Computer Vision (ECCV)*, pp. 126–138, 2004.

[129] S.-I. Lee, V. Ganapathi and D. Koller. Efficient Structure Learning of Markov Networks using L1-Regularization, *Advances in Neural Information Processing Systems (NIPS)*, 2006.

[130] F. Lerasle,G. Rives andM. Dhome. Tracking of Human Limbs byMultiocular Vision, *Computer Vision and Image Understanding*,Vol. 75, Issue 3, pp. 229–246, 1999.

[131] R. Li, M.-H. Yang, S. Sclaroff and T.-P. Tian. Monocular Tracking of 3D Human Motion with a Coordinated Mixture of Factor Analyzers, *European Conference on Computer Vision (ECCV)*, Vol. 2, pp. 137–150, 2006.

[132] Y. Li, S. Ma and H. Lu. Human Posture Recognition Using Multi-Scale Morphological Method and Kalman Motion Estimation, *International Conference on Pattern Recognition (ICPR)*, Vol. 1, pp. 175–177, 1998.

[133] Y. Li, A. Hilton and J. Illingworth. A relaxation algorithm for real-time multiview 3d-tracking. *Image and Vision Computing*, 20(12):841–59, 2002.

[134] T. Lindeberg. Feature detection with automatic scale selection, *International Journal of Computer Vision (IJCV)*, Vol. 30, No. 2, pp. 77–116, 1998.

[135] D. G. Lowe. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision (IJCV)*, Vol. 60, No. 2, pp. 91–110, 2004.

[136] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking, *European Conference on Computer Vision (ECCV)* Vol. 2, pp. 3–19, 2000,

[137] D. J. C. MacKay. Information Theory, Inference, and Learning Algorithms, *Cambridge University Press*, Cambridge, U.K., 2003.

[138] D. J. C. MacKay. Introduction to Monte Carlo methods, *Learning in GraphicalModels*, pp. 175–204, 1999.

[139] D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three dimensional structure, *Proceedings of the Royal Society of London B*, Vol. 200, pp. 269–294, 1978.

[140] S. McKenna, S. Jabri, Z. Duric and H. Wechsler. Tracking interacting people, *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000)*, pp. 348–358, 2000.

[141] N. Metropolis and S. Ulam. The Monte Carlo method, *Journal of the American Statistical Association*, Vol. 44(247), pp. 335–341, September 1949.

[142] I. Miki, M. Trivedi, E. Hunter and P. Cosman. Human Body Model Acquisition and Motion Capture Using Voxel Data, *Proceedings of the Second InternationalWorkshop on Articulated Motion and Deformable Objects (AMDO)*, pp. 104–118, 2002.

[143] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 19, No. 7, pp. 696–710, July 1997.

[144] A. Mohan, C. Papageorgiou and T. Poggio. Example-based object detection in images by components, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 23, No. 4, pp. 349–361, April 2001.

[145] P. Moreels, M. Maire and P. Perona. Recognition by probabilistic hypothesis construction, *European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 55–68, 2004.

[146] G.Mori. Guiding Model Search Using Segmentation, *IEEE International Conference on Computer Vision (ICCV)*, Vol. 2, pp. 1417–1423, 2005.

[147] G. Mori, X. Ren, A. Efros and J. Malik. Recovering Human Body Configurations: Combining Segmentation and Recognition, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 326–333, 2004.

[148] G. Mori and J. Malik. Estimating Human Body Configurations using Shape Context Matching, *European Conference on Computer Vision (ECCV)*, Vol. 3, pp. 666–680, 2002.

[149] D. Morris and J. Rehg. Singularity Analysis for Articulated Object Tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 289–296, 1998.

[150] L. Muendermann, S. Corazza and T. Andriacchi. Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[151] L. Mundermann, S. Corazza, A. M. Chaudhari, T. P. Andriacchi, A. Sundaresan and R. Chellappa. Measuring human movement for biomechanical applications using markerless motion capture, *IS&T/SPIE 18th Annual Symposium: Electronic Imaging*, San Jose, California, 2006.

[152] K. Murphy, A. Torralba and W. Freeman. Using the forest to see the trees: A graphical model relating features, objects, and scenes, *Advances in Neural Information Processing Systems 16*, 2003.

[153] K. Murphy and S. Russel. Rao-blackwellized particle filtering for dynamic bayesian networks, *Sequential Monte Carlo Methods in Practice*, pp. 499–515, Springer, 2001.

[154] C. Musso, N. Oudjane and F. LeGland. Improving regularized particle filters, *SequentialMonte Carlo Methods in Practice*, Springer-Verlag, 2001.

[155] R. Neal and G. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. *Learning in Graphical Models*, pp. 355–368.MIT Press, 1999.

[156] M. Niskanen, E. Boyer and R. Horaud. Articulated Motion Capture from 3-D Points and Normals, *British Machine Vision Conference (BMVC)*, Vol. 1, pp. 439–448, 2005.

[157] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little and D. Lowe. A Boosted Particle Filter: Multitarget Detection and Tracking, *European Conference on Computer Vision (ECCV)*, Vol. 1, 28–39, 2004.

[158] A. Opelt, A. Pinz and A. Zisserman.A Boundary-Fragment-Model for Object Detection, *European Conference on Computer Vision (ECCV)*, Vol. 2, pp. 575–588, 2006.

[159] R. Gross and J. Shi. The CMU motion of body (MoBo) database. Robotics Institute, Carnegie Mellon University, 2001.

[160] D. Ormoneit, M. J. Black, T. Hastie, and H. Kjellstr¨om. Representing cyclic human motion using functional analysis. *Image and Vision Computing*, 23(14):1264–1276,Dec. 2005.

[161] C. Papageorgiou and T. Poggio. A Trainable System for Object Detection, *International Journal of Computer Vision (IJCV)*, Vol. 38, No. 1, pp. 15–33, June 2000.

[162] S. Park and J. K. Aggarwal. A Hierarchical Bayesian Network for Event Recognition of Human Actions and Interactions, *ACM Journal of Multimedia Systems*, 10(2), pp. 164–179, 2004.

[163] S. Park and J. K. Aggarwal. Simultaneous Tracking of Multiple Body Parts of Interacting Persons, *Computer Vision and Image Understanding*,Vol. 102, Issue 1, pp. 1–21, April 2006.

[164] V. Pavolvi´c, J. Rehg, T-J. Cham and K. Murphy. A dynamic Bayesian network approach to figure tracking using learned dynamic models, *IEEE International Conference on Computer Vision (ICCV)*, pp. 94–101, 1999.

[165] R. Plankers and P. Fua. Articulated Soft Objects for Video-based Body Modeling, *IEEE International Conference on Computer Vision (ICCV)*, pp. 394–401, 2001.

[166] R.W. Poppe andM. Poel. Comparison of Silhouette Shape Descriptors for Example-basedHuman Pose Recovery, *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2006)*, pp. 541–546, 2006.

[167] R. Quandt and J. Ramsey.A new approach to estimating switching regressions, *Journal of the American Statistical Society*, Vol. 67, pp. 306–310, 1972.

[168] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models, *IEEE ASSP Magazine*, pp. 4–15, January 1986.

[169] D. Ramanan, D. Forsyth and A. Zisserman. Strike a Pose: Tracking People by Finding Stylized Poses, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 271–278, 2005.

[170] D. Ramanan,D. Forsyth and K. Barnard. Detecting, Localizing, and Recovering Kinematics of Textured Animals, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 635–642, 2005.

[171] D. Ramanan and D. Forsyth. Automatic Annotation of Everyday Movements, *Neural Information Processing Systems (NIPS)*, 2003.

[172] D. Ramanan and D. Forsyth. Using temporal coherence to build models of animals, *IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 338–346, 2003.

[173] D. Ramanan and D. Forsyth. Finding and tracking people from the bottom up, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 467–474, 2003.

[174] X. Ren, A. Berg and J. Malik. Recovering Human Body Configurations using Pairwise Constraints between Parts, *IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 824–831, 2005.

[175] C. P. Robert and G. Casella.Monte Carlo Statistical Methods (2nd edition), *Springer-Verlag*, 2004.

[176] T. Roberts, S. McKenna and I. Ricketts. Human pose estimation using learnt probabilistic region similarities and partial configurations, *European Conference on Computer Vision (ECCV)*, Vol. 4, pp. 291–303, 2004.

[177] J. Rodgers, D. Anguelov, H.-C. Pang and D. Koller. Object Pose Detection in Range Scan Data, *IEEE ComputeSociety Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 2445–2452, 2006.

[178] R. Ronfard, C. Schmid and B. Triggs. Learning to parse pictures of people, *European Conference on ComputerVision (ECCV)*, Vol. 4, pp. 700–714, 2002.

[179] R. Rosales and S. Sclaroff. Learning Body Pose Via Specialized Maps, *Neural Information Processing Systems (NIPS)*, 2002.

[180] R. Rosales and S. Sclaroff. Learning and Synthesizing Human Body Pose and Motion, *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000)*, 2000.

[181] R. Rosales and S. Sclaroff. Inferring Body Pose without Tracking Body Parts, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 721–727, 2000.

[182] B. Rosenhahn, U. Kersting, K. Powell, R. Klette, G. Klette and H.-P. Seidel. A system for articulated tracking incorporating a clothing model, *Machine Vision and Applications (MVA)*, Vol. 18, No. 1, pp. 25–40, 2007.

[183] B. Rosenhahn, T. Brox, U. Kersting, D. Smith, J. Gurney and R. Klette. A system for marker-less human motion estimation, *Kuenstliche Intelligenz (KI)*, No. 1, pp. 45–51, 2006.

[184] B. Rosenhahn, U. Kersting, D. Smith, J. Gurney, T. Brox and R. Klette. A system for marker-less human motion estimation, *Symposium of the German Association for Pattern Recognition (DAGM)*, Springer-Verlag, LNCS 3663, pp. 230–237, 2005.

[185] S. Roth, L. Sigal and M. J. Black. Gibbs likelihoods for Bayesian tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 886–893, 2004.

[186] S. R. Sain. Multivariate locally adaptive density estimation, *Source Computational Statistics & Data Analysis archive*, Vol. 39(2), pp. 165–186, April, 2002.

[187] M. Siddiqui and G. Medioni Robust Real-Time Upper Body Limb Detection and Tracking, *ACM International Workshop on Video Surveillance & Sensor Networks (VSSN)*, 2006.

[188] R. E. Schapire. Theoretical Views of Boosting and Applications, *Algorithmic Learning Theory*, 1999.

[189] G. Shakhnarovich, P. Viola and T. Darrell. Fast Pose Estimation with Parameter Sensitive Hashing, *IEEE Intlernational Conference on Computer Vision (ICCV)*, Vol. 2, pp. 750–757, 2003.

[190] J. Shi and C. Tomasi. Good features to track, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600, 1994.

[191] K. Shoemake.Animating Rotations with Quaternion Curves, *Computer Graphics (SIGGRAPH)*, Vol. 19, pp. 245–254, 1985.

[192] H. Sidenbladh and M. J. Black. Learning the statistics of people in images and video, *International Journal of Computer Vision (IJCV)*, Vol. 54, No. 1–3, pp. 183–209, 2003.

[193] H. Sidenbladh, M. J. Black and D. Fleet. Stochastic tracking of 3D human figures using 2D image motion, *European Conference on Computer Vision (ECCV)*, Vol. 2, pp. 702–718, 2000.

[194] L. Sigal and M. J. Black. HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion, *Techniacl Report:* CS-06-08, Brown University, 2006.

[195] L. Sigal and M. J. Black. Predicting 3D People from 2D Pictures, *IV Conference on Articulated Motion and Deformable Objects (AMDO)*, 2006.

[196] L. Sigal and M. J. Black. Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 2041–2048, 2006.

[197] L. Sigal, S. Bhatia, S. Roth, M. J. Black and M. Isard. Tracking Loose-limbed People, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 421–428, 2004.

[198] L. Sigal, Y. Zhu, D. Comaniciu and M. J. Black. Tracking Complex Objects using Graphical Object Models, *1st InternationalWorkshop on Complex Motion*, 2004.

[199] L. Sigal, M. Isard, B. H. Sigelman and M. J. Black. Attractive people: Assembling loose-limbed models using non-parametric belief propagation, *Neural Information Processing Systems (NIPS)*, Vol. 16, pp. 1539–1546, 2003.

[200] R. Silva and R. Scheines. Bayesian learning of measurement and structural models, *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pp. 825–832, 2006.

[201] R. Silva, R. Scheines, C. Glymour and P. Spirtes. Learning the structure of linear latent variable models, *Journal of Machine Learning Research*,Vol. 7, pp. 191–246, February 2006.

[202] B.W. Silverman. Density Estimation for Statistics and Data Analysis, *Chapman & Hall*, London, 1986.

[203] S. Sinha,M. Pollefeys and L. McMillan. Camera Network Calibration from Dynamic Silhouettes, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 195–202, 2004.

[204] J. Sivic, B. Russell, A. Efros, A. Zisserman and W. Freeman. Discovering object categories in image collections, *IEEE International Conference on Computer Vision (ICCV)*, 2005.

[205] C. Sminchisescu,A. Kanajujia and D. Metaxas. Learning Joint Top-Down and Bottom-up Processes for 3D Visual Inference, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 1743–1752, 2006.

[206] C. Sminchisescu, A. Kanaujia, Z. Li and D. Metaxas. Discriminative Density Propagation for 3D Human Motion Estimation, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 390–397, 2005.

[207] C. Sminchisescu, A. Kanaujia, Z. Li and D. Metaxas. Learning to Reconstruct 3D Human Motion from Bayesian Mixtures of Experts: A Probabilistic Discriminative Approach, *TechnicalReportCSRG-502*, University of Toronto, 2004.

[208] C. Sminchisescu and A. Jepson.Variational mixture smoothing for non-linear dynamical systems, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 608–615, 2004.

[209] C. Sminchisescu and B. Triggs. Kinematic Jump Processes for Monocular 3D Human Tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 69–76, 2003.

[210] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling, *International Journal of Robotics Research*,Vol. 22, No. 6, pp. 371–393, 2003.

[211] C. Sminchisescu and B. Triggs. Hyperdynamics Importance Sampling, *European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 769–783, 2002.

[212] C. Sminchisescu and B. Triggs. Building Roadmaps of LocalMinima of Visual Models. *European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 566–582, 2002.

[213] P. Srinivasan and J. Shi. Bottom-up Recognition and Parsing of the Human Body, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[214] R. Srinivasan. Importance sampling - Applications in communications and detection, *Springer-Verlag*, Berlin, 2002.

[215] J. Starck and A. Hilton. Model-based multiple view reconstruction of people, *IEEE International Conference on Computer Vision (ICCV)*, pp. 915–922, 2003.

[216] C. Stauffer andW. Grimson. Adaptive backgroundmixture models for real-time tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 246–252, 1999.

[217] T. A. Stephenson. Conditional GaussianMixtures, *Technical Report IDIAP-RR 03-11*, 2003.

[218] E. Sudderth. Graphical Models for Visual Object Recognition and Tracking, *PhD. Thesis*,Massachusetts Institute of Technology,May 2006.

[219] E. Sudderth, M. Mandel, W. Freeman and A. Willsky. Distributed Occlusion Reasoning for Tracking with Nonparametric Belief Propagation, *Neural Information Processing Systems (NIPS)*, 2004.

[220] E. Sudderth, A. Ihler, W. Freeman and A. Willsky. Nonparametric belief propagation, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 605–612, 2003.

[221] J. Sun, H. Shum and N. Zheng. Stereo matching using belief propagation, *European Conference on Computer Vision (ECCV)*, pp. 510–524, 2002.

[222] C. J. Taylor. Reconstruction of Articulated Objects from Point Correspondences in a Single Image, *Computer Vision and Image Understanding*,Vol. 80, No. 3, pp. 349–363, 2000.

[223] A. Torralba, K. P.Murphy andW. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 762–769, 2004.

[224] A. Torralba, K. P. Murphy, W. T. Freeman and M. A. Rubin. Context-based vision system for place and object recognition, *IEEE Intlernational Conference on Computer Vision (ICCV)*, Vol. 1, pp. 273–280, 2003.

[225] L. Torresani, P. Hackney and C. Bregler. Learning Motion Style Synthesis from Perceptual Observations, *Neural Information Processing Systems (NIPS)*, 2006.

[226] R. Urtasun, D. J. Fleet and P. Fua. 3D People Tracking with Gaussian ProcessDynamicalModels, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 238–245, 2006.

[227] R. Urtasun, D. J. Fleet, A. Hertzmann and P. Fua. Priors for People Tracking from Small Training Sets, *IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 403–410, 2005.

[228] R. Urtasun and P. Fua. 3D Human Body Tracking Using Deterministic Temporal Motion Models, *European Conference on Computer Vision (ECCV)*, Vol. 3, pp. 92–106, 2004.

[229] V. Vapnik. Statistical Learning Theory, *John Wiley and Sons, Inc.*, New York, 1998.

[230] V. Vapnik. The Nature of Statistical Learning Theory, *Springer-Verlag*,New York, 1995.

[231] S. Vedula, S. Baker and T. Kanade. Image-Based Spatio-Temporal Modeling and View Interpolation of Dynamic Events, *ACM Transactions on Graphics*, Vol. 24, No. 2, pp. 240–261, April 2005.

[232] S. Vedula, S. Baker, P. Rander, R. Collins and T. Kanade. Three-Dimensional Scene Flow, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 27, No. 3, pp. 475–480,March 2005.

[233] J. Vermaak, A. Doucet and P. Perez. Maintaining Multi-Modality through Mixture Tracking, *IEEE International Conference on Computer Vision (ICCV)*, Vol. 2, pp. 1110–1116, 2003.

[234] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, April 1967.

[235] M. Cheng, et al. Gait analysis for human identification through manifold learning and HMM. *Pattern Recognition*, Vol. 41, pp. 2541-2553, 2008.

[236] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 511–518, 2001.

[237] S. Wachter and H. H. Nagel. Tracking Persons in Monocular Image Sequences, *Computer Vision and Image Understanding*, 74(3):174–192, 1999.

[238] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Technical Report 649*, Department of Statistics, UC Berkeley, September 2003.

[239] M. Wainwright, T. Jaakkola and A. Willsky. Tree-based reparameterization for approximate estimation on loopy graphs, *Neural Information Processing Systems (NIPS)*, 2002.

[240] J. Wang, D. Fleet and A. Hertzmann. Multifactor Gaussian process models for style-content separation, *International Conference on Machine Learning (ICML)*, Vol. 227, pp. 975–982, 2007.

[241] P.Wang and J. M. Rehg. A Modular Approach to the Analysis and Evaluation of Particle Filters for Figure Tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 790–797, 2006.

[242] S.Waterhouse. Classification and Regression using Mixtures of Experts, *Ph.D. thesis*, Department of Engineering, Cambridge University, U.K, 1997.

[243] M. Weber, M. Welling and P. Perona. Unsupervised Learning of Models for Recognition, *European Conference on Computer Vision (ECCV)*, pp. 18–32, 2000.

[244] Y.Weiss andW. T. Freeman. Correctness of belief propagation in Gaussian graphicalmodels of arbitrary topology, *Neural Computation*, Vol. 13, pp. 2173–2200, 2001.

[245] G.Wywill and T. L. Kunii. A functional model for Constructive Solid Geometry, *The Visual Computer*, Vol. 1, Pt. 1, pp. 3–14, July 1985.

[246] W.Wiegerinck. Variational approximations between mean field theory and the junction tree algorithm, *Conference on Uncertainty in Artificial Intelligence*, Vol. 16, pp. 626–633, 2000.

[247] 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 (PAMI)*, Vol. 19, No. 7, pp. 780–785, July 1997.

[248] Y. Wu, G. Hua and T. Yu. Tracking Articulated Body by Dynamic Markov Network, *IEEE International Conference on Computer Vision (ICCV)*, pp. 1094–1101, 2003.

[249] B. Xie, D. Comaniciu, V. Ramesh, M. Simon and T. Boult. Component fusion for face detection in the presence of heteroscedastic noise, *Annual Conf. of the German Society for Pattern Recognition (DAGM'03)*, pp. 434–441, 2003.

[250] Y. Yacoob and L. S. Davis. Learned Models for Estimation of Rigid and Articulated Human Motion from Stationary or Moving Camera, *International Journal of Computer Vision (IJCV)*, Vol. 36, No. 1, pp. 5–30, 2000.

[251] H. Yalcin, M. J. Black and R. Fablet. The dense estimation of motion and appearance in layers, *IEEE Workshop on Image and Video Registration*, 2004.

[252] J. S. Yedidia,W. T. Freeman and Y.Weiss. Generalized belief propagation, *Neural Information Processing Systems (NIPS)*, pp. 689–695, 2000.

[253] J. S. Yedidia, W. T. Freeman and Y. Weiss. Understanding belief propagation and its generalizations, *Technica Report*, TR2001-22,MERL, 2001.

[254] J. S. Yedidia. An idiosyncratic journey beyond mean field theory, *Advanced Mean Field Methods*. MIT Press, 2001.

[255] J. S. Yedidia, W. T. Freeman and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms, *IEEE Transactions on Information Theory*, Vol. 51(7), pp. 2282–2312, July 2005.

[256] S. Yonemoto, D. Arita and R. Taniguchi. Real-time human motion analysis and IK-based human figure control, *Proceedings of the Workshop on Human Motion (HUMO)*, 2000.

[257] J. Zhang, J. Luo, R. Collins and Y. Liu. Body Localization in Still Images Using Hierarchical Models and Hybrid Search, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 1536–1543, 2006.

[258] N. Zhang and D. Poole. Exploiting causal independence in Bayesian network inference, *Journal of Artificial Intelligence Research*,Vol. 5, pp. 301–328, 1996.

[259] T. Zhao and R. Nevatia. TrackingMultiple Humans in Complex Situations, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 26, No. 9, pp. 1208–1221, September 2004.

[260] T. Zhao and R. Nevatia. Bayesian Human Segmentation in Crowded Situations. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 459–466, 2003.

[261] T. Zhao, T. Wang and H. Shum. Learning A Highly Structured Motion Model for 3D Human Tracking, *Asian Conference on Computer Vision (ACCV)*, 2002.

[262] L. Zhu, Y. Chen and A. Yuille. Unsupervised Learning of a Probabilistic Grammar for Object Detection and Parsing, *Neural Information Processing Systems (NIPS)*, 2006.

[263] L. Zhu and A. Yuille. A Hierarchical Compositional System for Rapid Object Detection, *Neural Information Processing Systems (NIPS)*, 2005.

[264] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 849-865, 1988.

[265] T.J. Cham and J. Rehg. A multiple hypothesis approach to figure tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 239-245, 199.

[266] G. Cheung, S. Baker and T. Kanade. Shape-From-Silhouette of aritculated objects and its use for human body kinematics estimation and motion capture, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol 1., pp. 77-84, 2003.

[267] L. Saul and S. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, Vol. 4, pp. 119-155, 2002.

[268] N. Lawerance. Gaussian process latent variable models for visualization of high dimensional data. *Advances in Neural Information Processing Systems*, Vol. 16, pp.329-336, 2003.

[269] M. Tipping and C. Bishop. Probabilistic principle component analysis. *Journal of the Royal Statistical Society, Series B,* Vol. 3, pp.611-622, 1999.

[270] I. Mikic and M. Trivedi. Human Body Model Acquisition and Tracking Using Voxel Data. *International Journal of Computer Vision*, Vol. 3, pp. 199-223, 2003.

[271] Y. Wang and K. Cheng. Two-Stage Bayesian Network Method for 3D Human Pose Estimation from Monocular Image Sequences. *EURASIP Journal on Advances in Signal Processing*, Vol. 2010, pp. 1-16, 2010.