DETERMINATION OF NORMAL OR ABNORMAL GAIT USING A TWO-DIMENSIONAL VIDEO CAMERA

Benjamin A. Smith

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

Master of Science
In
Mechanical Engineering

Dr. Michael Roan, Chair
Dr. Thurmond Lockhart
Dr. Martin Johnson

April 6, 2007
Blacksburg, Virginia

Keywords: Gait, Abnormality Detection, Pattern Recognition, Computer Vision

DETERMINATION OF NORMAL OR ABNORMAL GAIT USING A TWO-DIMENSIONAL VIDEO CAMERA

Benjamin A. Smith

Abstract

The extraction and analysis of human gait characteristics using image sequences and the subsequent classification of these characteristics are currently an intense area of research. Recently, the focus of this research area has turned to the realm of computer vision as an unobtrusive way of performing this analysis. With such systems becoming more common, a gait analysis system that will quickly and accurately determine if a subject is walking normally becomes more valuable. Such a system could be used as a preprocessing step in a more sophisticated gait analysis system or could be used for rehabilitation purposes. In this thesis a system is proposed which utilizes a novel fusion of spatial computer vision operations as well as motion in order to accurately and efficiently determine if a subject moving through a scene is walking normally or abnormally. Specifically this system will yield a classification of the type of motion being observed, whether it is a human walking normally or some other kind of motion taking place within the frame. Experimental results will show that the system provides accurate detection of normal walking and can distinguish abnormalities as subtle as limping or walking with a straight leg reliably.

Contents

List of Figures

List of Tables

CHAPTER 1 - INTRODUCTION

SECTION 1.1 – MOTIVATION AND OBJECTIVE

Analysis of human walking movements, or gait, has been an on going 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 in which considerable research has been done that exploits the analysis of this motion. These areas include clinical gait analysis [46], used for rehabilitation purposes, and biometric gait analysis for automatic person identification [14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 30]. These two areas dominate the field of gait research at the present. Although both 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 which are used specifically for motion capture. Biometric goals of human gait analysis are much different in that the goal is to unobtrusively observe and analyze an individual's gait in a variety of different areas and scenarios. Because of these limitations, gait analysis for use in biometric systems are largely based on visual data capture and analysis systems which process video of walking subjects in order to analyze gait. The limitations that are inherent in these techniques necessitate the use of sophisticated computer vision systems to generate variables which describe the gait motion.

Although considerable research has been done in the above areas, very limited research has been done in the area of gait analysis in a global sense. A system that observes and analyzes human gait for determination of abnormalities would provide a considerable amount of useful information for applications including surveillance, rehabilitation and various other unobtrusive gait analysis systems. In order for such an unobtrusive gait analysis to become feasible, a classification of the observed actions would serve as a crucial first step in any gait analysis system. This would ensure that the desired actions to be analyzed by the system are being observed by the system.

SECTION 1.2 – 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 [25, 27, 18, 43] in order to classify a limited number of grossly different human motions (such as various gymnastic movements [21]). In this work, a system is proposed which will determine if a subject moving in a video sequence is either walking normally or moving in a way that differs from normal human gait. The system presented in this thesis is unique in the way that the gait variables that are extracted from the video sequence, a variety of computer vision techniques are utilized by the system that equate to traditional gait descriptors (such as hip angle and height). The assumption that human gait motion behaves as a damped harmonic oscillator that reacts in sum fundamental way to certain stimuli is exploited in a pattern recognition scheme used to classify the observed gait as normal or abnormal. The data collected from the system for this normality analysis can then be passed onto a higher level system for a more in-depth analysis of the gait data.

In the experimental results of the system it is shown that the approach yields an accurate classification of normal and abnormal gait when presented with a number of different walking subjects exhibiting a variety of abnormal gait motions and as well as normal gait. 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 in order to classify a variety of different abnormalities.

SECTION 1.3 – ORGANIZATION

The remainder of this thesis is organized as follows. In Chapter 2, the various traditional computer vision techniques used in human gait analysis are introduced. Emphasis is placed on those elements that provide a significant resource for the analysis performed in the current system. Chapter 3 describes the theory and history behind the idea that the human walking motion acts as a damped harmonic oscillator which is repeatable and predictable. The final background chapter, Chapter 4, describes areas of the field of pattern recognition that are applied in this thesis. With the background of the individual fields of study used in the system described, Chapter 5 describes the newly developed system in detail. This chapter is divided into various sections each of which describes a technique or algorithm used in determining the final system output. Chapter 6 presents the results of the final system as well as discussion of the results obtained from the system. Finally, in Chapter 7, conclusions and future research are presented.

CHAPTER 2 – COMPUTER VISION REVIEW

SECTION 2.1 - INTRODUCTION

This chapter describes the various methods used to extract necessary variables and descriptors from a video sequence in order to provide them to a classification system. The following sections describe optical flow, background determination and extraction as well as the basic computer vision operations which are utilized in the system.

SECTION 2.2 – OPTICAL FLOW

Moving objects exist almost everywhere within our world. The nature of the motion that is occurring in a scene provides a vast source of information about what is being observed. A measure of the motion found within a scene in a two-dimensional image 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. This measure is found using two or more 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 estimation of motion can then be used to accomplish a wide variety of tasks related to video compression, scene analysis, navigation, and three-dimensional scene reconstruction [4]. This technique 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]. In a paper which yielded the most influential research on optical flow computation, the paper, "Determining Optical Flow", provided the first real definition of what optical flow was and gave direction on how to calculate this measure. 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 the accurate estimation of optical flow 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 following basic equation of optical flow found in [2] as

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + 1) \qquad (1)$$

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 various 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]. Many techniques have been developed in order to calculate the optical flow found within a scene. The two most dominant methods are those that use differential techniques or those that use area matching techniques. Differential techniques use the spatiotemporal derivatives of the image intensity in order to calculate the motion taking place in the vertical and horizontal directions of the image sequence. Area based techniques utilize the matching of small portions of an image from one image to another within the sequence in order to find the motion taking place in the video from frame to frame. 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 are then moved around in the previous frame in order to find where in the previous frame the image block was located [4]. This location information is then used to form the estimation of the motion from the previous frame to the current frame for each block. This technique also allows the motion estimation of the sequence to be partially robust to certain assumptions made in the basic optical flow calculation. Block matching motion estimation techniques allow for the violation of the spatial and temporal consistency assumptions as well as the constant intensity assumption. Since block matching techniques do not retain motion information beyond the current image frame, the temporal consistency assumption is not imposed, this means that motions can be erratically changing in time. The spatial consistency requirement can be tightened or relaxed by defining the search area of the block searching algorithm. This is due to the fact that the algorithm can only detect motions equal to or smaller than the defined search parameter. 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 in turn makes block matching algorithms more accurate for real images taken using normal cameras (cameras which introduce a large amount of noise in the pixel intensity values) since the search is now more robust to overall image changes, such as those due to lighting changes noise added by the image acquisition device. The properties mentioned above make this technique especially appropriate for gait analysis because the motion is not uniform in nature throughout the scene or throughout the object, in this case the subject being analyzed, and the images are taken using relatively low quality cameras (ie noisy, low resolution). Block matching algorithms provide an ideal search technique that will account for the types of motion and environments encountered in gait analysis.

Since the early development of block matching algorithms, many different techniques have been proposed in order to increase the accuracy of the motion estimation as well as the efficiency of the computation. Since the middle of the 1980's block matching algorithms have been developed in order to estimate 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. These cost functions vary in computational complexity and accuracy. The two most popular cost functions [3] are the mean absolute difference given by

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

And the mean squared error given as

$$MSE = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(C_{ij} - R_{ij}\right)^2 \qquad (3)$$

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. These measures are especially attractive over statistical matching measures for the application of gait analysis because rotational as well as translational motion can take place within the scene. 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 begins to rotate around the hip. This is illustrated in Figure 2.1. This will cause most severe errors at edge boundaries of objects that are rotating or objects which are not rigid (deformable). These errors will result in a block not matching correctly in a statistical matching scheme.

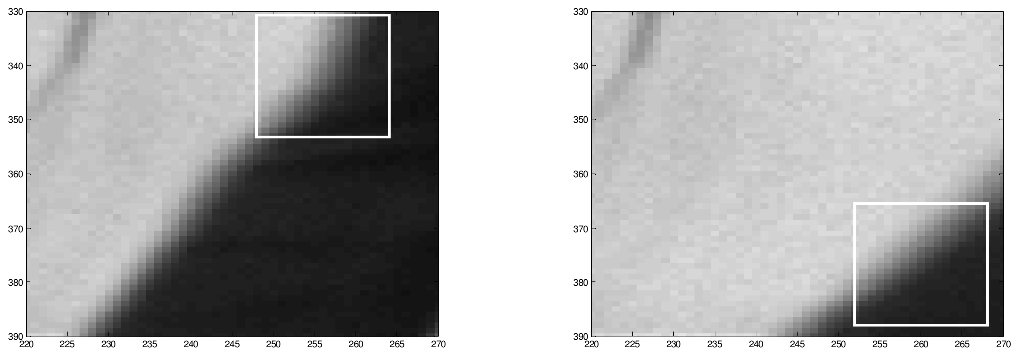


Figure 2.1: Block Errors, the Block area is outlined in the white box, as can be seen the region being tracked stays the same however there is more background being introduced at the edge and the edge shape is changing enough so that the block will not be tracked properly by statistical matching criteria

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 rotational or non-rigid motion, while a statistical matching scheme provides superior performance for scenes containing rigid objects experiencing translational motion. 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 to be used.

Choosing the block size is typically a process that is tailored to the specific application of the motion estimation scheme. Larger blocks yield more accurate motion estimations for large objects moving within the scene, but can lead to poor motion resolution within the frame. Smaller blocks yield more dense motion estimation fields, but can be inaccurate if the size of the block is so small that blocks can no longer be reliably matched using the chosen cost function. For most applications, a block size of 16x16 is used [4] in order to provide a large enough sample of pixels to reliably match blocks while still retaining an accurate motion field. The block size must be tailored to the specific application, hardware and algorithm to be used for estimate motion. In addition to block size, a search window size must be defined. Again, this quantity is very much application and algorithm specific, a larger search window increases the computational cost of the algorithm while allowing for larger motions, but can also account for erroneous motion calculations (violation of spatial consistence assumption). A smaller search window size provides less area to search and as a result a less computationally expensive algorithm. However, motions larger than the search window could result in large errors in the motion estimates. The search window parameter must be selected so that it provides a large enough search window to calculate all desired motions, but not so large as to provide unnecessary computational burden. The next aspect of block matching motion estimation algorithms is how blocks are searched for from one frame to another.

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, this results in the most accurate motion estimation that can be obtained with the chosen algorithm components. Beyond the exhaustive search method, many different algorithms have been developed that search specific points within the search area instead of all the individual points contained in the search area. These algorithms are more efficient due to the reduced number of search points, but can introduce a larger amount of error in the final motion estimation. The most popular of these search algorithms that have been developed during the last two decades are the 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 motion estimation 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 certain location it is highly likely that a block located near this block moves in the same way. In order to accomplish this task, 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 2.2.

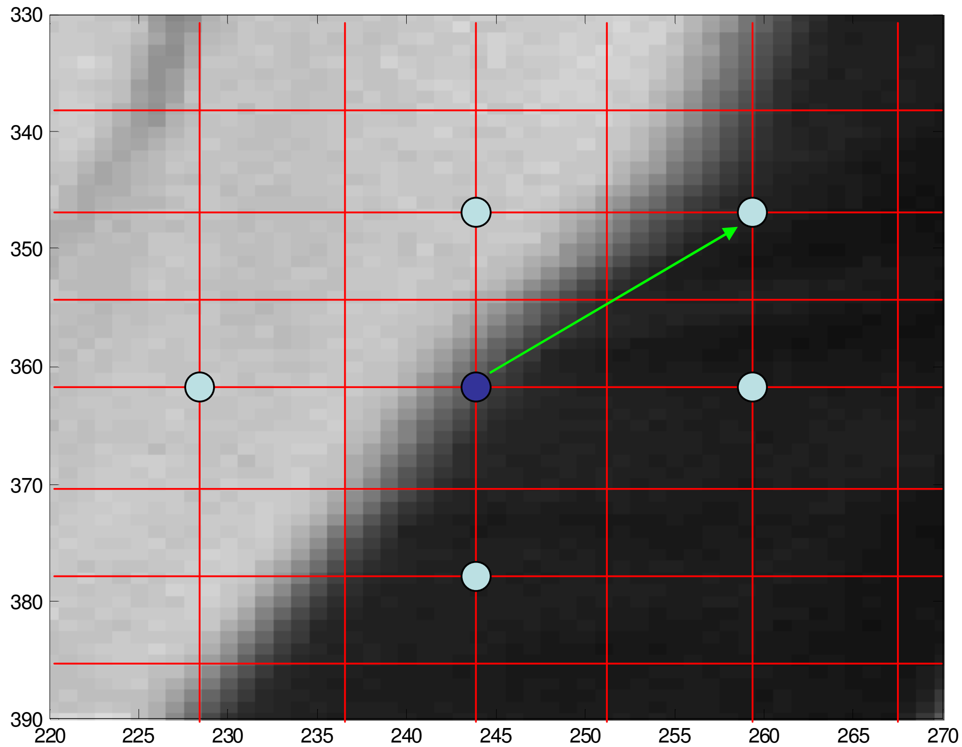


Figure 2.2: The search pattern for the Adaptive Rood Pattern Search for a block size of 8x8 pixels (block boundaries shown as red lines), where the green arrow is the previous block's motion vector, the step size is the maximum value of the motion vector in any one direction (16 pixels in this case), 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 the equation

$$S = Max(|X|, |Y|) \tag{4}$$

where X is the x-vector of the predicted motion vector and Y is the y-vector of the predicted motion vector. Upon searching each of the five points defined above, the point which has the corresponding lowest value of the chosen cost function is then set as the origin for the subsequent search operations [4]. In addition to the above five points, the center point of the search pattern, the current block location, is also searched. The next step in the algorithm is to perform a small diamond search around the point chosen in the above operation. This is not done however if the point chosen above is found to be the current location of the block. This is because it is determined that the block does not move from the reference to the current frame. The small diamond search pattern is shown in Figure 2.3.

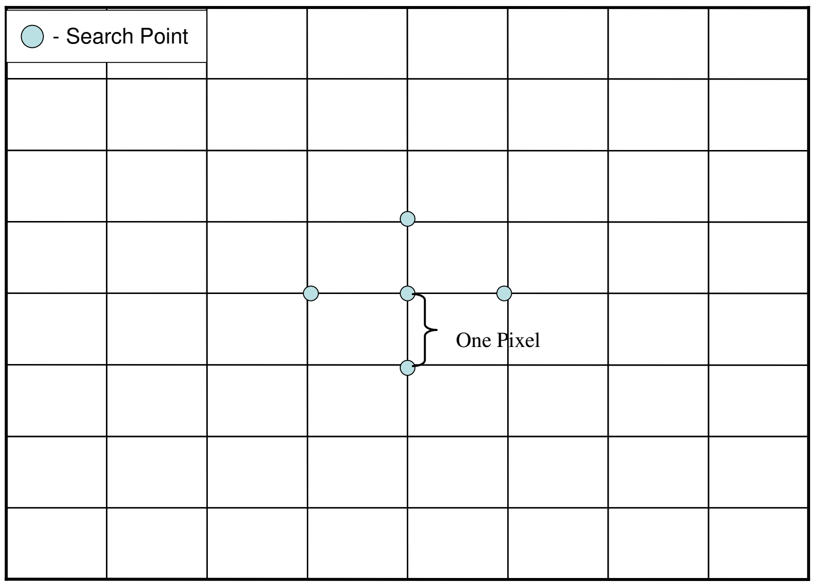


Figure 2.3: The Small Diamond Search Pattern

This search pattern then finds the point that again 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.

SECTION 2.3 – SPATIAL IMAGE ANALYSIS

Motion information provides a large amount of data about a scene; however, motion information is not the only information available in a scene. In addition to spatiotemporal data, spatial data, that is to say shape information, also provides a large amount of data that can be exploited in order to classify human gait. Spatial information comprises a vast majority of the computer vision field and includes, but is not limited to, image filtering, edge detection, background determination, and shape analysis [3]. Spatial image data is used in many different ways and in many different applications in order to analyze gait [6]. For many years spatial data has been used in order to develop various systems for human gait identification as a biometric [6]. The main types of spatial analysis used in these systems are background extraction and shape analysis from the resulting binary silhouettes. In order to find a suitable silhouette of the subject, background extraction must first be done to separate the complex background from the subject located in the foreground.

Background extraction consists of removing the background, or any objects not of interest in the scene, from an image while retaining the foreground objects of interest. In human gait analysis, the foreground objects of interest are the human subjects walking through the scene 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 2.4.

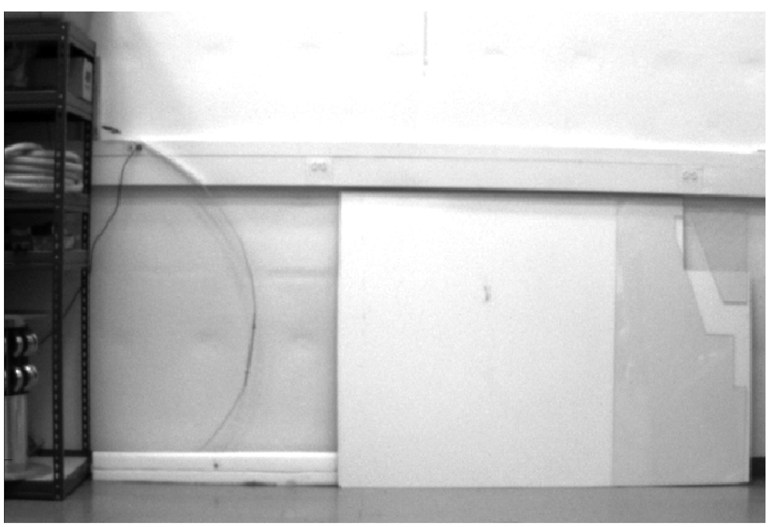


Figure 2.4: A Background Containing Different Objects

Background determination is typically accomplished by comparing an image to a model of a scene's background. This can be done 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, a threshold is then applied to the results. This leads to a background subtraction that cannot adapt to changes in the background (ie. background movement) or noise due to the image capture device and results in a high number of false positives. An alternative approach to this simple method is the use of statistical means in order to create a background model and determination scheme. An example of such a method would be to use an average of a number of background frames in order to create a model of the background that could account for image noise due to the image capture device. This background model can then be compared to the current image in much the same way as described above to determine the presence of foreground objects. While the described model provides limited 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. Another class of statistical background extraction methods is those which utilize non-parametric statistics to create the background models [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 has been done using a static background. However, a background model that can account for various changes, such as lighting and small background movements, is very important when utilizing a system in uncontrolled environments outside of a laboratory. The robustness in a variety of applications and situations is why statistical background determination schemes are so desirable. The methods utilize statistical models that do not require a uni-modal intensity distribution. This means that a non-parametric background model can account for background changes beyond the noise introduced by the image capture device. A system such as this provides an amazing amount of robustness when used in a scene that has a variety of intensity changes, such as those due to a tree blowing in the wind. This type of motion in the background of a scene causes large changes in pixel intensity over a very short amount of time. In order to account for this type of multi-modal distribution in the pixel intensity of the background, a model that can adapt to any distribution of the pixels is required. In the presence of changes described above a multi-modal probability distribution for each pixel in the background allows for a more accurate calculation of the probability that a given pixel is part of the background. This probability can then be compared to a predefined threshold value. If this probability is found to be below the threshold then the pixel is determined to be part of the foreground, if above the threshold, the pixel is established as part of the background.

CHAPTER 3 – NATURE OF GAIT

SECTION 3.1 - INTRODUCTION

In order to establish that abnormality in human gait can be globally detected, is the idea a model of normal human gait is needed. This implies that normal human gait is a repeatable process and has predictable characteristics across individuals that differ in height, weight and gender. This can be witnessed by anyone who observes an individual walking, predictable motions are observed when a human is walking normally and disruptions in this normal gait can be recognized and classified by an observer. It is this fundamental principle that is exploited in our system and explanation of human gait as a damped harmonic oscillator is given in the following section.

SECTION 3.2 – GAIT AS OSCILLATORY MOTION

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 of motion. Bipedal motion requires a very specific rhythmic pattern in order to avoid a number of instabilities and unnecessary energy expenditures. Because of this particular rhythmic motion, gross similarities are found in human gait even between individuals that are dissimilar in proportion, ethnicity, or any other attribute. This rhythmic motion can be broken down into a number of movements that dominate the motion. 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 in order 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 in order to accomplish forward motion of the entire body. There remain two specific factors that 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.

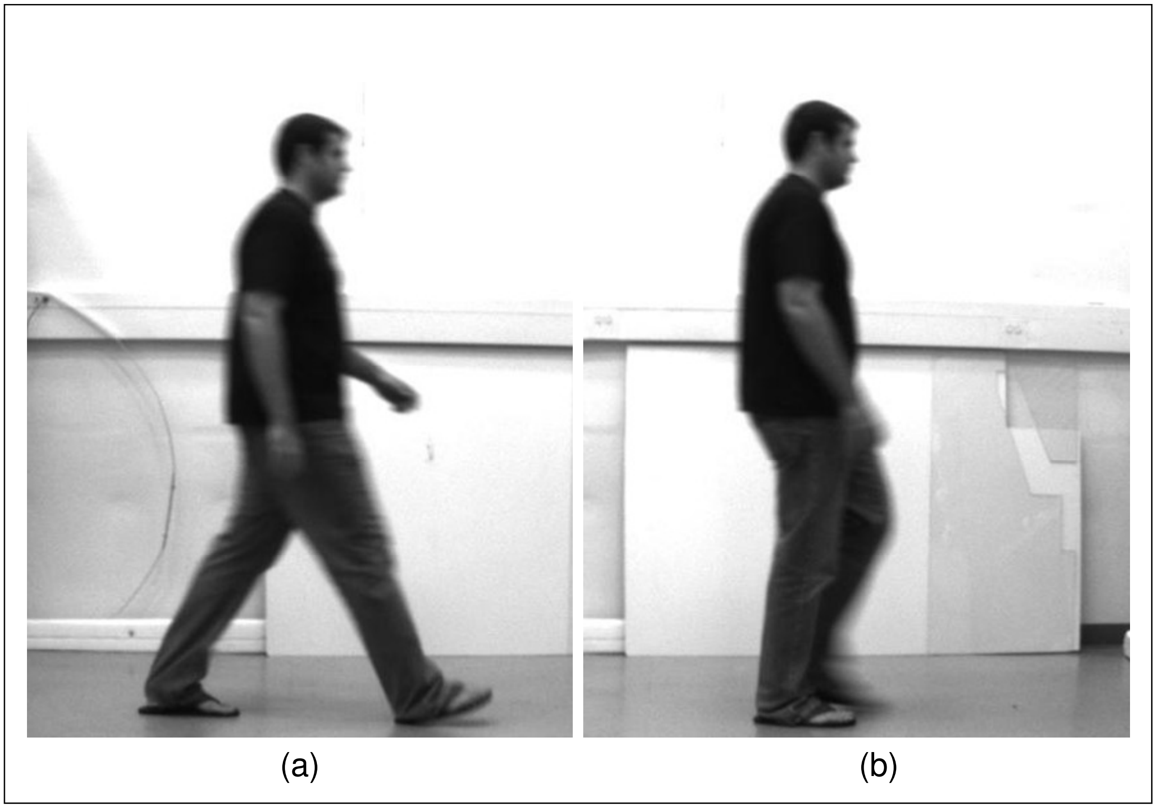


Figure 3.1: Comparison of double stance phase and single stance phase, notice how the configuration of the legs results in a change in vertical height from one phase as compared to the other

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 results 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.

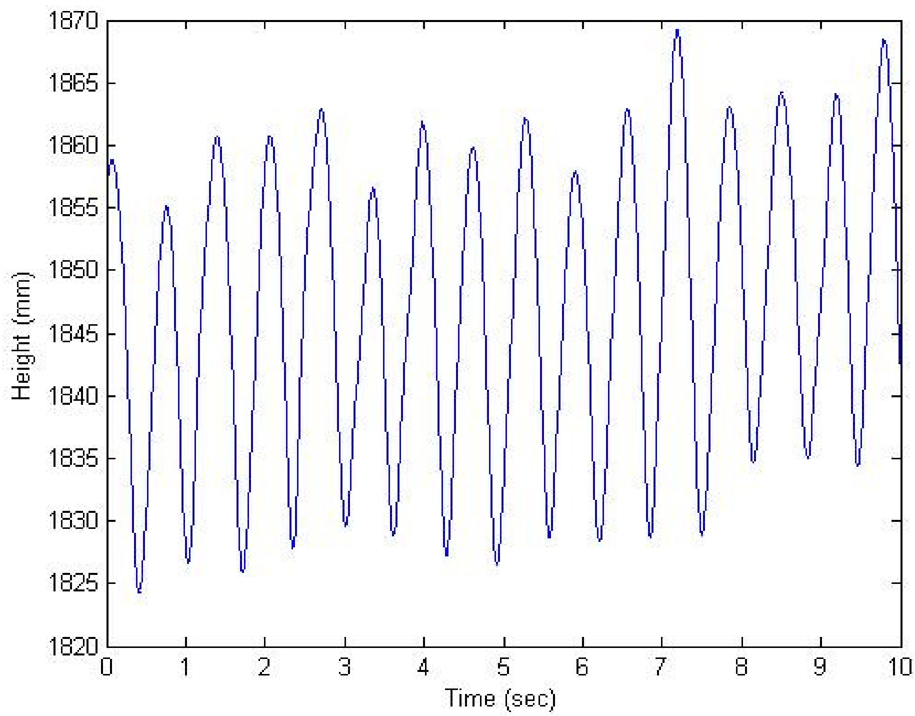


Figure 3.2: Vertical motion of the top of the head measured during normal walking; notice that the motion is dominated by a roughly sinusoidal component

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.

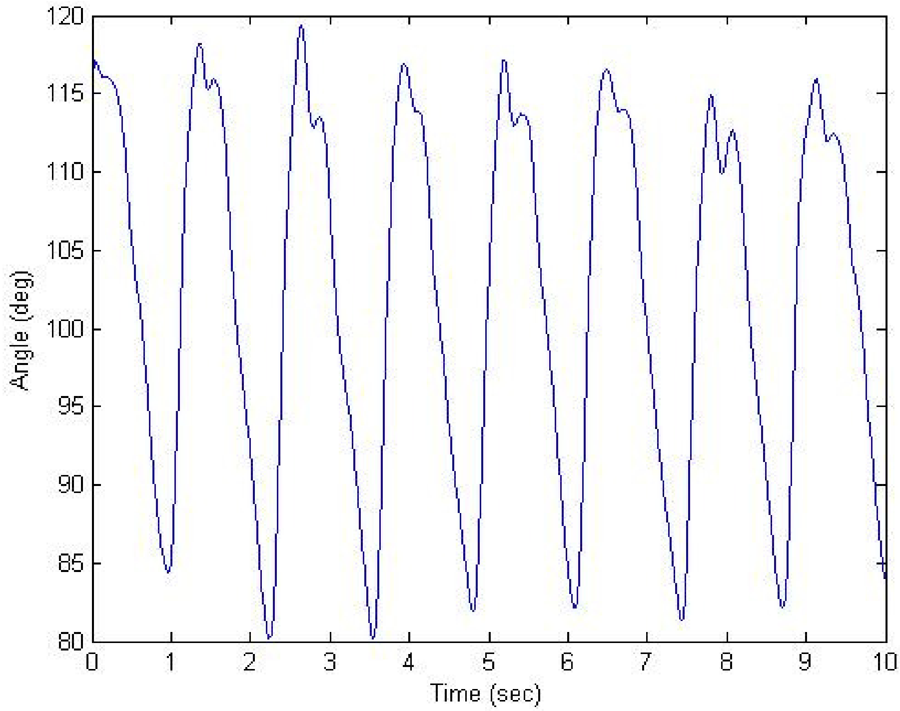


Figure 3.3: Hip angle of a subject calculated using motion capture data, the angle is degrees down for the plane parallel to the ground at the height of the hip, notice how the motion is dominated by a sinusoidal component with subtle, consistent preturbances

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.

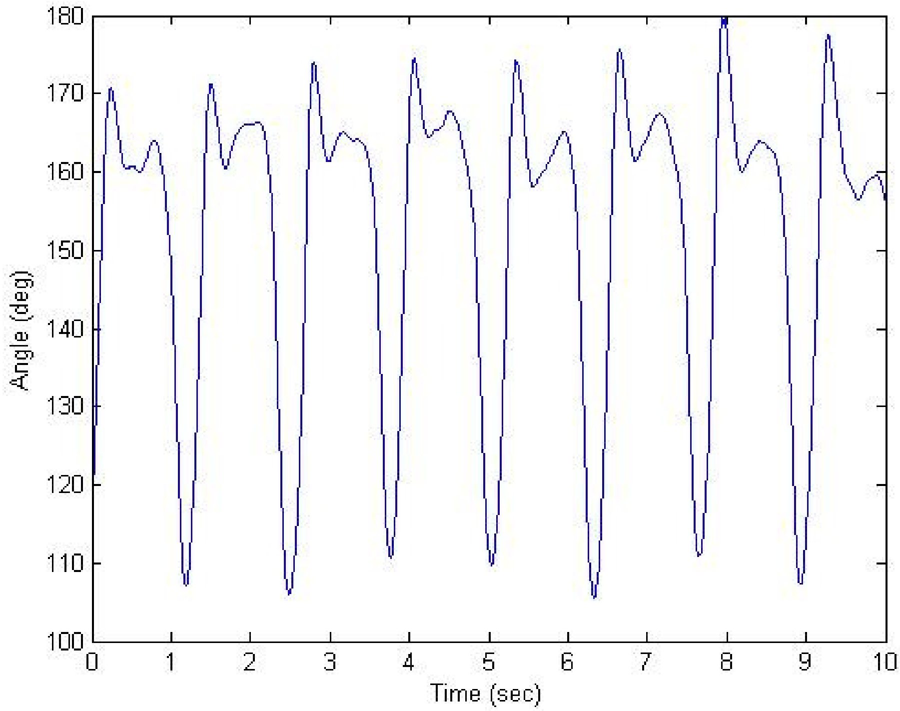


Figure 3.4: Knee angle, 180 degrees is a straight leg, notice the consistent oscillatory motion present during normal walking

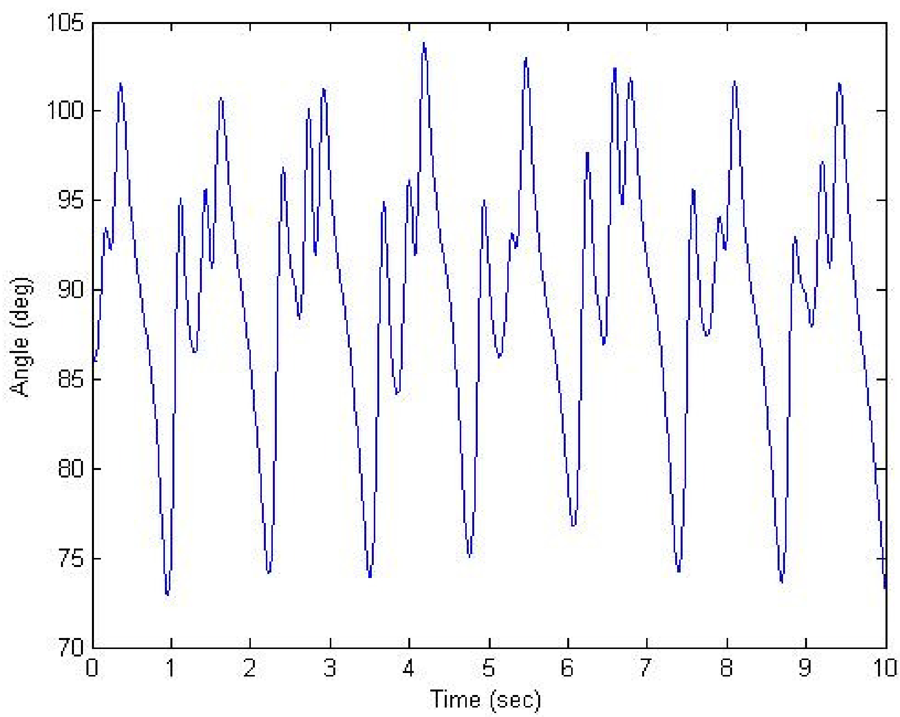


Figure 3.5: Ankle angle of a subject during normal walking

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.

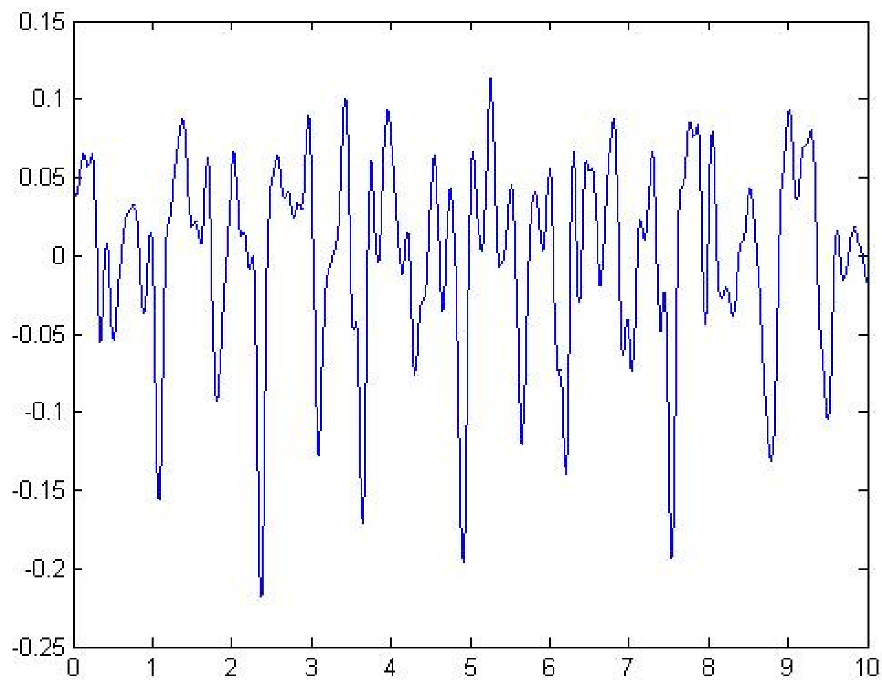
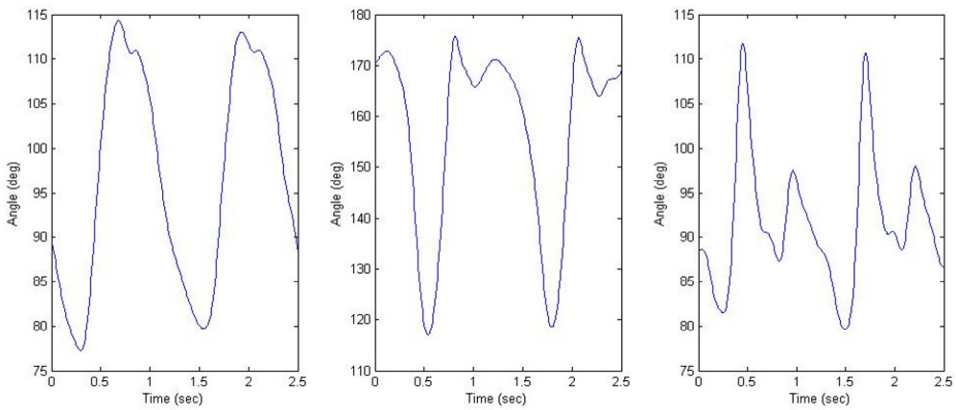


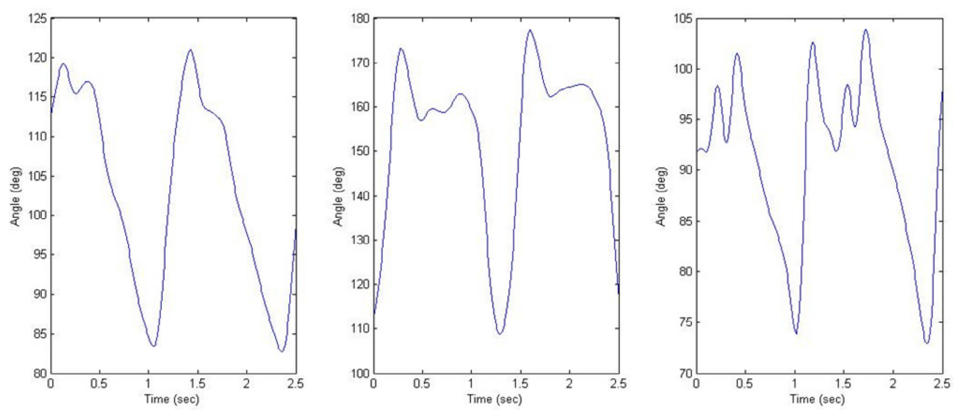
Figure 3.6: Upper body acceleration of a subject walking normally on a treadmill. Calculated using the second derivative of the shoulder position measured in the direction of movement. Notice small acceleration changes present in each cycle

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.



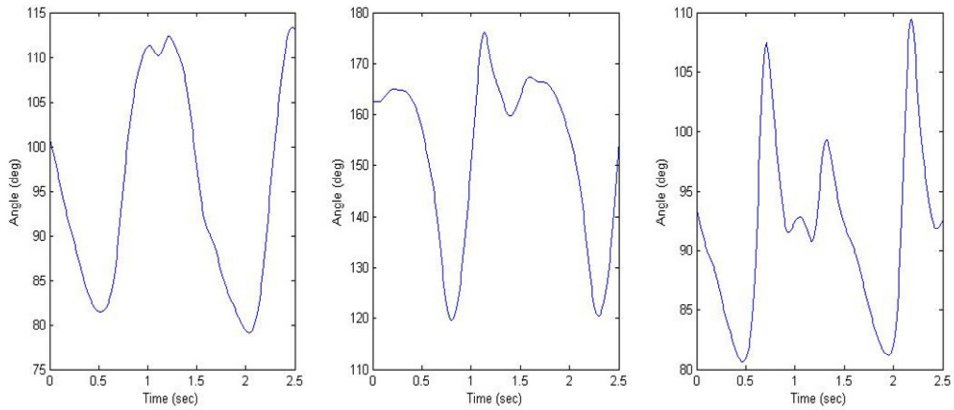
(a) (b) (c)

Figure 3.7: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle



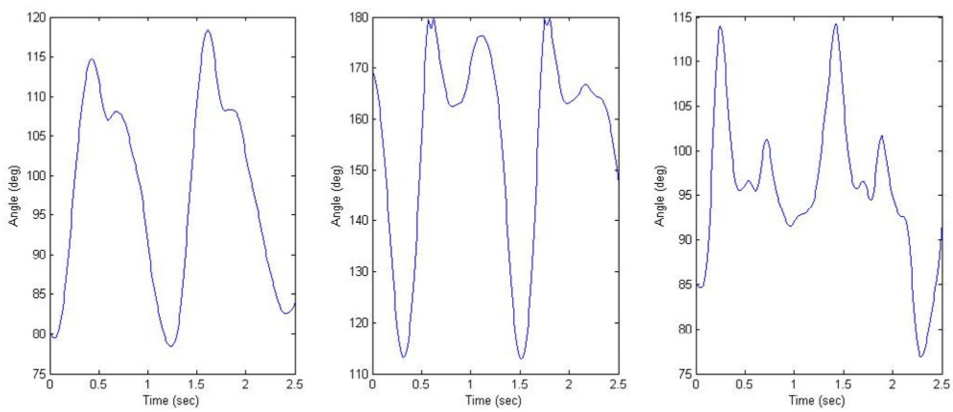
(a) (b) (c)

Figure 3.8: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle



(a) (b) (c)

Figure 3.9: Subject walking normally with Hip (a), Knee (b), and Ankle (c) angle



(a) (b) (c)

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.

CHAPTER 4 – PATTERN RECOGNITION REVIEW

SECTION 4.1 – INTRODUCTION

Patter recognition is a field which focuses on the description of objects by classification of those objects into various desired categories. The objects which are attempted to be classified are most often images or signal waveforms, but can be generalized to almost any type of measurement or variable [8]. This is of utmost importance in development of a system used to classify normal and abnormal gait patterns, as such; the field of pattern recognition will provide the necessary tools which can be used to classify the variables provided by the video processing.

The figure shown below taken from [8] demonstrates the various stages that must be accomplished in an effective pattern recognition system.

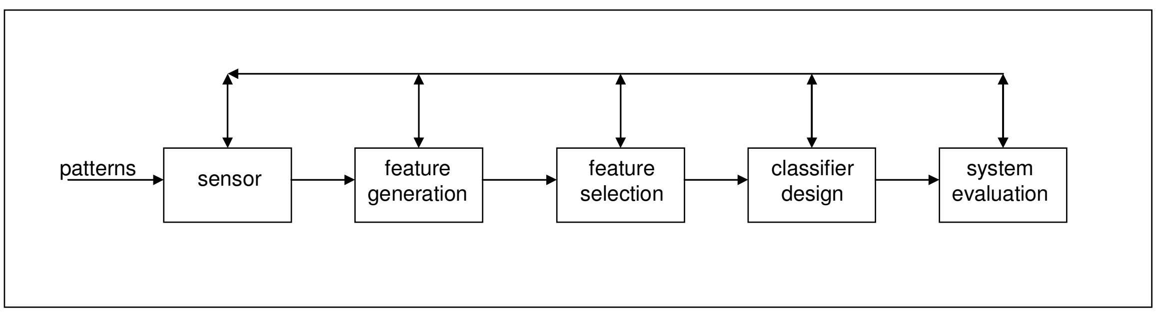


Figure 4.1: Block diagram of the design stages of a pattern recognition system

The figure above shows each of the various processes involved in the creation of a pattern recognition system, the first step is the sensor actually measuring the desired pattern to be recognized. After collection, features may be generated which aptly describe the signal and give the final classifier good discriminatory power, once features have been created using the incoming signal, a number of features should now be chosen in order to maximize the efficiency of the classification system. In order to accomplish this, features with a large amount of discriminatory information should be used and features that provide little classification power to the final system should be discarded, this way the final system is optimally efficient in performing the classification task. With the features chosen, the actual classification algorithm must be chosen and implemented. There are a large number of classifiers that are available ranging from linear to non-linear systems as well as supervised systems, classifiers requiring information regarding how the features are to be classified, and unsupervised systems, those which require no known feature information in order to perform classification. With such a wide variety of factors to chosen from the system design must be done recursively in order to maximize the effectiveness and efficiency of the overall system.

SECTION 4.2 – FEATURE GENERATION

The first task which must be accomplished when designing a pattern recognition system is the formation of variables which aptly describe the signals to be classified. These variables are known as features and must be selected so that they provide as much information about the original signal as possible in addition to providing a large amount of class discrimination so that they may be used effectively in the final system. A wide variety of methods may be utilized in order to generate features which can be used to classify a signal. The goal of this generation process is to break down a signal vector into a scalar value or values which will adequately describe the signal. This will allow signals which are two different lengths to be analyzed by the same system without depending on signal length normalization; this directly applies to a gait classification system since each person will walk at slightly different frequencies causing the length of one cycle to vary in length. A wide variety of different methods can be used to accomplish this feature generation task; some of the simplest measures are simple first and second order statistical measures. Use of mean and variance of a signal provides a wide variety of information of a signal; however, this measure will not provide a good metric of the actual waveform shape which is of utmost importance in the gait analysis application. While there are almost infinitely more measures of a signal that are more complicated and provide more in-depth information regarding the actual shape of the waveform, only a few allow for a simple calculation and thus an efficient classification. Some of these are classifications of the waveform via the discrete wavelet transform, which decomposes a signal into a number of basis waveforms, or a Fourier analysis, which yields frequency information of the signal for time-invariant signals. Other methods that can be distinctly advantageous in a gait classification system is an actual measure of similarity between a signal generated for normal gait and the incoming signal that is to be analyzed by the pattern recognition system. This proves to be a desirable method when compared to traditional waveform features because of the small time durations of the signal that are available for analysis. The gait that is to be analyzed by the system only provides a very limited amount of information, in most cases only one cycle, which makes a Fourier or wavelet analysis unfeasible. An example of a signal which will be input into this gait classification system is given below.

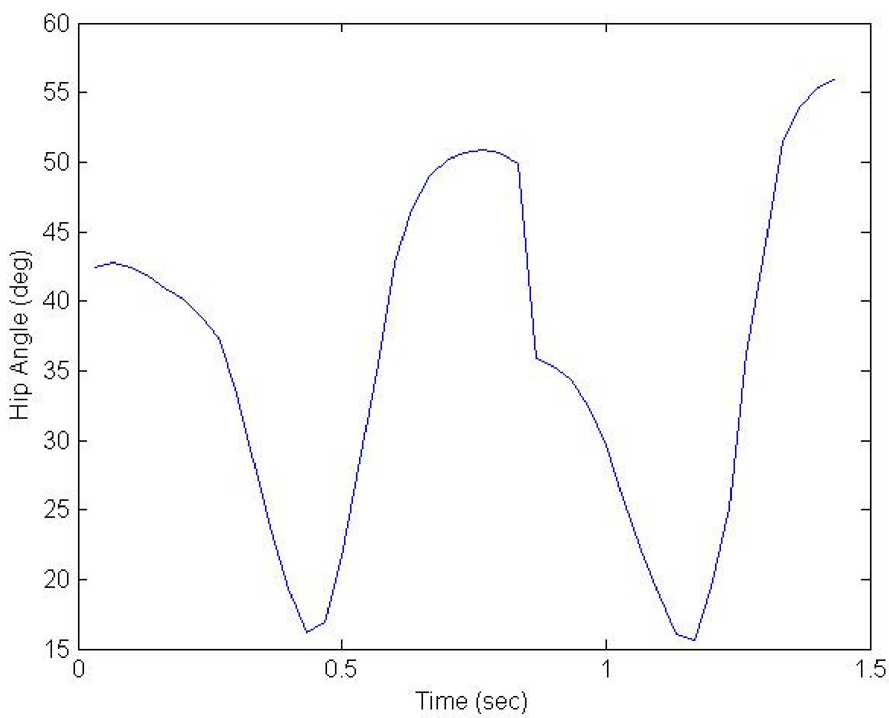


Figure 4.2: Sample of an input waveform for the system

As can be seen in the above figure a Fourier analysis will only result in a meaningless set of values since there are only roughly 30 points available for analysis of each signal.

With this limitation known, a method of comparing two signals of different lengths must be found which will provide a good feature for the classification system. One of the first methods which can be utilized efficiently is a measure of the error between an incoming signal and the reference, or normal gait, signal. The most common way of quantifying the error between two signals involves comparing the reference signal to the incoming signal and then quantifying the difference between them, usually by taking the average of the squared error between each point, the mean square error, or by simply taking the average difference between each point in the signal. Alternative ways to quantify the error between two signals, such as those that are generated by gait analysis, are to find the difference in the overall power of the signal, where power of the signal is defined as.

$$P_f = \lim \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left(\left| f(t) \right| \right)^2 dt \qquad (5)$$

Other methods of quantifying error can also be found using a combination of the above methods. In addition to error calculation another measure of the difference between two signals is to find the correlation coefficient between the two signals. The correlation coefficient is a normalized measure of how strong the two signals relate to each other This can be used to quantify how well an input signal matches the reference, or normal gait signal. It is this measure which can then be fed into the classifier.

SECTION 3.3 – NEURAL NETWORKS

A neural network is a data processing algorithm which 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 which is used in order to determine the output of the individual node; this output is then processed in conjunction with all the other nodes of the network in order 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.

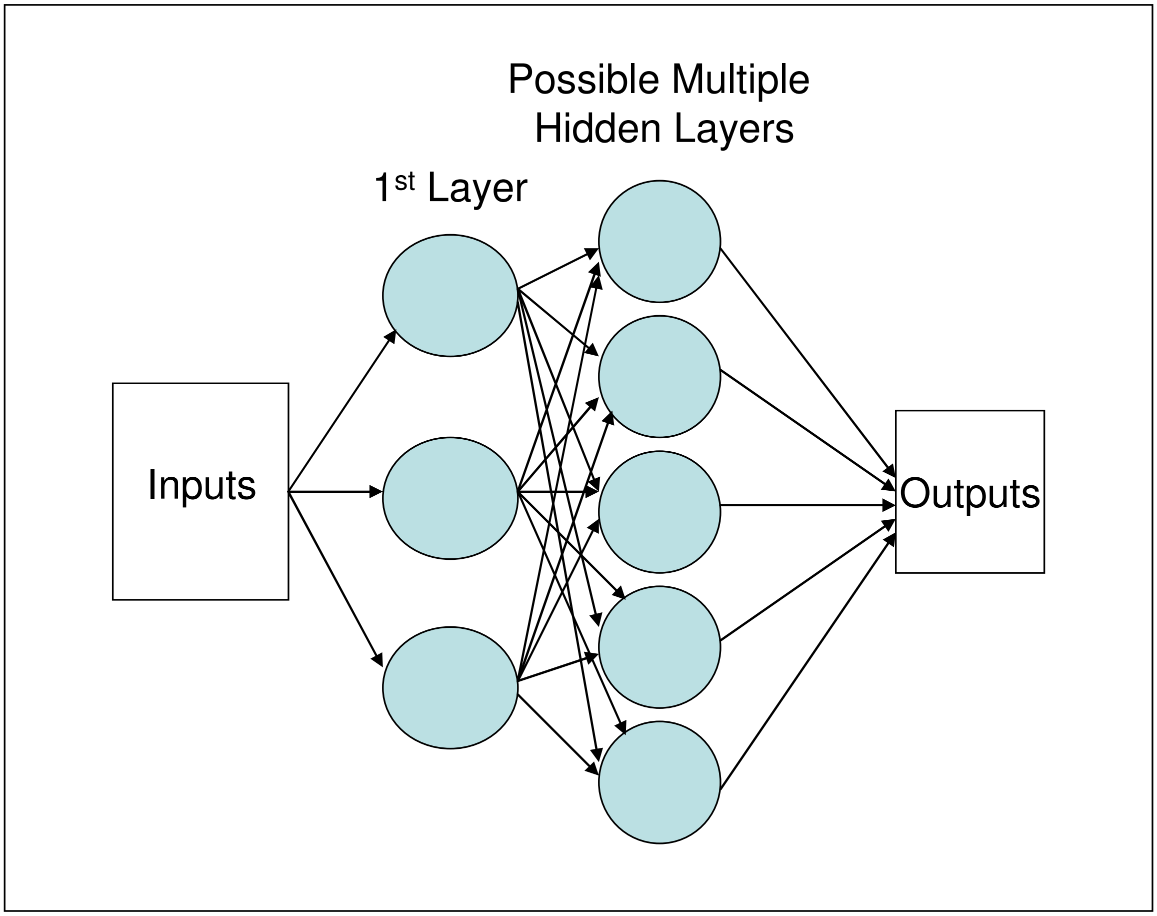


Figure 4.3: Conceptual Diagram of a Neural Network, a number of inputs are fed into the network, which then pass the data through a number of layers of differing numbers of nodes to form the final output

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, which 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]. One property, the property which makes neural networks such an attractive tool, that all neural networks have in common is the ability for the algorithm to learn. What this means in a practical sense is that the neural network has the ability to adjust its parameters in order to arrive at the optimal solution to the problem. There are a number of training algorithms available which take advantage of this learning ability to arrive at the optimal solution. Each of these training algorithms depend heavily on the type of features which are to be classified as well as the type of neural network being utilized for classification, the most common of these methods are evolutionary, simulated annealing, expectation-maximization and non-parametric methods.

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 provides versatility to be expanded upon further as well as an output and training method that are suited to the system being created, which is a determination of the observed gait as normal or abnormal. The type of network implemented consists of one layer of nodes, with each node corresponding to one feature input type, these nodes make up the input and output layer of the network. In this type of network, the nodes each perform their specified function and are summed together so that if the output value of the combined nodes is above a threshold value then the output takes on activation value, if the value is calculated to be below the threshold the output takes on the deactivation value. The typical single-layer 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 most simple of the training methods. 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 (target value), in this specific algorithm the error between the actual output value and the target value, which will either be 1 or -1 depending on if the output is a false negative or positive (0 when it should be 1 results in an error of 1), is multiplied by the current input and this value is added to the current weight. This calculation is shown below.

$$dw = (t - a)p^T \tag{6}$$

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, which can be a scalar or vector. This is done recursively until the calculated outputs of the system match the target outputs of the system or until a specified number of iterations through all the training inputs, also known as epochs, are completed.

CHAPTER 5 – SYSTEM DESCRIPTION

SECTION 5.1 – INTRODUCTION

As has been previously established, a system which can determine whether an individual is walking normally or abnormally presents itself as a valuable first step in many high level gait recognition and classification systems. Such a system can utilize information gathered for use in a higher level classification system to quickly and accurately determine whether the subject is walking in a way which can be reliably analyzed by the proceeding system. Further application could develop a system which would determine the proper analysis of the observed gait regardless of the type of gait observed. Such a preliminary analysis system, a proposed in this thesis, 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.

SECTION 5.2 – ACQUISITION

The acquisition phase consists of taking images of the scene and processing them so that they may be analyzed in the following steps. 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 in order to maintain versatility in the observable area of the camera. A manually adjusted lens is chosen instead of an automatic focus lens in order to preserve frame to frame consistency which is of paramount importance in any video analysis system as automatic zooming and focusing can change aspect ratios and cause large errors in measurement. To connect the camera to the video acquisition system a IEEE-1394, Firewire, interface is utilized. This interface is ideal because it allows the largest data transfer rates of traditional external interfaces and also provides a high degree of data transfer with minimal addition of noise. 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 due to Matlab's built-in functions and various toolboxes which greatly simplify the video processing methods by eliminating the need to create code to accomplish basic imaging tasks, such as image subtraction and conversion, and eliminates the need to write custom drivers in order to use the camera chosen for use. With Matlab used to collect the incoming video data from the camera the necessary processing steps can then be used to analyze the video data.

SECTION 5.3 – FOREGROUND DETERMINATION

Foreground determination is the first step in the processing of video data for this gait classification system. In order to determine which parts of the video are foreground and which are background, a preliminary background determination step must be completed. This is done using a non-parametric background extraction scheme similar to that discussed above in section 3 of chapter 2. For this implementation of the background extraction scheme, the background will be modeled based on a selection of background frames which do not contain subjects moving through the scene. This is done because with this gait system's testing and implementation, a limited number of frames are taken, roughly 100 per subject, which is an insufficient number of frames for an online background extraction. Basically, there are not enough frames available which can be used to build an accurate background model since the subject is moving through the entirety of the video frame over the 50 or so frames in which the subject is in the scene therefore the background model will account for changes due to the subject walking as changes in the background causing a high number of false positives and false negatives. Due to this limitation, the background extraction method, as it is implemented here, accounts only for the noise present due to the video acquisition system which is acceptable due to the short duration of acquisition time coupled with the use of a purely static background. Slight modification of the algorithm can be used in order to make the system applicable 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. This can be done manually by confirming the frame to be used does not contain a subject, or a number of the first frames can be taken in the beginning of the video which are known to not contain a subject. The latter method is utilized in this algorithm; the first 30 frames of every video sequence are used. With the frames to be used for the background model established, the first step that must be done is calculating the standard deviation of the individual pixel values found in the frames chosen to build the model. The standard deviation for the overall image if found by using the following equations.

$$\sigma_{i,j}{}^2 = \left(\frac{1}{N-1} \sum_{f=1}^{N} x_f(i,j) - \bar{x}(i,j) \right)^2 \tag{7}$$

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.

$$\sigma = \sqrt{\frac{1}{c*r} \sum_{i=1}^{c} \sum_{j=1}^{r} \sigma_{i,j}^2} \tag{8}$$

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

This standard deviation is then utilized in the statistical tests that are performed to determine if the individual pixel values of the incoming frame are part of the background. With the pixel wise standard deviation for the video known, a test for each pixel in the incoming image can be devised to accurately determine the portions of an image that are statistically different from the background. Again, the background model used is created using the first 30 frames of the video data and therefore 30 frames are used for the following background extraction operations. 30 frames are chosen in order to give suitably accurate results while still maintaining computational efficiency. The test which is to be used in this implementation is taken from [6] and is shown below.

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x_t - x_i)^2}{\sigma^2}} \tag{9}$$

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

With the probability of the incoming pixel being part of the background known, the pixel can be determined to be foreground in the probability value is found be below some threshold value, *th*. This threshold value can be adjusted depending on the video data being processed, for the data processed for this system a threshold value of 5 is chosen. This threshold value is very high in relation to many statistical tests (a threshold value of 0.05 is used in order to obtain a 95% confidence in the results), however, the data is such that best results, in terms of the least number of false positives, are obtained when using a high threshold value. The effect of different thresholds on the number of false positives is shown in the figure below.

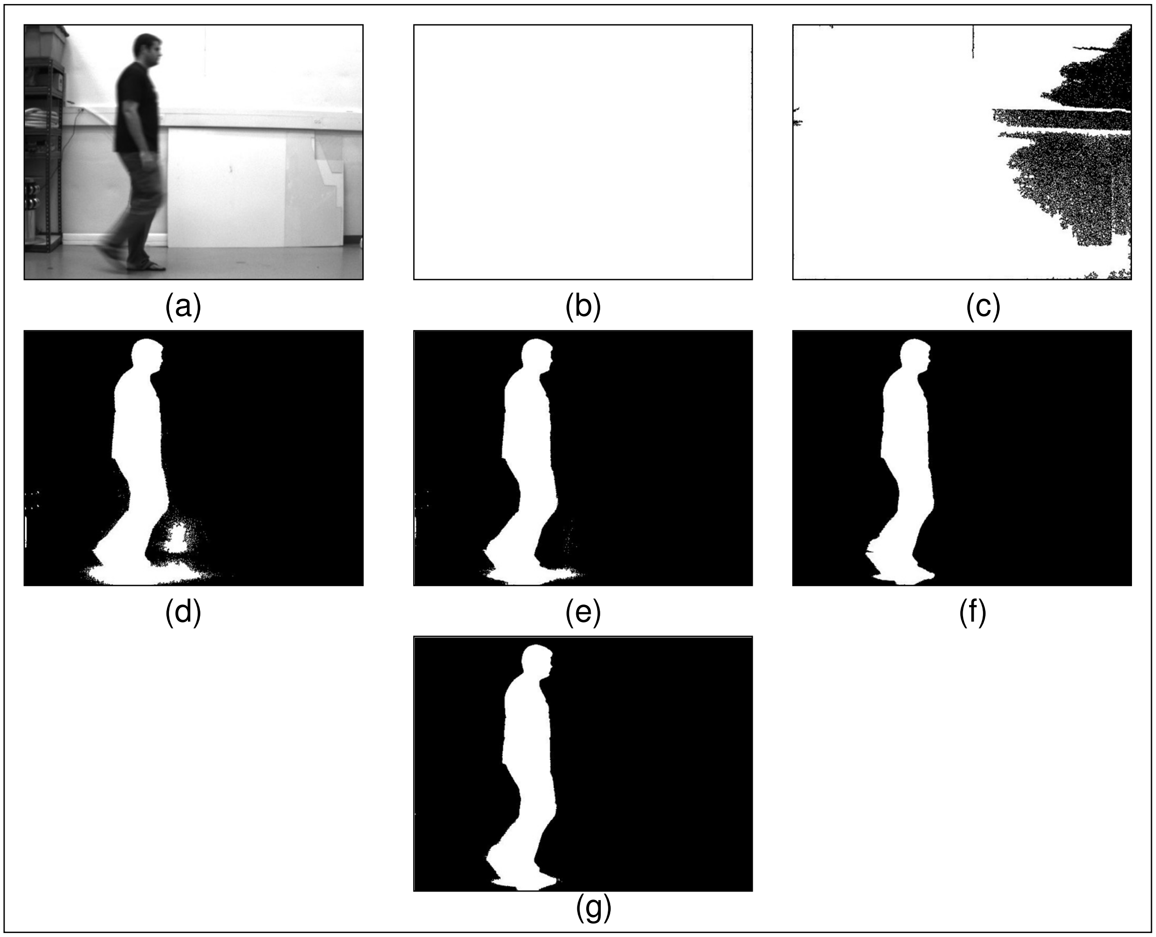


Figure 5.1: Results of the non-parametric background extraction algorithm, (a) shows the original image while the remaining images are results of the background extraction algorithm with the threshold set at (b) 0.05, (c) 0.10, (d) 0.50, (e) 1.00, (f) 2.50, and (g) 5

As can be seen there is a distinct difference in the output of the algorithm that depends on what the threshold value is set to, this threshold value does not necessarily 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 was chosen to minimize the amount of false positives and therefore increase the accuracy of the foreground detection algorithm.

Once the background has successfully been extracted from the input frame, a determination of foreground must be determined. While this step may seem unnecessary in light of the background extraction data shown above, often the binary information given from background extraction can contain small regions of false positives as well as regions of false negatives located within the silhouette of the subject. These may be overcome using a number of simple steps which serve to join parts of the silhouette that are disconnected in the image due to false negatives and also can eliminate erroneous areas, caused by false positives, from the final binary foreground image. The first step that is performed on the image data given by background extraction is a simple binarizations of the data in which all values above 0 are set to 1. This forms the initial binary silhouette of the data. Then binary dilation and fill operations are performed in order to attempt to compensate for any false negative regions within the subject silhouette. Dilation is the process of gradual enlargement of the boundaries of regions which serves to fill small gaps in the silhouette. This is accomplished by placing a structuring element over each of the pixels in the image, if any foreground pixels (1's in this case) 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 below.

1
1
1
1
1

Figure 5.2: 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 in a certain region; therefore, if a small area within the silhouette has a value of 0 the value is changed to 1 after the fill operation which makes the silhouette more coherent. An example of the fill and dilation operation being performed is shown below.

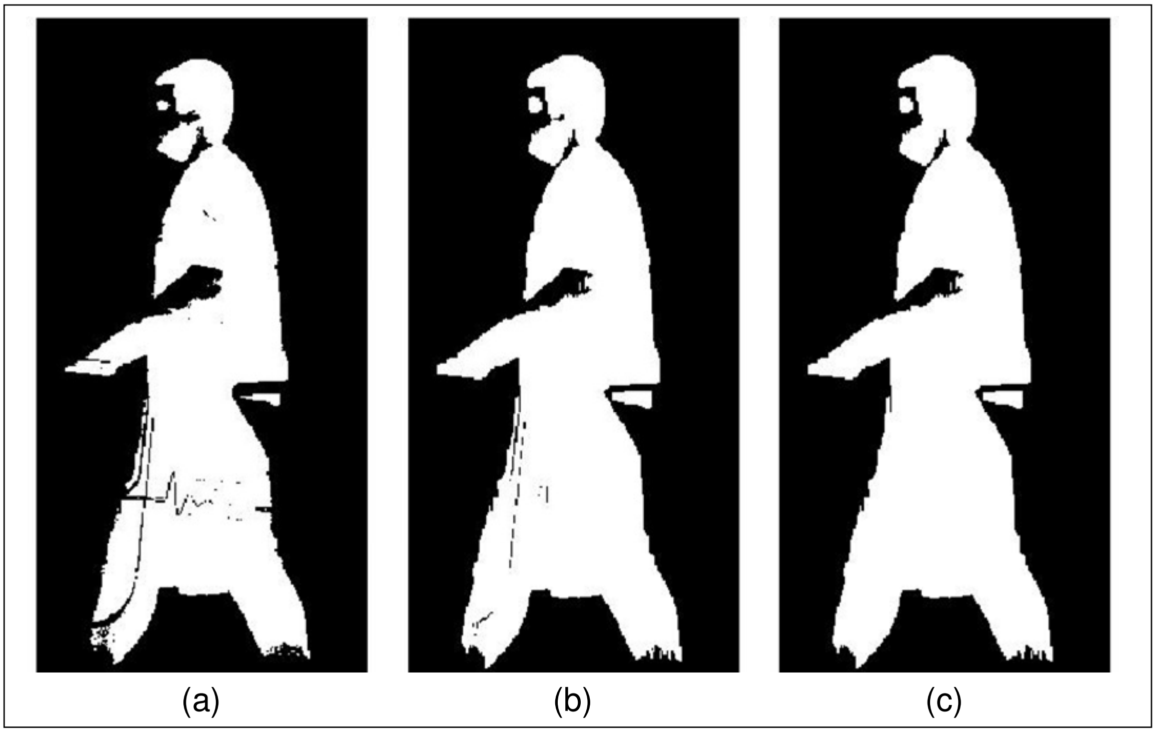


Figure 5.3: An example of a binary silhouette generate for use with the gait classification system, (a) is the binary silhouette after background extraction and application of a binary threshold, (b) shows the same silhouette after performing a dilation operation (c) is after performing a fill and dilation operation

SECTION 5.4 – FOREGROUND OPERATIONS

With a reliable silhouette of the subject established, the next stage of the system is to process the binary data in order to generate a number of values which are utilized at later stages of analysis in the system. The first information that is to be found using the binary silhouette is its centroid. This provides information on how the overall subject is moving in time as well as an approximate location of the hip. This information can be exploited 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.

$$\frac{\int xf(x)dx}{\int f(x)dx} \qquad (10)$$

Approximations of the centroid are used in this system in order to reduces the computational complexity of the over all algorithm. The centriod of the silhouette can be 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 results which lie close to each other on the silhouette, the method using area yields a result that is more consistent with the location of the center of the body, when viewed in the sagittal plane, and therefore closer to the actual hip location of the subject. The first method, that utilizing the boundary locations, is done by first finding the overall 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 found in the background extraction process. Finally 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 generated in the background extraction process, but can lead to a less accurate measure of the centroid in relation to the hip location and true center of the centroid due to false negatives in the background extraction process which can separate the silhouette in to a number of smaller parts which are excluded by the thresholding operation. A flow chart of the above method is found in the following figure.

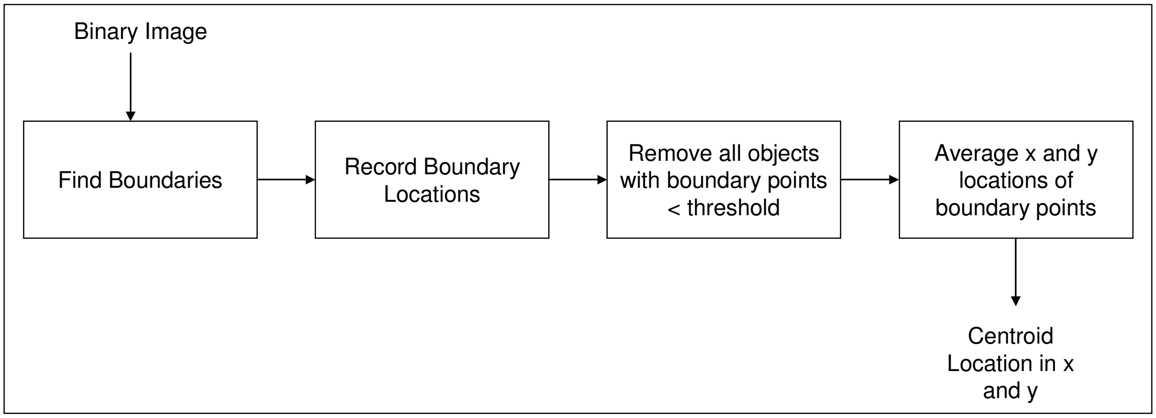


Figure 5.4: Flowchart of operations to approximate centroid location using boundary positions

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), these 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 small false positives generated by the background extraction process. The figure shown below shows the centroid location of a subject using both of the above described methods.

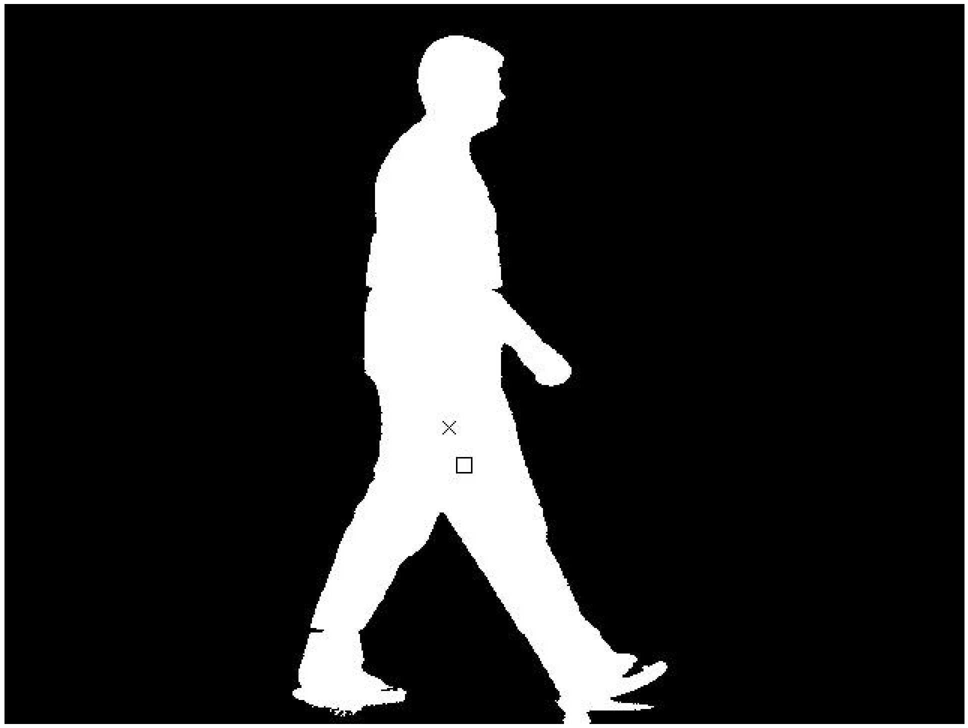


Figure 5.5: A binary silhouette with a relatively low noise level, the ‘X’ is the centroid found using the entire area of the silhouette and the square is the centroid found using the boundary method

Figure 5.6: 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 the figure, the centroid calculated by using the entire area encompassed by the centroid yields a result that is located much closer to the hip and also appears to be more accurate intuitively. The next step in the process is finding the location of the feet, this can be done a number of ways but in order to provide a more accurate hip angle the centroid of the feet is used. The centroid of a foot will provide an approximate location of the ankle; this position can in turn be used to calculate an estimated 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 to the area in which the foot should be located. This search area is chosen based on the limitation of movement during gait, the search area for the rear foot is located behind the body's centroid, up to 150 pixels, and up 60 pixels from the floor. The search area for the front foot is located from the body's centroid up to 150 pixels ahead of the centroid, and up 60 pixels from the floor. The search areas are shown in the superimposed boxes in the figure below.

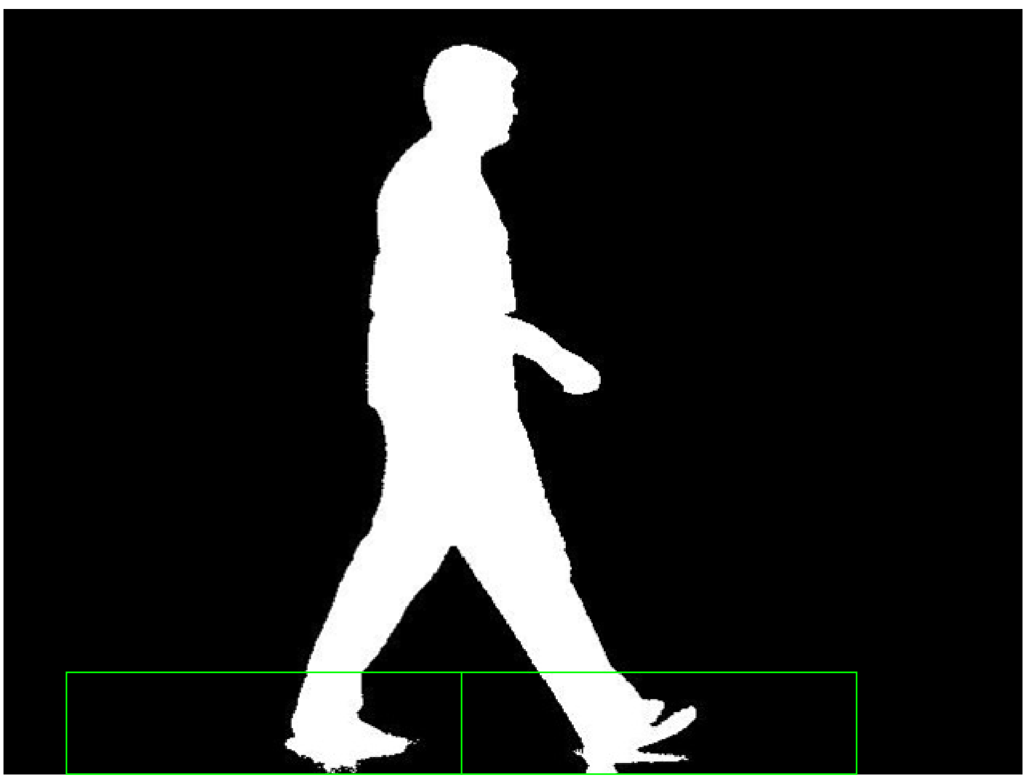


Figure 5.7: Search area for foot centroids superimposed on a binary silhouette

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 the figure below.

Figure 5.8: Foot centroids shown as an 'x' superimposed on a binary silhouette

As can be seen in the figures, the resulting foot centroids give a good approximation of the ankles of the leading and trailing foot. These foot centroid locations can then be used in conjunction with the body centroid in order to approximate the hip angle. This is done by utilizing the rule of sines as shown in the figure below along with associated equations.

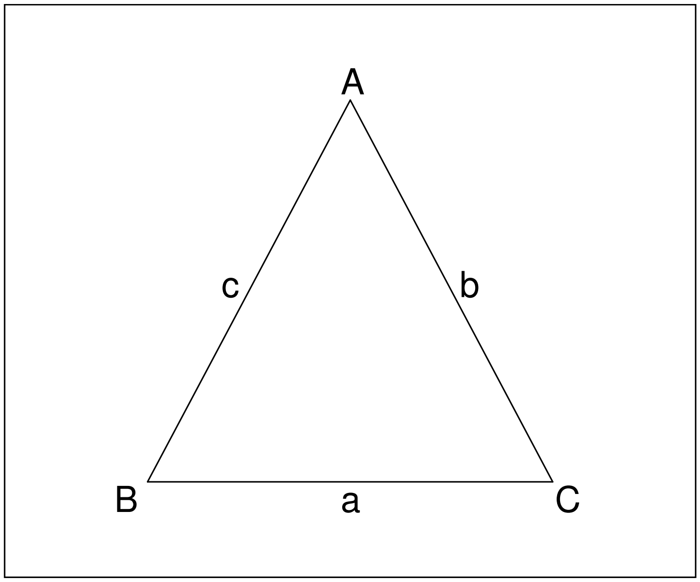


Figure 5.9: Segments Used for Hip Angle Calculation, where point B corresponds to the trailing foot, point C the leading foot and A the centroid, with the angle at A being the hip angle, c and b are the lengths of the legs, and a is the distance between the leading and trailing foot

$$\Theta = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \qquad (11)$$

where

Θ - hip angle

a,b,c – defined above in Figure 26 description

While this does not provide an exact measure of the hip angle, such as one that could be found using a commercial motion capture system, it does provide a good approximation to the actual hip angle and provide a sufficient description of how the feet are moving in relation to the center of the body. The centroid locations are connected and superimposed on the binary image of the subject's silhouette in the figure below.

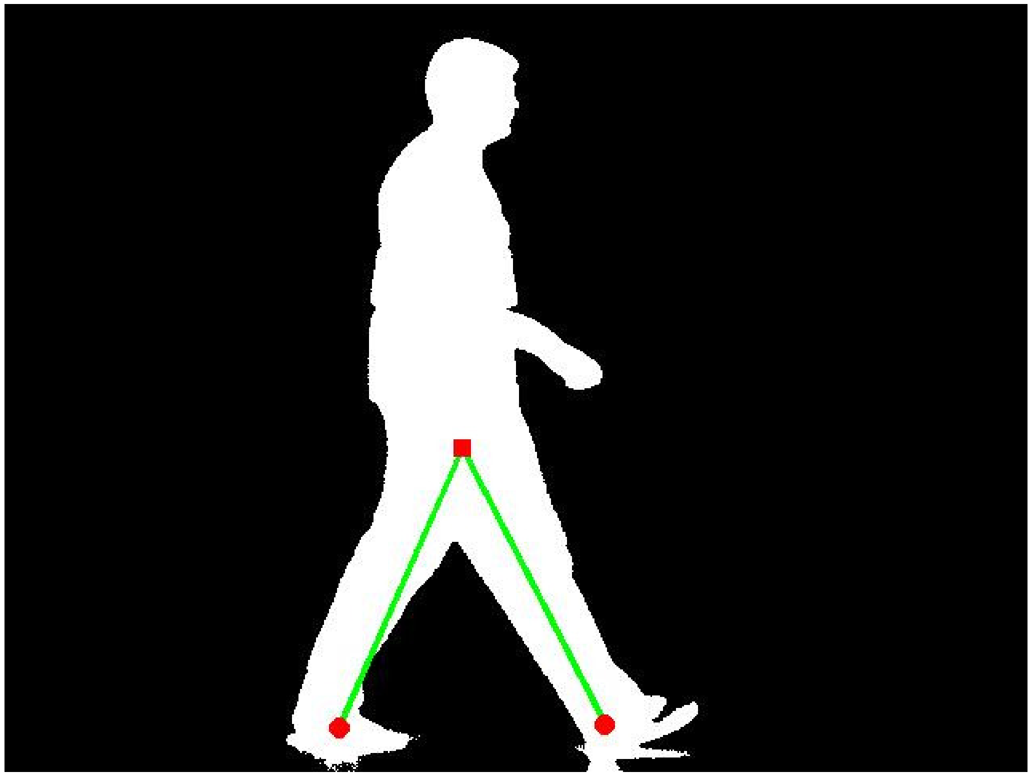


Figure 5.10: The binary silhouette with foot and body centroids superimposed as red circles and a red square respectively. The line segments used to estimate the hip angle are shown as green lines.

The time history of the hip angle measurement found for a normally walking subject is shown below in Figure 5.11; 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 can be used to provide a measure of the torso angle of the subject while walking. This is done by calculating the angle between the line formed by the highest point on the silhouette, approximately the top of the subject's head, and the body's centroid and the plane parallel to the ground. The points used to calculate the angle are shown in Figure 5.12.

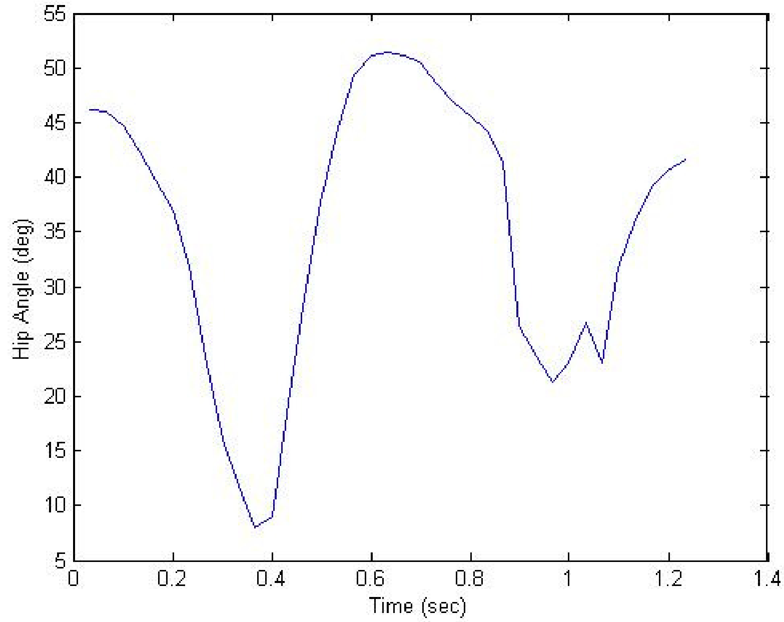


Figure 5.11: Hip Angle in degrees plotted versus time for a normally walking subject taken from video

Figure 5.12: 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 torso angle

Again, this does not provide an exact measure of the actual torso angle of the subject, but it does provide a good approximation of the orientation of the subject's upper body. The final values which are calculated using the foreground 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 with the max stride length being found by taking the maximum value of the stride length over the entire length of the gait video. The points used to find the height and stride length are shown in the figure below.

Figure 5.13: 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

SECTION 5.5 – Motion Estimation

In addition to spatial, or shape, information, motion provides another layer of descriptors which can be utilized to classify gait. Motion information is derived from a video sequence for use in this system using a block matching motion estimation scheme. The type of motion estimation scheme used allows for a quick estimation of all motion within the scene in the plane perpendicular to the camera's axis, if any motion is taking place in a different plane then the projection of that motion onto the plane perpendicular to the camera's axis is computed. This occurs since the camera's images are just the three-dimensional scene projected into the two-dimensional image space of the camera. An example of the layout that is used for testing of the system is shown below, where the optimal plane for proper motion estimation is labeled in relation to the camera's position.

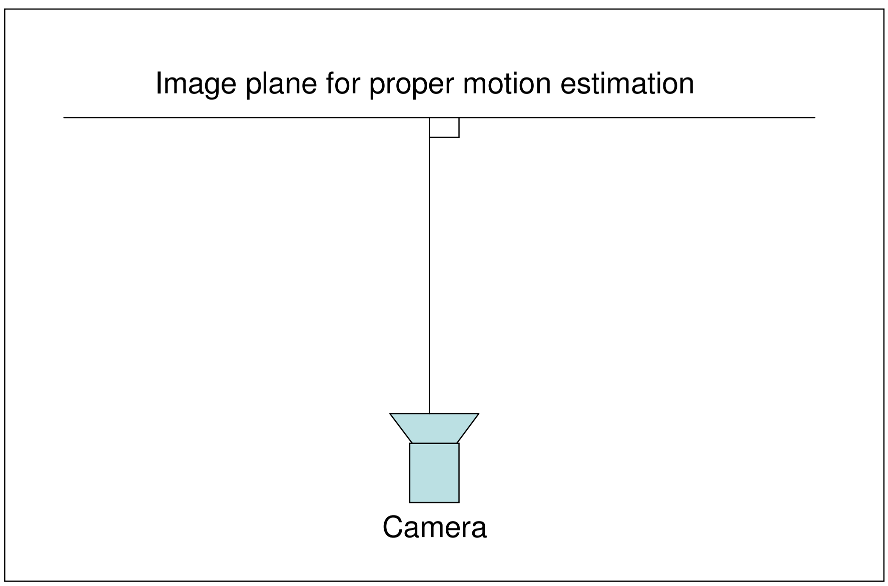
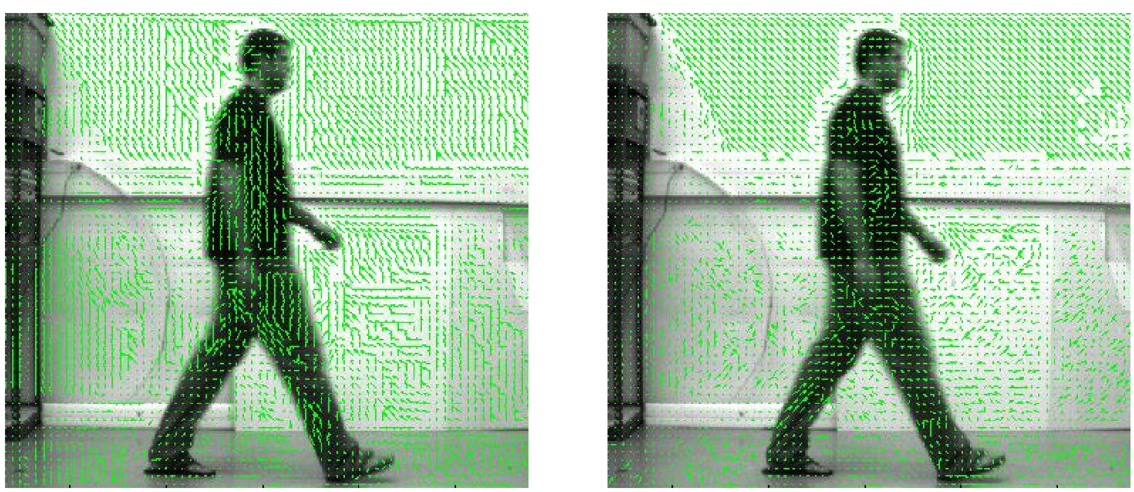


Figure 5.14: Overhead view of how the camera is setup for image capture

If motion does not take place in this plane then the motion calculated is no longer accurate, but is really a projection of the true motion onto this plane which does not equate directly to the motion calculated for the proper plane. However, the motion could be transformed to equate the motion taken in the proper plane if the plane of the observed motion is known along with other associated geometries of the scene.

With motion known to be taking place in the proper plane within the scene, a block matching scheme can then be utilized to estimate this motion. As mentioned in Section 2 of Chapter 2 there are many different features which can be chosen in a block matching scheme which will determine how the effective overall system is in estimating the motion taking place within the scene. The first of these features which must be chosen is the cost function which will be utilized by the search scheme. In Chapter 2 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 was chosen due to its simplicity and computational efficiency as compared to the various other cost functions, additionally the motion estimation calculated by the final algorithm when using the Mean Square Error cost function is found to be less accurate than the Mean Absolute Difference. The figure below shows two quiver plots of the motion calculated using the two cost functions described above.



(a) (b)

Figure 5.15: 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, and most important, of the chosen features is the search algorithm. In Chapter 2 the most popular of the various searching schemes were briefly discussed, with one of the more recent and computationally efficient algorithms, the Adaptive Rood Pattern Search algorithm, described in more detail. In order to choose an effective algorithm for use in this system, various search algorithms were tested in order to determine the most effective scheme for use with a gait classification system. 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 of the motion into account, it would seem that search schemes which utilize a diamond type search pattern would be the most effective at efficiently and accurately calculating the motion taking place within the scene. Because of this, two different search schemes are attempted, the Diamond Search and Adaptive Rood Pattern Search, additionally the exhaustive search algorithm is used to provide a motion accuracy baseline to compare the various systems since the exhaustive search algorithm provides a "best case" for motion estimation using a block searching scheme [1]. While accuracy is of paramount importance with this motion estimation system, computational efficiency is also a concern, with this in mind the number of computations, as measured by number of iterations needed by the search pattern to estimate the motion fir 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 used to calculate the motion; 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 2.

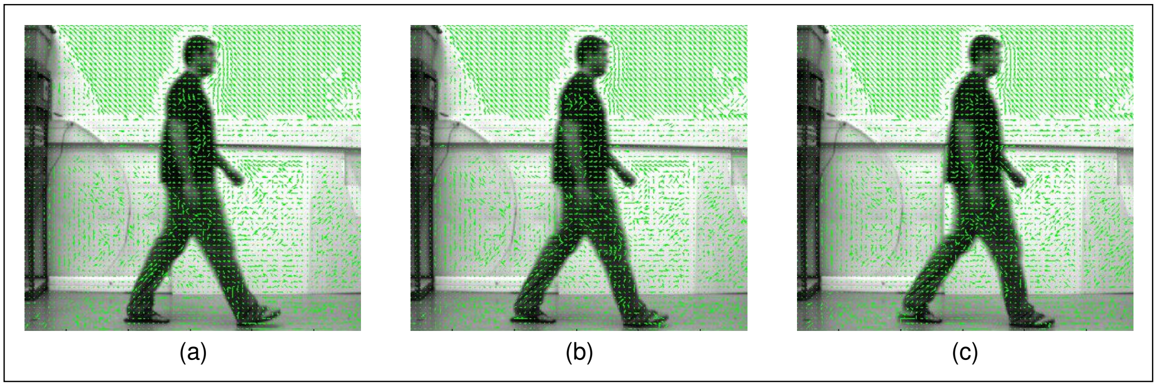


Figure 5.16: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Exhaustive Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video, notice the large amount of errors

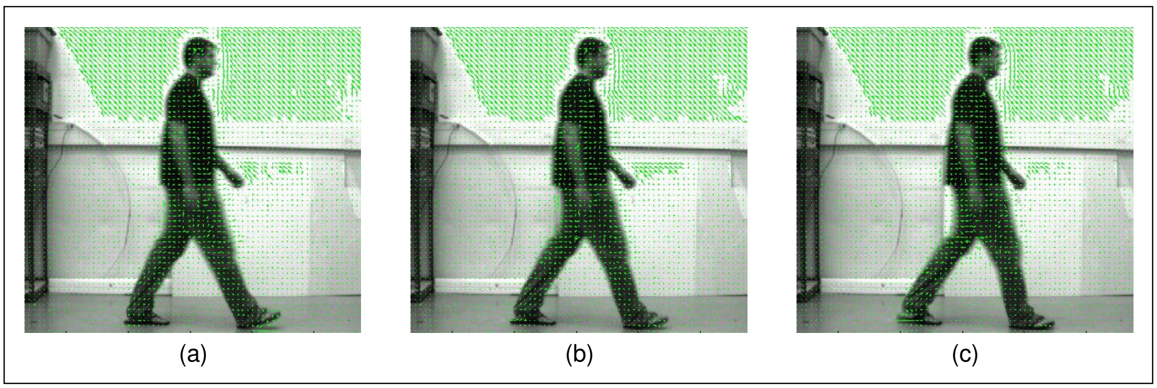


Figure 5.17: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Diamond Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video, notice the large amount of errors

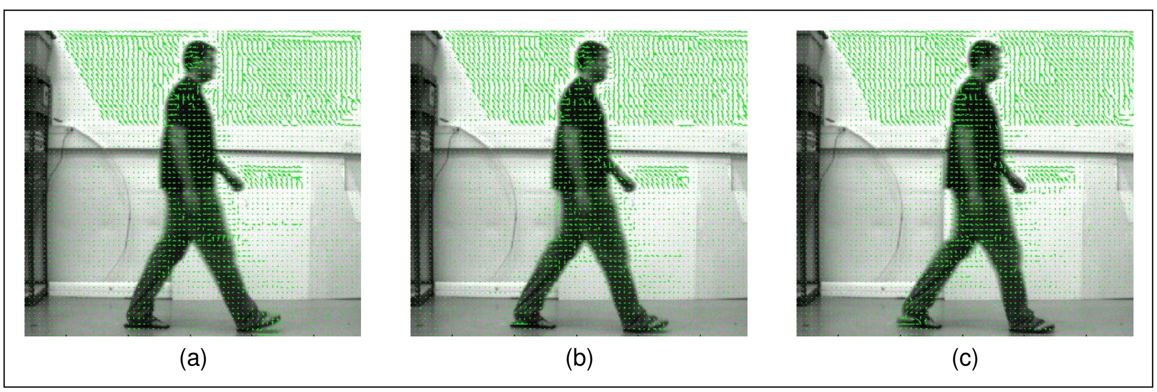


Figure 5.18: The figure shows three consecutive frames (a, b, c) taken from a subject walking normally; the motion estimation is calculated using an Adaptive Rood Pattern Search block matching algorithm, the motion for each block is plotted as an arrow overlaid on the frame of the video

As can be seen 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 provide a more concrete measure of performance. To accomplish this, the peak-signal-to-noise ratio is calculated for the motion compensated image as described in [1], basically the motion information is used to construct the next frame in the video sequence and then the error in the motion compensated image is calculated. This is done by using the following equation.

$$PSNR = 10\log_{10}\left[\frac{(p_N)^2}{MSE}\right] \tag{12}$$

where the MSE is calculated as

$$MSE = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(I(i,j) - M(i,j)\right)^2 \tag{13}$$

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 motion compensated image is created by moving the blocks in the first image in the sequence to the points specified by the motion information calculated by the block matching algorithm. The peak-signal-to-noise ratio found for the image sequence used above and the three search methods chosen for analysis is given below.

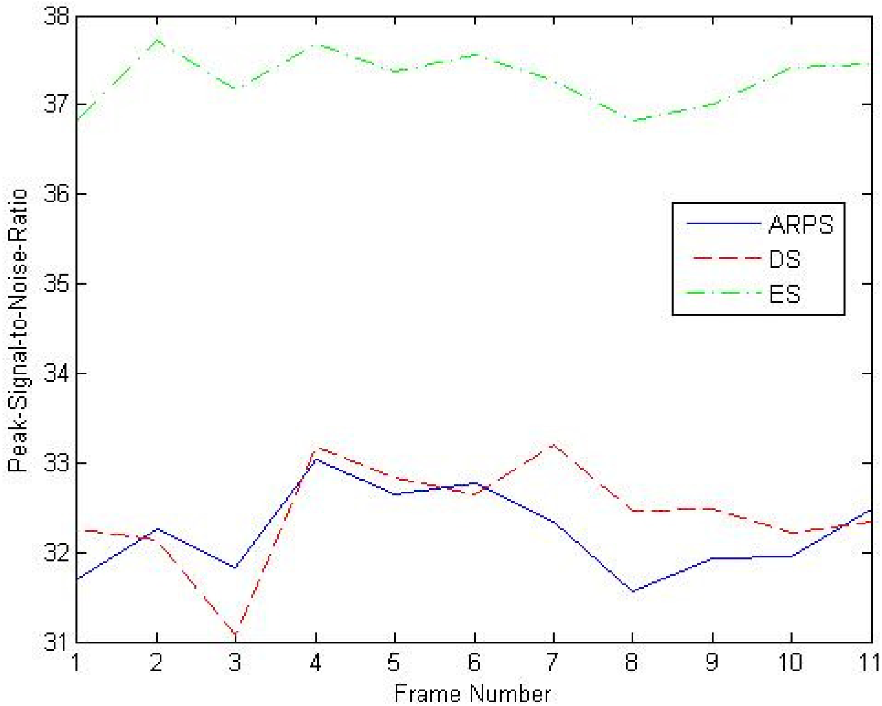


Figure 5.19: Plot of the peak signal to noise ratio for the three search methods, Adaptive Rood Pattern Search (ARPS), Diamond Search (DS) and Exhaustive Search (ES)

While the accuracy of the motion calculation is of utmost importance, the computational efficiency of the search algorithm should also be considered and an analysis of the trade-off between algorithm accuracy and efficiency should be done. In order to measure the computation efficiency of the above search algorithms the number of search iterations taken by the algorithms to arrive at the motion estimation for the image is measured. A comparison of this measure for each of the above algorithms is shown in the figure below.

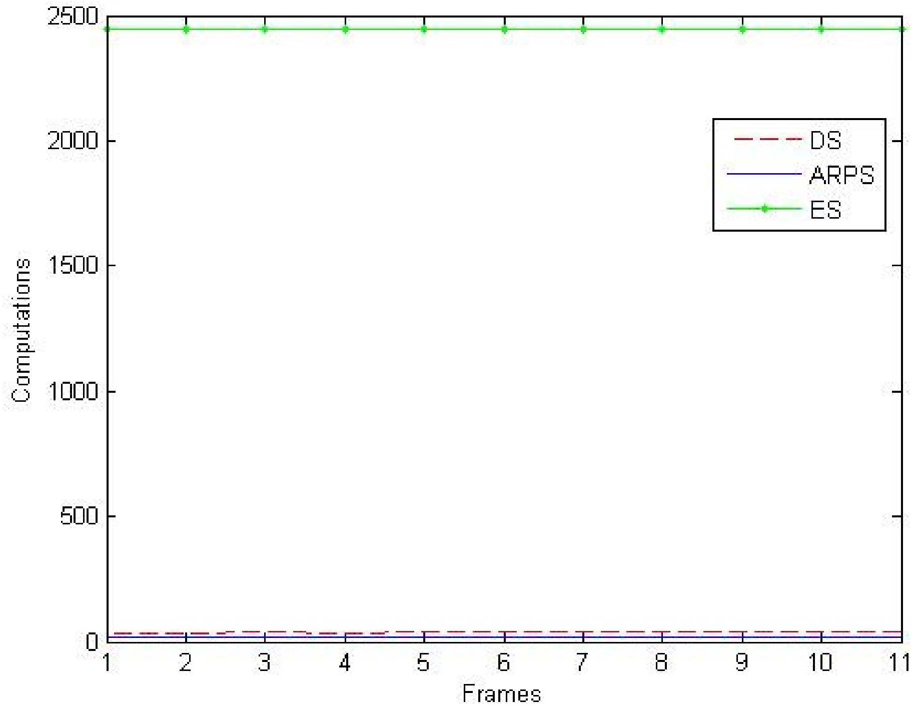


Figure 5.20: Plot of the required number of computations for each search method to arrive at its respective motion estimation for each frame

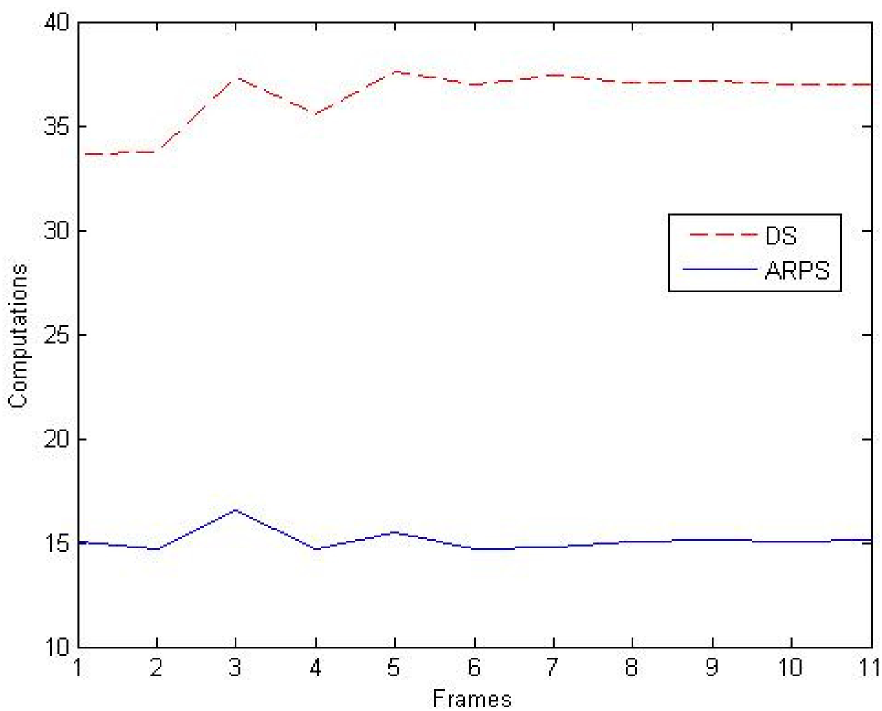


Figure 5.21: Plot of the required number of computations for ARPS and DS search methods to arrive at their respective motion estimations for each frame

As can be seen in the above figure, there is a large difference in the computationally efficiency of the three algorithms. As can be seen in Figure 5.20, 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 to arrive at its motion calculation. 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. Taking the above data into account, the use of the Adaptive Rood Pattern Search algorithm for the proposed system is obvious.

With the search algorithm selected, the next step is to select the block size as well as the size of the search pattern to be used in the block matching scheme. The search pattern must be selected so that none of the motion taking place within the scene occurs at a distance greater than the search pattern from one frame to the next. However, a larger search pattern will also result in a more computationally expensive algorithm, so care must be taken to not set the search pattern to large. From observation of normal human gaits using video analyzed by the above search algorithms, a search pattern of 25 is chosen in order to totally encompass all motions taking place during normal and abnormal walking while at the same time not taxing the system with unnecessary computations. The block size is the next parameter which must be chosen. Block size determines how accurate the motion estimation operation is while at the same time governing the motion resolution which can be obtained by the algorithm. A larger block size will provide more information for the block matching algorithm at each block, but objects and areas moving within the scene which are smaller than a block could actually cause a large amount of error in the motion estimation. Additionally, a smaller block size causes the algorithm to perform more searches over the image space which in turn makes the algorithm computationally more costly. 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. With all search parameters established, the next step in creating the final motion estimation algorithm is to utilize the information found by the system thus far to create a more accurate and efficient gait analysis system.

SECTION 5.6 – ALGORITHM FUSION

With all information gathered by the system thus far, an efficient fusion of all the above techniques, background extraction, foreground determination and operations, and motion estimation, is desired. Each method should not operate individually when information could be passed from one algorithm to another in order to create a system which operates solely on regions of interest while not spending computational resources on areas of an image which remain unchanged from frame to frame. To accomplish this, foreground areas determined in the previous steps of the system are to be utilized in order to define a search area for the optimized motion estimation algorithm. This is done in order to reduce computational time used to calculate the motion taking place within the frame and also to reduce errors associated with edge regions found when using block matching schemes for motion estimation.

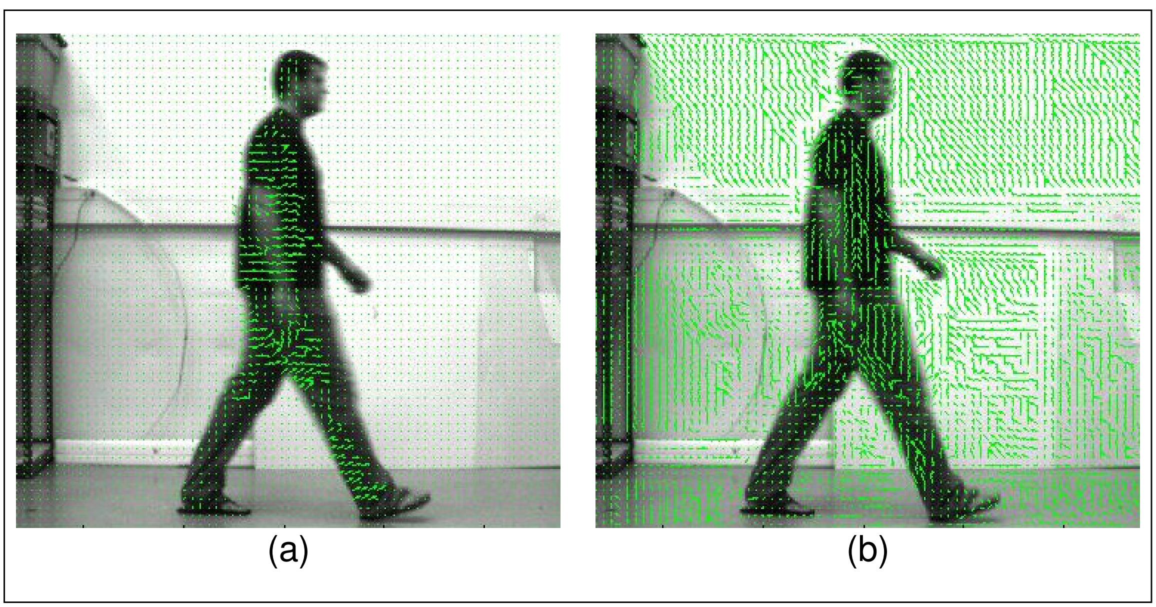


Figure 5.22: The result of algorithm fusion for this system, both images have the associated motion information superimposed, (a) is the algorithm used in the system and (b) is the traditional ARPS algorithm

In the implementation of this system, the foreground location information determined in the previous phases of the system is passed to the motion estimation algorithm. This information is used in order to define a search area for each frame of the image; this succeeds in eliminating errors caused by the edge effects described above and also eliminates search operations in areas of the frame where no foreground object exists. This serves to eliminate all the errors in the motion estimation caused by the background of the scene which in turn makes the motion estimation algorithm much more accurate. Due to these improvements, the motion information calculated from the video sequence is much more valuable because of a great reduction in noise. This allows further changes to be made which can in turn make the data easier to analyze in subsequent system operations.

An additional operation which is carried out on the motion data using values calculated in previous steps is the windowing of the motion data. What this entails is correcting the motion location for each image frame so that the locations of the upper body stay stationary throughout the video sequence. This way all motion of the subject is located relative to the upper body. This allows motion of a specific region or part of the body to be analyzed solely on its motion information and not its movement within the scene. A number of variables are used to accomplish this, including stride length and centroid information found in the analysis of the foreground objects. The window is created using the entire height of the frame, but is limited in the horizontal direction using the maximum stride length observed in the video sequence. This allows all motion taking place in the scene for a given subject to be wholly contained by the window. The next step in the analysis is to correct 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. This is accomplished by using the centroid location of the subject for each frame in order to calculate an offset which will 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 for each of the motion vectors which results in all motion vectors being located in the window. Resulting from the operation is a video sequence consisting wholly of motion vectors which only move in relation to the centroid, this provides a valuable source of information that can be utilized by successive steps of the system for pure motion analysis. A block diagram of the system operations up to this point in the processing is shown below.

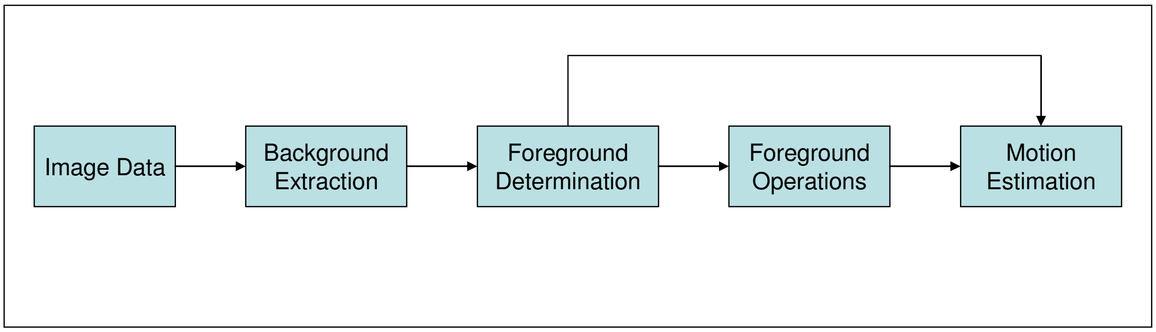


Figure 5.23: Flow chart of the operations performed by the system to form descriptors of human gait

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

SECTION 5.7 – SIGNAL RECOGNITION

Once all the descriptors of the subject's walking have been calculated and complied by the system, the next step is to analyze the available information to accurately determine if the subject walking in the scene is indeed walking normally. As mentioned above in Chapter 4, there are a variety of steps which must be taken in order to analyze data such as this. A recognition scheme consists of a number of steps including, feature generation and selection as well as selection of a classification scheme. Each of these steps determines the effectiveness of the overall pattern recognition system's data classification.

The first step of designing this pattern recognition scheme is to decide on the type of classifier that is to be used in the system. Of the many types available, the perceptron neural network algorithm is chosen for use in this system. This type of algorithm is chosen based on its simplicity, ease of implementation and ability to be trained using actual data gained from the system. The perceptron algorithm is described in detail in Chapter 5. 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 conjunction with the chosen pattern recognition scheme. Because the gaits are to be analyzed based on there departure from normal gait, a comparison between the incoming gait and normal gait would be a favorable feature for use in this pattern recognition scheme. A number of these types of quantification schemes are described in Chapter 4, due to the limitations imposed on the type of analysis which can be performed, based on the incoming data, a measure of how well the incoming data matches a normal signal using correlation or some kind of error calculation is desirable. In order to provide an accurate measure of how well an incoming signal correlates to a normal walking signal, the incoming signal is compared to a number of normal walking signals, four for this implementation, and the similarity values generated for each of the normal signals is averaged in order to arrive at the overall similarity measure for the incoming gait signal. Taking the above information into account, three types of signal quantification are chosen to be tested for use with the system, cross-correlation at zero lag, correlation coefficient and error calculated between the two signals. The error term which is to be used is shown below.

$$\frac{\int_1^N (X_n - N_n)^2 \, dn}{\int_1^N N_n^2 \, dn} \tag{14}$$

where

X_n - n$^{\text{th}}$ point of the incoming signal

N_n - n$^{\text{th}}$ point of normal signal

N – total number of points in the signal

Each of these quantification methods are tested using a set number of features in the perceptron architecture in order to determine which method provides the highest level of accuracy in the system. This so called suboptimal search method is utilized because it can often be difficult to estimate the effectiveness of the final classification system using seperatability criteria when a large number of features are utilized by the classification system [8]. An example of a number of scatter matrices is shown for two feature values generated using the correlation coefficient.

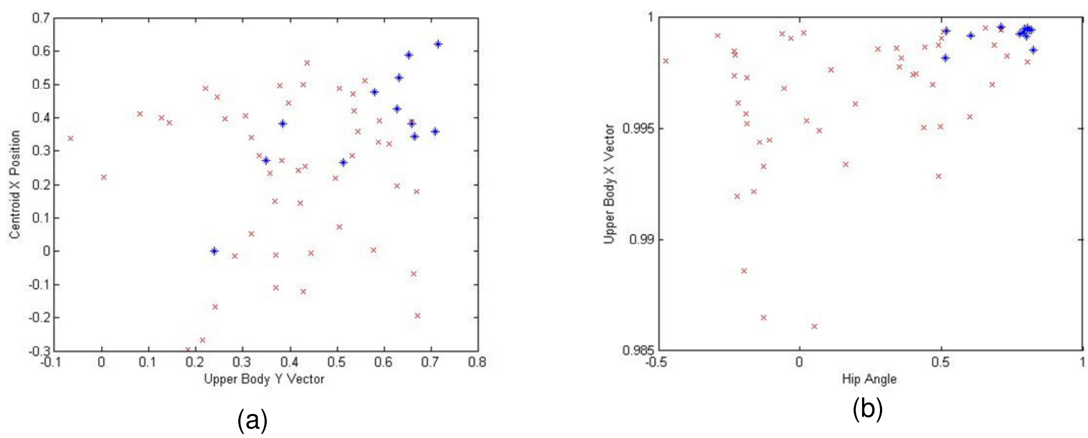


Figure 5.24: Scatter plots of feature values for use in the classification system, (a) shows a plot of Upper Body Y Vector vs Centroid X Position and (b) shows Hip Angle vs Upper Body X Vector. The blue stars represent normal walking data and the red x's represent abnormal walking data. The most effective 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

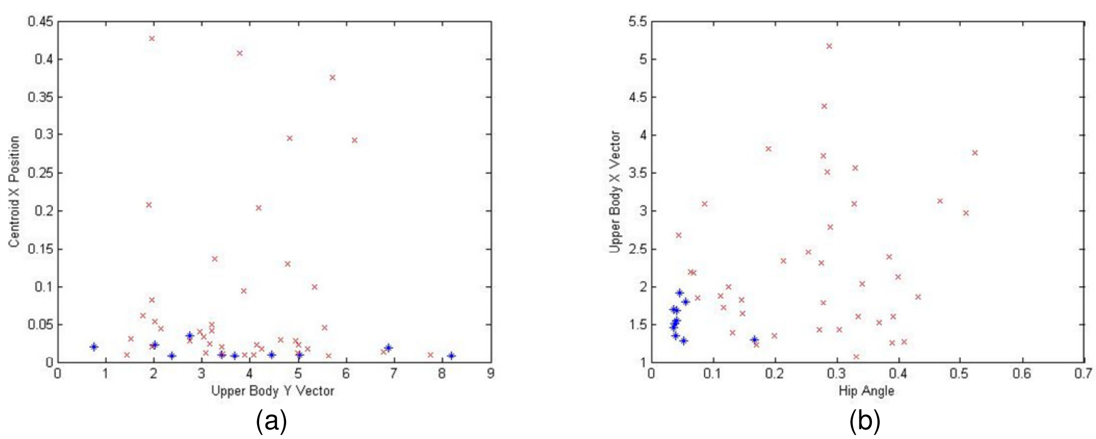


Figure 5.25: Scatter plots of feature values found using the error term, (a) shows a plot of Upper Body Y Vector vs Centroid X Position and (b) shows Hip Angle vs Upper Body X Vector. The blue stars represent normal walking data and the red x's represent abnormal walking data.

Figure 5.25 shows the scatter plots using the error term, as can be seen in the above figures, it is very difficult to see which feature generation method will provide a better linear separation boundary (since a one layer perceptron is being used) when using more than two variables. It is because of this that a set number of features are used in conjunction with the perceptron to determine the effectiveness of the feature generation methods, the method which yields the most accurate classification system is the feature generation method that provides the greatest classification effectivness and is therefore the method chosen. 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 then used to test the effectiveness of the overall training system. The results from the test are shown below.

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

Table 5.1: Results of Feature Generation Testing

As can be seen in the table of results found above, when the correlation coefficient is utilized in the feature generation stage 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 utilized for this classification system.

With the feature generator selected, the next step is to decide which combination of features yields the most accurate classification system while utilizing the least number of features. This is again done using a suboptimal search method due to the reasons stated in the above paragraph regarding the complications of predicting the classification of the final system based on various selection criteria. 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 which 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 again the combination yielding the most accurate system is then 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. When this process is completed the most accurate system is found utilizing the desired number of features, for this 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.

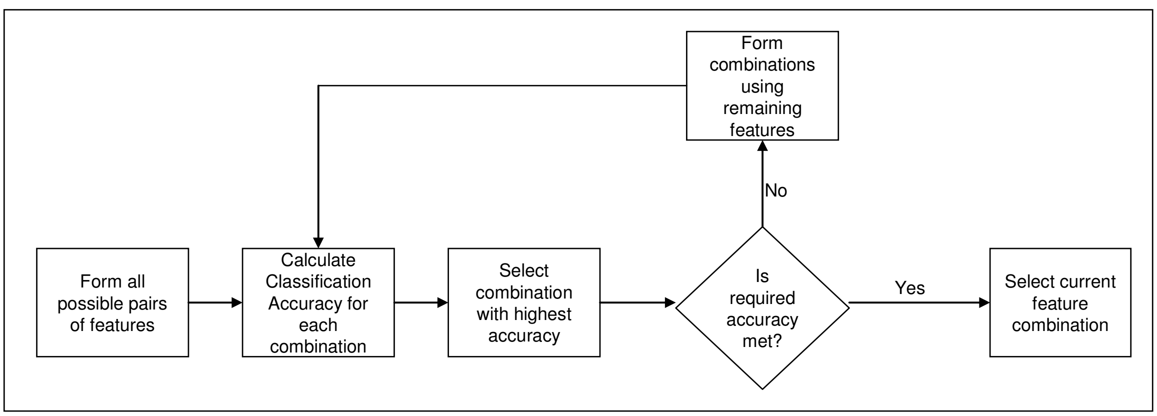


Figure 5.26: Flowchart of suboptimal search method technique used for feature selection

CHAPTER 6 - RESULTS

SECTION 6.1 – INTRODUCTION

With the system designed as described above in Chapter 5 using data obtained from the vision system, the system must now be verified and tested in order to determine the effectiveness of gait classification. This is done by taking videos of a number of subjects walking normally as well as abnormally in order to create a baseline set of data with which to train the system and then later to test the effectiveness of the system.

SECTION 6.2 – SETUP

In order to test the system properly it is desired to train and test the system using video of actual people walking in front of a camera. This is done because the complexity of the human body does not allow for easy modeling of how various individuals walk in an exact way. Additionally, abnormal walking again provides a scenario in which individuals will react differently to different movements which is difficult to model with enough accuracy to capture subtle individual differences. Due to these facts, testing and training of the system is done in a laboratory environment in which various human subjects are instructed to walk in front of a camera in such a way that will provide data consistent with what is needed by the system for classification.

All data collected for this system was done in a controlled indoor environment in which lighting was kept relatively constant. A walking path was created in an open area over a flat, clean floor in order 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 other to create a straight walking path for the subjects. The camera used for image acquisition was placed so that it was perpendicular to the plane created by the walking path; the area seen behind the walking path was kept static during the duration of each data collection in order to reduce errors created by a moving background. One subject's data was collected in front of a background which induced noise due to movement in order to determine the robustness of the system to noise. The background used was a white drop cloth which moved in a manor which was not accounted for by the background extraction algorithm when the subject walked in front of it. The setup for this data collection processes is shown in the following figure.

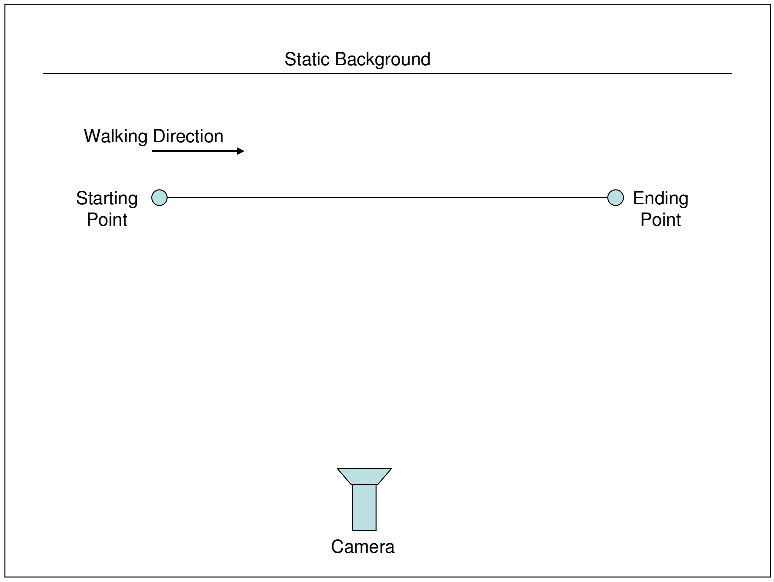


Figure 6.1: The experimental setup, an overhead view

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 which could be analyzed using the classification system.

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

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

These trials result in having one normal walking data set for each subject and four walking sets with different abnormalities being exhibited in each individual video sequence. These five different walking styles were done twice by each subject, this resulted in a total of ten separate data sets that were created for each individual subject. This provided a large amount of data for each subject which could be compared between and within subjects in order to classify the observed gait. Because of the number of trials per subject, 6 subjects were chosen for use in this project. In order 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 the table below.

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 6.1: 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.

All data collection and processing was done by the system described in Chapter 5, an Imaging Source DMK21AF04 CCD camera configured to collect video at 30 frames per second with a resolution of 640x480 pixels per frame is used to collect video data. This data is subsequently imported into a computer via Matlab's Image Acquisition Toolbox. The data is then fed into the system where background extraction, foreground determination, spatial processing, and motion estimation processes are completed. With this data computed by the system, it is ready to be fed into the classifier for classification.

The classifier is trained using 5 of the normal data sets and 5 of the abnormal data sets from various test subjects, the remaining data from the test is then processed using the system and the final classification decision for each of the subject's walking trails is computed. This output is then compared to the known gait classifications and the effectiveness of the system can then be determined based on the accuracy of the final system classification, as given by a percentage of correct classifications.

SECTION 6.3 – RESULTS

The final system accuracy obtained for the testing described above using the system described in Chapter 5 is found to be 93.22%. This means that the system will classify an observed gait correctly roughly 93% of the time. The results for each individual subject can be found in the table below, the subject walking with the 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 6.2: 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 which is wholly accounted for by the background extraction algorithm used by the system, 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 in order to model the type of background movement taking place in the scene; Subject 6's normal gait is incorrectly classified due to the error caused by the background movement. 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.

SECTION 6.4 – CONCLUSIONS

As can be seen in the above results, the proposed system provides an efficient and accurate determination scheme for determining if a human is walking normally or abnormally. Even though the abnormalities used for evaluation of this system are not "true" gait abnormalities, they are instead feigned by individuals who walk normally, they provide subtle difference in the human gait pattern which 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 as well as being more repeatable from cycle to cycle. Greater repeatability would provide more stable data to the classification system and theoretically better results from the system as a whole. A system which can determine if a subject is walking normally would have applications in a variety of systems.

A system such as this would be critical in the design of many complicated gait analysis systems, such as those used for biometrics or other gait identification systems as well as other gait classification systems. A system which can determine if a subject is walking in a certain way could assist a more in-depth gait analysis system by providing information on whether the gait should be analyzed by the system at all or if different criteria should be implemented to analyze the data based on a certain abnormality. A computationally efficient and fast system providing accurate classification of an observed gait that can be done before a more exhaustive system is utilized to examine the gait would prove a valuable first step to many gait assessment algorithms.

CHAPTER 7 - CONCLUSIONS

SECTION 7.1 – 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, harmonic system analysis as well as pattern recognition in order 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 in order to produce an effective and efficient algorithm which generates a wide variety of variables which describe the observed gait. The data generated from the video using the above methods is then fed into a classification system, which uses all available gait descriptors to accurately arrive at a determination for the observed gait.

The performance of the system is tested using real data captured from people performing various motions in front of the camera. The system achieves good classification results and proves to be robust to the noise induced in real image sequences as a classification accuracy of 93% is achieved when using real video data.

SECTION 7.2 – FUTURE RESEARCH

The system performs well in this limited classification task of determination whether observed human gait is normal or abnormal. However, a more in-depth classification system would provide a more valuable addition to any number of applications in the area of gait analysis. The expansion of the classification introduced in this system to include abnormality classification instead of just a abnormal or normal classification would prove to be a valuable addition for use in a variety of applications, including unobtrusive clinical gait analysis, automated surveillance in addition to a variety of others. This system will continue to be developed to provide a more robust and in-depth classification of gait using video data which is unobtrusively obtained from walking subjects.

APPENDIX A: FEATURE SELECTION VALUES

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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