**STATISTICAL APPROACH TO BACKGROUND SUBTRACTION FOR
PRODUCTION OF HIGH-QUALITY SILHOUETTES
FOR HUMAN GAIT RECOGNITION**

THESIS

Jennifer J. Samler, Captain, USAF

AFIT/GAM/ENC/06-04

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GAM/ENC/06-04

**STATISTICAL APPROACH TO BACKGROUND SUBTRACTION FOR
PRODUCTION OF HIGH-QUALITY SILHOUETTES
FOR HUMAN GAIT RECOGNITION**

THESIS

Presented to the Faculty of the

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science

Jennifer J. Samler, B.S.

Captain, USAF

September 2006

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT/GAM/ENC/06-04

STATISTICAL APPROACH TO BACKGROUND SUBTRACTION FOR PRODUCTION OF HIGH-QUALITY SILHOUETTES FOR HUMAN GAIT RECOGNITION

Jennifer J. Samler, B.S.

Captain, USAF

Approved:

_____   _____
Samuel A. Wright, Maj., USAF   Date
Thesis Advisor

_____   _____
Dr. Steven C. Gustafson   Date
Committee Member

_____   _____
David C. Kaziska, Maj., USAF   Date
Committee Member

Acknowledgments

I would first like to thank my family. Without your help I would not have made it this far. To my husband, I thank you for your support and understanding while I was suffering through school. To our child, I would like to thank you for waiting till I was done with my thesis and graduation. To my parents, thank you for all the late night conversations, pre-proof reading and encouragement along the way. To my other parents, thank you for your support and encouraging words. You have all been behind me and believed in me every step of the way. If it was not for all your support I would not have the honor of graduating. I will always be grateful.

I would also like to thank Maj. Sam Wright for being my thesis advisor and my mentor through out my graduate studies at AFIT. It has been my pleasure learning under you. I would also like to extend my thanks to Dr. Steven Gustafson and Maj. David Kaziska for taking the time to be on my thesis committee. Both your mentorship and willingness to help me during my studies inspired me to succeed. My thanks are also extended to Dr. Rick Martin and Maj. Kyle Novak. Without your help and patience I would have never learned all the MATLAB I needed.

Jennifer J. Samler

Table of Contents

List of Figures

List of Tables

AFIT/GAM/ENC/06-04

Abstract

This thesis uses a background subtraction to produce high-quality silhouettes for use in human identification by human gait recognition, an identification method which does not require contact with an individual and which can be done from a distance. A statistical method which reduces the noise level is employed resulting in cleaner silhouettes which facilitate identification.

The thesis starts with gathering video data of individuals walking normally across a background scene. The video is then converted into a sequence of images that are stored as joint photographic experts group (jpeg) files. The background is subtracted from each image using a developed automatic computer code. In those codes, pixels in all the background frames are compared and averaged to produce an average background picture. The average background picture is then subtracted from pictures with a moving individual. If differenced pixels are determined to lie within a specified region, the pixel is colored black, otherwise it is colored white. The outline of the human figure is produced as a black and white silhouette. This inverse silhouette is then put into motion by recombining the individual frames into a video.

STATISTICAL APPROACH TO BACKGROUND SUBTRACTION FOR PRODUCTION OF HIGH-QUALITY SILHOUETTES FOR HUMAN GAIT RECOGNITION

I. Introduction

Background

Consider the idea of distinguishing individuals based on their gait. The human gait is a potentially valuable biometric for distinguishing people from a distance, in part because it requires no physical contact with the individual and is unlikely to be obscured, so a person walking along a public street could easily be the subject of a gait classification study. Since gait recognition requires no contact (like face recognition), privacy issues in traditional biometrics can be avoided (Lie, 2005: 767). In addition to the privacy issues, gait recognition can be done from a distance and the individual does not have to be aware of the procedure. The individual does not have to initiate or even have to be distracted by the process (Lie, 2005: 767). Therefore, the human gait has potential for widespread use within the Department of Defense and the Air Force. Being able to recognize someone without having to be close to the person could provide a great advantage, e.g., recognizing an individual from a reconnaissance aircraft mission recording.

For many years surveillance cameras and sound recording have been used as a security means in many public applications. In most applications, the collected video is monitored by security guards and stored for later evaluation, if needed (WP-11, 2006: 1).

In many cases the security guards have too much information to absorb and risk missing important clues. More detailed information of an individual can be found by their individual movements: gestures, body movements or facial expressions, to name a few (WP-11, 2006: 2). With this information surveillance cameras could highlight important details and allow much more complex interpretations.

Problem Statement

If the human gait is unique to every individual, a person can be identified by their gait. The classification of individuals' gait with the notion of distinguishing one person's walk from another is the overarching goal.

Research Objectives

There are many steps to take for the detection of an individual's gait. This thesis begins the process. The goal is to prepare the video stream so a gait classification algorithm may be performed. The video will be prepared by first removing the background.

This first step involves the removal of the individual from the background. The idea is to obtain a silhouette of the moving individual from the background to reduce the noise level from the background for better identification the individual by their gait. Performance of identification methods are greatly affected by the quality of silhouettes. This thesis provides an improvement in silhouette quality.

Research Question

The overall research question for this study is: If the human gait is unique to every individual, can a person be identified by their gait?

This thesis begins the process of answering this question. The specific research question which encompasses the work done in this thesis is:

Can the background scene of a video be effectively removed from the movement of the individual in the video?

Thesis Organization

Chapter II reviews the considerable body of literature about work that has already been done on gait recognition. Gait recognition is defined and its various uses is presented along with a brief history of the research. The chapter discusses the available literature concerning the effectiveness of using an individual's gait as a means of identification and the two primary classes of gait recognition: model-based and model-free.

Chapter III details the methodology used in the research of this thesis. Since the research focuses on people, the data collection phase is an important consideration, it is highlighted in the chapter. Descriptions of the purpose of each of the MATLAB programs that are central to the automated process of background removal are included. The chapter also contains some of the limitations of the research.

Chapter IV contains the results and analysis for the thesis. The chapter begins with a discussion of the advantages and disadvantages of the data collection. The rest of the chapter examines each of the MATLAB programs and their outputs in some detail.

The video streams of two individuals are examined frame by frame to show the effect of the employed methodology. An effective image of just the silhouette with less noise can be achieved using the developed MATLAB code together with the MATLAB medfilt2 filter.

Finally, Chapter V is the conclusions and recommendations of the thesis. In this method the moving individual is separated from the background. A successful sequence of frames with a white silhouette of the image on a black background is produced through implementation of our method. This thesis demonstrates the removal of the background behind a walking individual by means of an automated process is possible. After background removal, it does seem to be plausible to identify an individual based on their gait. Possible avenues for future research are also identified. This methodology yields an important step towards an automated gait recognition and identification program.

II. Literature Review

Chapter Overview

 The purpose of this chapter is to discuss the studies in the literature concerning gait recognition. "The Oxford Dictionary definition of gait is '*manner of walking, bearing or carriage as one walks*', suggesting that studies can concentrate on different facets of a person's walk" (Nixon, 1999:1). A gait cycle consists of the time intervals needed for successive 'heal strikes,' initial foot to floor contact for the same foot (Nixon, 1999:2-3). The idea of identifying someone by their gait is not new. In *The Tempest* (Act 4 Scene 1) by Shakespeare, Ceres states "High'st Queen of state, Great Juno comes; I know her by her gait." Also, in Shakespeare's *Troilius and Cressida* (Act 4 Scene 5), Ulysses states "Tis he, I ken the manner of his gait; He rises on the toe: that spirit of his in aspiration lifts him from the earth" (Nixon, 1999:2). This implies Shakespeare considers gait to be a distinguishing trait of an individual (Nixon, 2006:1).

 An individual's gait is also unlikely to be obscured and is hard to conceal. Most biometric techniques require a close sensing or physical contact, as in fingerprinting or retina scanning, whereas the gait feature can be perceived at a distance (Wang, 2003:1120). Gait recognition is non-contact and unobtrusive and therefore avoids most privacy issues involved in the use of biometrics (Lie, 2005:767). However, there are arguments that an individuals gait can be concealed. For different physical conditions such as injuries to joints, drunkenness, and pregnancy, an individual's motion may be altered significantly (Hayfron-Acquah, 2002:632).

Description

        The earliest studies of gait detection were done with lights attached to designated joints on a body; these experiments are called Moving Light Displays (MLDs) and they are still being used by some researches. From a single static image of a MLD an individual could not tell what object was in the picture, but with a sequence of MLDs an individual could identify a walking person. One study found an individual could tell if the person pictured was male or female, and if the person pictured was a friend, the individual could tell who was in the picture (Lie, 2005:767). Kale has found that people have the ability to recognize an individual from an impoverished display of gait (Kale, 2003: 1). Studies also showed that different types of motion, including jumping and dancing, could be discriminated (Nixon, 1999:4).

        Gait recognition has also been used in the medical field. It has been suggested that individual gait patterns mature by three years of age (Tingley, 2002:150). The main reason for gait studies in the medical field is the identification of pathological abnormalities in patients. There have been studies of a child's gait to detect birth defects or other abnormalities. Gait studies have been used to classify the components of gait for the treatment of these patients (Nixon, 1999:2).

        Gait recognition research can be broadly classified into two classes, model-free and model-based. The first, model-free, is often referred to as appearance-based (Boyd, 2001). This approach is mostly used on silhouette features, oscillations, or shape-of-motion to accumulate information on the gait (Lie, 2005:768; Boyd, 2001). In the silhouette features, the silhouette is encoded into a form and the dimensionality is

reduced by the use of principal components analysis (Lie, 2005:768). Oscillations in image intensities are performed by taking information on the motion and transforming the data into a motion energy image. By comparing the motion history of the motion energy image, gait recognition can be performed (Boyd, 2001). Alternatively, shape-of-motion methods analyze the shape of people as they walk (Boyd, 2001). These are just a few of the model-free methods used in gait recognition.

In the model-based approach, gait recognition is achieved by analysis of the structural shape of a person. The use of volumetric, ribbon, blob, and stick figure models are some of the common models. Volumetric and stick figure models are more commonly used than ribbon and blob models (Nixon, 1999:5). Volumetric models use a collection of sphere shapes for the representation of a "tree structured" skeleton (Nixon, 1999:5). The ribbon model is a two dimensional version of the volumetric model. In the blob model, a person is modeled as a set of circle blobs (Nixon, 1999:6). Each blob represents one class and has a given spatial color (Nixon, 1999:6). A map which indicates the pixels that are members of the different blob classes is also generated. In the stick figure model joints are connected by sticks to represent the shape of a person (Nixon, 1999:5). These models can represent the joints and limbs of people. By using a sequence of images, the joint angle information can be extracted and used for gait recognition (Nixon, 1999:6). This study prepares video data so that any of the above methods could be applied.

Relevant Research

Many researchers have studied the idea of an individual's gait being unique. In the research by Boyd and Little, both model-free and model-based approaches are examined using three requirements in gait recognition: frequency entrainment, phase locking, and physical plausibility (Boyd, 2001). Frequency entrainment is accomplished when the various components of gait share a common frequency. The phase of the components of the gait need to be relatively constant, and the locking component varies for different types of movement. Physical plausibility means the gait must have a stable solution to an equation of motion (Boyd, 2001). Boyd and Little believe oscillations are the center of gait analysis. Thus frequency entrainment and phase locking are important in both model-free and model-based methods (Boyd, 2001). To finish their study, Boyd and Little discuss the advantages and disadvantages of using oscillators in model-free and model-based approaches to gait recognition.

Two articles by Nixon and others present a model-based approach. These studies use computer vision techniques to find the individual in the picture and to derive motion characteristics to form a sequence of images that yield the gait signature (Nixon, 1999:1). They use an Eigenspace Transformation (EST) based on Principal Component Analysis (PCA) for a set of training images. This transformation rotates the original data coordinates along the maximum variance direction, and the new eigenvectors are used as a basis to span a new vector space. After the transformation, the original image is approximated by a linear combination of the new eigenvectors. A canonical space transformation (CST) reduces the data dimensionality and maximizes the class separation

of the new data. The trajectories in the eigenspace overlap then and the centroids are close together making it easy to accomplish recognition of the individual (Nixon, 1999, 12). They classify this approach Recognition by Statistical Measurement. They also conducted research using another approach: Recognition by Feature-Based Measurement (Nixon; 2006: 5). In this approach it is shown gait is not characterized by flexion alone but is also controlled by musculature, which controls the way the limbs move (Nixon, 2006, 5).

Wang, Tan, Hu and Ning also propose an automatic gait recognition algorithm using statistical shape analysis. A background subtraction procedure is used in each of the image sequences to extract a silhouette of the individual from the background (Wang, 2003: 1120). Then the detection of the temporal changes of the silhouettes is made into a complex vector configuration. Next *Procrustes* shape analysis which is a method in shape statistics, is portrayed (Wang, 2003: 1123). The method does not analyze the dynamics of the gait but rather uses the walking action of the individual to capture the characteristics of individual gaits (Wang, 2003: 1120). In their study different individual gaits were recognized in the silhouettes.

They also produced an effective study of a three-dimensional human body model, which is challenging because of the large number of free parameters. In their study they reduced the parameters to 12 with the assumption of walking parallel to the image plane (Ning, 2006: 1). With the objective of predicting the posture of human body model in the next frame, this prediction is then matched to the next frame in the sequence. The matched frames are then calculated and optimized to find the minimization of the match

error (Ning, 2006: 2). They attempted to find distinguishing characteristics in the individual gait posture vectors (Ning, 2006, 3).

An approach taken by Hayfron-Acquah, Nixon and Carter was a new method of spatio-temporal symmetry using the Generalised Symmetry Operator. This approach is motivated by the psychology view that human gait is a symmetrical pattern of motion (Hayfron-Acquah, 2002: 632). By adding temporal information into the calculations for gait recognition, an individual is not only recognized by body shape but also by motion (Hayfron-Acquah, 2002: 632). They start by using a symmetry extraction on the original image. From the original image the silhouette is extracted, the edges are found, and then the spatial symmetry map is detected by examining the symmetry from pairs of image points (Hayfron-Acquah, 2002: 633). Then recognition is accomplished by averaging all of the symmetry maps from the given image sequence to derive the gait signature (Hayfron-Acquah, 2002: 634). To produce the recognition by symmetry the *k*-nearest neighbour rule is applied. This approach gives a reasonable classification. An even better classifier may arise from a feature space classification or a classifier that is more sophisticated than the *k*-nearest neighbour (Hayfron-Acquah, 2002: 634). This approach has shown results which agree with earlier results that human gait has symmetrical properties and is unique to an individual (Hayfron-Acquah, 2002: 632).

Fourier series is another approach that has been used in gait recognition. In studies by Yu and others of human identification, the spatio-temporal characteristic of the moving silhouettes are analyzed. A set of key Fourier descriptors (KFDs) is found to reduce the gait data dimensionality and lessen the cost of the computation (Yu, 2004:

10

282). The KFDs are from the discrete Fourier transform. Fourier descriptors are invariant to translation, scale, and rotation (Yu, 2004: 283). When doing the classification of the data the leave-one-out cross-validation rule and a nearest neighbour classifier are used (Yu, 2004: 283). They discovered that sixteen points are enough to represent a human silhouette for recognition (Yu, 2004: 285).

Tingley and others also used Fourier series in their study of gait in young children. They do not identify individual children but classify the child's gait cycle as being within bounds of normal or abnormal gait (Tingley, 2002: 151). They use eleven functions that involved hip angle, knee angle, and ankle angle. From these eleven functions the coefficients that describe the Fourier curves are found (Tingley, 2002: 152). The variation of the child's gait pattern from that of a normal hip angle, knee angle, and ankle angle pattern are approximated as a linear combination using PCA (Tingley, 2002: 153). Then a child can be classified as having a normal or abnormal gait.

One approach combines three others into one. Begg and Kamruzzaman study gait cycle changes by using three types of machine learning approaches of gait measures: basic temporal/spatial, kinetic, and kinematic (Begg, 2005: 401). This study compared the gait cycle of twelve young individuals and twelve elderly individuals. Again this study does not attempt to identify individuals based on their gait but rather classifies the individual into an age group. The classifications of the two groups use neural networks and fuzzy clustering techniques (Begg, 2005: 402). A machine classifier, support vector machines (SVM), helps in the classification and regression of the data (Begg, 2005: 402). The results show that SVMs can identify the differences between the young and elderly

walking gait cycles. SVMs also show the underlying data structure in the models relating to young and old (Begg, 2005: 406). This study did not distinguish different individuals but it was able to distinguish between two different groups of individuals.

Another approach with kinetic or kinematic characteristics is by Yoo, Nixon and Harris. This study proposes a new method to extract body points by linear regression and topological analysis in different areas on the body based on anatomical knowledge (Yoo, 2006: 1). With this knowledge a two-dimensional stick figure is used to represent the human body. The angles of the relative points in the gait cycle are then compared to that of medical data figures (Yoo, 2006: 2). This approach shows it is possible to recognize the body in motion and the body structure while in motion (Yoo, 2006: 2).

Fujiyoshi, Lipton and Kanade analyzed the motion of humans in video streams to make an image skeletonization. As a first step for real-time target extraction they attempt to use a background subtraction that is more adaptable to environmental changes (Fujiyoshi, 2004: 114). The types of image motion stated as significant to moving target detection are: slow dynamic changes in the environment, "once-off" independently moving false alarms, movement of environment clutter, and the moving target (Fujiyoshi, 2004: 114). Because the first step only removes the background, the next step is to process the target, which removes everything in the frame that is not a part of the human target and then produces the star skeleton formation of the image (Fujiyoshi, 2004: 114). A star skeleton formation is where the extremities of the individual are joined to a center point, the centroid, by a line, producing a star pattern. Analyzing human motion from a

video is a complex problem and would have to be computationally inexpensive to detect the small amount of target data (Fujiyoshi, 2004: 119).

The width vector gait representation approach is used by Kale and others. In this approach the width of the outer contour of the silhouette is used as the feature vector, so that the physical structure of the individual and the swing of the limbs are retained in the data (Kale, 2003: 2). The gait signature is regarded as the variation of the components in the width vector for the individual (Kale, 2003: 2). Over the period of the gait cycle the width vector changes but with a high degree of correlation within the gait cycle where most changes are in the hand and leg regions (Kale, 2003: 3). It is found that changes in the individual dynamic and stride could lead to poor performance in gait recognition (Kale, 2003: 6). As in other studies, they are reasonably effective at identifying individual gaits.

Summary

Gait recognition using computational techniques has been studied for an only relatively short amount of time. The lack of a common database and evaluation methodology has previously limited the development of gait recognition research. This situation has recently improved with the introduction of large databases, such as the Large Gait Database at the University of Southampton and USF HumanID Database (Yu, 2004:285). The model-free approach is often referred to as appearance-based and is mostly used on silhouette features, oscillations, or shape-of-motion to accumulate information on the gait. By contrast, in the model-based approach, gait recognition is achieved by analysis of the structural shape of a person. The use of volumetric, ribbon,

blob, and stick figure models are some of the common models. There are still many

different views on how gait recognition should be done and what the important features

are in human gait, but researchers agree that classification of human gait will provide an

important advancement in biometrics for distinguishing people from a distance. Clearly

gait recognition research needs to be continued and the technology developed further.

III. Methodology

Chapter Overview

This chapter describes how the research was conducted. From Chapter I, the overall problem statement suggests that human gait is unique to every individual and therefore, a person can be detected by their gait. The primary goal is to identify individuals by their gait. This thesis addresses the first step in this goal, separating a moving person through a video from the background scene. This chapter first gives an explanation of how the gait data was collected. The devised method of background removal is described next. Finally, the chapter identifies limitations inherent in the devised methodology.

Data Collection

The data collection was in accordance with the AFIT guidelines for Human Experimentation Requirements of AFI 40-402. After attaining a request for exemption from Human Experimentation Requirements AFI 40-402, the data collection began. A video camera was set up in front of a static background that consisted of a background wall and a blacktop, level ground adjoining the wall. No vegetation or sky was present in the scene. The camera was placed on a tripod to minimize the movement of the camera. Participants were asked if they would volunteer to be taped in order to conduct a study on human gait. It was explained to the individual what would be done to the data and how their identity would not be revealed. When a volunteer was identified they were asked to make a series of three passes in front of the camera, each pass was a distance of

approximately five, ten, and fifteen feet away from the camera. The order of the distances was chosen at random, but each participant completed a pass at each distance. Video data was collected on two different days. The first day some of the scene was in shade and some was sun lit. The second day (in the afternoon), the entire scene was shaded by the building against which the video was taken.

Multivariate Analysis

When presented with a set of data, even a small set of data, some striking features of the data may remain hidden. Multivariate analysis reveals these features. "Multivariate analysis is the analysis of observations on several correlated random variables, for a number of individuals." (Kshirsagar, 1972: 1). Most of the theory of multivariate analysis is based on linear transformations of the original data, mainly because when using normal variables the distribution of a linear function is normal (Kshirsagar, 1972: 2). It is not always evident that a set of numbers is different from another set of numbers. Yet when the data is graphed the separation may become clear (Krzanowski, 1988: 4).
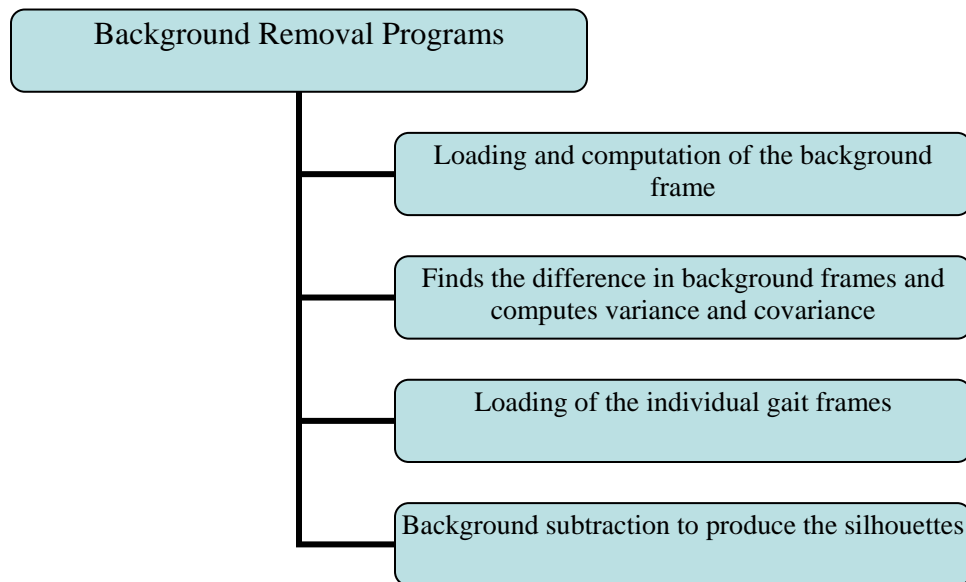
Research Step

A deficiency of gait recognition is the lack of a standard automated way to remove the background from the individual in a video sequence. Background removal has been done by either hand tracing the individual out of the background or by computer imaging techniques. Unfortunately the finished silhouettes are often of poor quality

because the methods are not very sophisticated. This thesis increases the sophistication of the background removal which will help solve the silhouette problem.

The video of an individual walking across the screen was loaded into the computer as a series of jpeg files, each of which displays a single frame of the video as a still picture. Each second of video produces thirty jpeg files. Since the video contains both the individual and background data, computer programs were developed to remove the background data from the individual data. Table 1 describes the purpose of each of these computer programs which can be found in Appendix B.

Table 1. Background Removal Program Outline

Background Removal Programs

Loading and computation of the background frame

Finds the difference in background frames and computes variance and covariance

Loading of the individual gait frames

Background subtraction to produce the silhouettes

A typical video stream has several frames of background only, followed by many frames of a person walking across the scene, followed by several more frames of background only. An example may be ten frames of only background, 130 frames of the person walking, followed by fifteen frames of background. After the video is converted

to jpeg files, the first task is to identify those frames in the beginning and at the end which contain only background. These files are identified to the first MATLAB program which averages the red, green, and blue (RGB) intensities of each individual pixel across all these frames. For instance, in the above example there are a total of twenty-five frames of only background. The RGB intensity of pixel 1,1 in each of these twenty-five frames is averaged. This is done for each of the 345,600 pixels (image size 480x720). These average background pixels are then combined into a new image that represents the average background scene. The program also displays a subplot of the RGB intensities.

The second MATLAB program subtracts the average background frame from each of the background frames and produces a scatterplot of a select number of the differences. The reduced frames are used to compute the variance and covariance in the RGB intensities which will be used in the fourth program to determine whether the pixel is solely background or person.

The third MATLAB program loads the files with the person in the frames. Again, these frames need to be specifically identified but then this program prepares the files to be used by the next program.

The fourth MATLAB program compares the RGB intensity of each pixel to the average background intensity of the same pixel. Those pixels within the covariance region are deleted and those pixels outside the covariance region are assumed to be the individual. Thus, the background of the picture has been removed from the person.

Limitations

A few limitations of this research are due to the camera used, and some are due to the computer. These limitations were known issues before the study was done, but do not prohibit the use of the method.

The video is subjected to compressions done by the camera. This interlacing in the camera presented a limitation. The camera used was an ordinary personal-use video camera, not a high definition or infrared camera. The interlacing in the video made the edges of objects in motion (e. g., a person walking across the screen) ill-defined. Thus, the image of an individual is shown with jagged edges.

The video is downloaded into the computer as a series of jpeg files snap shots. The video is already degraded by compression performed internal in the camera and is now further compressed by being converted to a series of jpeg files. Due to the compression, the created jpeg file deviates slightly from the original. Jpeg files are intended for images that are examined through human eyes so that small color changes are perceived less accurately than small changes in brightness.

Even with these known limitations the jpeg files are still used for two reasons. First, the jpeg files store the images as twenty-four bit-per-pixel color data instead of eight bit-per-pixel. Second, jpeg compression makes the image files smaller. This study deals with a large data set. If the still pictures were not compressed into jpeg files, the computer might not be able to handle the inputs.

Summary

The chosen methodology looks at the start of the overall problem statement: The human gait is unique to every individual and therefore, a person can be detected by their gait. This chapter explains how the gait data was collected. It also describes how background removal is achieved through the use of hypothesis test performed on millions of pixels as they compare to the background scene. The chapter briefly explains multivariate distributions and the known limitations of the study, through the limitations are generally with the data, and do not affect performance of the method presented.

IV. Results and Analysis

Chapter Overview

The purpose of this chapter is to demonstrate the results and analysis yielded by the research. The primary goal of gait recognition is to identify individuals by their gait. This thesis does not provide a method to distinguish different individuals but starts the process by removing the background scene from video containing a person walking. This data may be used in a variety of ways to perform continued gait analysis. This chapter discusses the data process and presents the results of applying the methodology outlined in the previous chapter on the gathered data. There is much greater detail and some of the problems encountered are resolved. Finally, the chapter answers the research question, the background is able to be removed by the proposed methodology.

Results of Research

This research first started out with the gathering of data. Two different days were used to gather data by use of a video. The first day was hot and sunny and the second day was overcast with a storm on the way. On the first day eleven different people were recorded walking a straight line along a building. On the second day five people were recorded. When the video data was loaded into the computer there was no noticeable difference in the quality of the video data from one day to the other.

As the research progressed, the different days were found to have both advantages and disadvantages. On the first day, the bright sun produced a shadow of the image, a disadvantage, while the calm weather helped keep the camera still, an advantage. The

shadow of the individual was not averaged into the background. The method for background subtraction is able to take the shadow due to the building out; it is only the shadow of the person not remove. Thus, there was noise in the picture that was not taken out. The second day was overcast and the individual did not produce a visible shadow which seemed to compensate for this problem, an advantage, but an approaching storm produced winds which could have slightly moved the camera, a disadvantage. However, when the data was put through the computer programs and the building background was removed, a small shadow by the feet of the individual is seen. Despite the overcast the individual did produce a minimal shadow.

The video of the individual was loaded into the computer as a series of jpeg files. There are thirty jpeg files per second. Video was collected with the individual walking from the left and from the right side of the frame. If the individual took five seconds to walk across in front of the video camera (about three cycles of the individual's gait) that is 150 jpeg files. Plus twenty-five frames of background (fifteen in the beginning and ten at the end). Each file is presented as a frame of size 480x720x3 (480 rows of pixels by 720 columns of pixels and each pixel has three RGB dimensions). That is there are 1,036,800 pieces of information for each frame or a total of 181,440,000 pieces of information for a single individual video. Therefore the data gets large very quickly. This chapter uses example data from persons on each day of recording.

Figure 1 is an example of the average background taken from the first day of recording. To produce this picture the first five frames and the last twenty-two frames of the data (when the person was not in the frame) were used. This picture was produced

from averaging the individual pixels from the twenty-seven frames together and passing

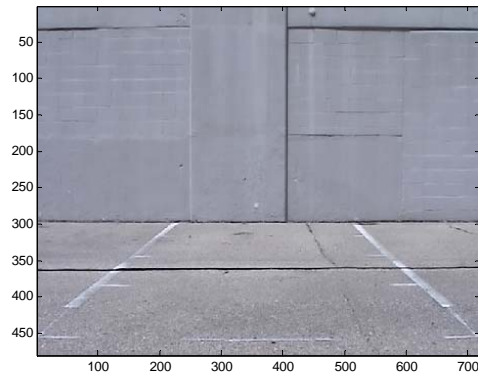the average RGB intensities to one picture. Only the first five frames at the beginning

Figure 1. Average Background

were used because after the first five frames the individual's shadow began to appear on

the ground. The first frame to include any person or shadow (in this case the shadow)

was considered a frame where the person was walking. Any jpeg file in the series with a

hint of a shadow or the person was not used in the calculations to produce the average

background. This was done to reduce the amount of noise in the background picture.

The white lines in the picture are the lines drawn for the individual to know where

the recording was actually taking place. They were instructed to start walking three feet

before the line on the left and to continue for three feet past the line on the right or in the

other direction. (Additional lines were to indicate these points.) Starting from the bottom

of the picture, the first horizontal chalk line is to mark the bottom of the video frame and

there is another line at five feet, ten feet and fifteen feet to indicate the distance the

subject is from the background wall.

Now that the average background has been found the average RGB intensity of each pixel is computed. Throughout this example, data from the first subject on the first day is used. Figure 2 is a picture of RGB intensities of the average background. The top
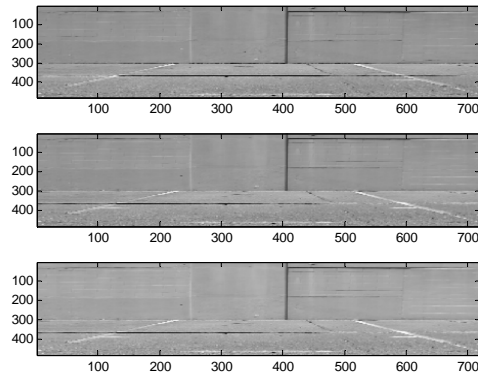

Figure 2. RGB Intensity

picture is the average red intensity. The middle picture is the average green intensity and the bottom is the average blue intensity. There does not appear to be a readily visible difference in the three intensities, implying a high degree of correlation between the intensities for this particular background. This would be reasonable since the building was primarily gray, a mixture of red, green, and blue colors.

Since the differences in the intensities is hard to distinguish the RGB variation between the average and each individual background picture is found This difference between RGB of the average background and each background frame should lead to pixels that have RGB intensity values which are approximately zero. These differenced pixel values should describe an approximately normal distribution in each of the RGB color direction. This can be seen in the histogram of the three intensities presented in Figure 3.
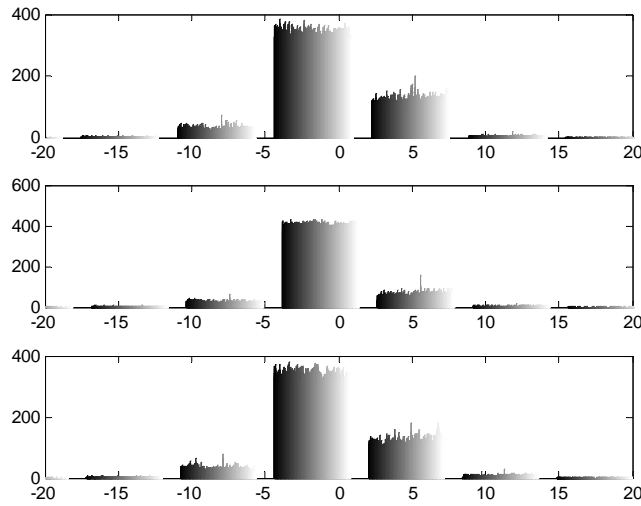
Figure 3. RGB Intensity Histograms

Considering the correlated nature of the RGB intensities, the differenced pixel values should fall into a highly correlated trivariate normal distribution, confidence contours of which should be ellipsoidal in shape. This ellipsoidal shape, which will be discussed in much greater detail later in the chapter, will be used to create a rejection region for a hypothesis test to be performed on each pixel to gauge whether individual pixels are part of the background scene or part of a moving body. To begin, only uncorrelated variances are used to determine whether the found covariance matrix will be required. The pixels determined to be not in the background are the pixels determined to be the individual walking through the screen. With this said the RGB variance in the first set of background data is 8.0781, 7.2496 and 8.5232, respectively.

The average background picture has been produced and the RGB variance of the average background has been found. Now the entire data set of the individual walking across the screen is loaded into the MATLAB program. Figure 4 shows the 117 frames

25

of the video stream that were considered to be not background frames, image frames.

The first few frames of this data set does not show the person, but the shadow has already

Figure 4. Image Frames

started protruding into the picture so they are not used in the development of the average

background. Within the MATLAB program these jpeg files are assembled into an avi

file in order to show motion in the figure.

With the image frames of the individual walking across the screen loaded into the

MATLAB programs, the background removal can begin. First, the average background

picture is subtracted, pixel by pixel, from each image frame. Now, each background

pixel should be part of the trivariate normal distribution created using just the differences

between the average background and each individual background frame, creating a

contour at a specified number of standard deviations away from the mean values of the

differenced RGB intensities, produces an ellipsoid. Any differenced pixel within the

ellipsoid is assumed to be a background pixel, and any differenced pixel not contained

within the ellipsoid is assumed to be part of the walking person. The ellipsoid equation

used is:

$$X = R^2 /(k*V(R)) + G^2 /(k*V(G)) + B^2 /(k*V(B)) \tag{1}$$

where R is the red intensity for a differenced pixel, G is the green intensity for a

differenced pixel, and B is the blue intensity for a differenced pixel, V(R) is the variance

found for the red pixel, V(G) is the variance found for the green pixel, and V(B) is the

variance found for the blue pixel, and k is a constant based on the variances of the

distributions. If X is less than one, fail to reject the null hypothesis of the test that the

pixel is part of the background. The pixel is assigned a value of zero. If X is greater than

or equal to one, reject the null hypothesis of the test in favor of the alternative hypothesis

that the pixel is not in the background and must therefore be part of the person walking

across the frame. The pixel is assigned a value of one.

After testing every pixel in every frame against equation (1) and assigning a one

or zero to each pixel in the picture, a silhouette of the image was made by coloring the

zero pixels black and the one pixels white. Thus a series of images was created with a

black background and white silhouette image. As in the image frames of the individual

walking across the screen, this series of silhouette frames is put into an avi file in the

MATLAB program in order to show motion in the figure.

Figure 5 shows the series of resulting silhouette frames used to produce the avi file of the moving silhouettes. Each frame also shows the image of the individual with the background removed. When the images are shown in this small of a subplot, random noise in the pictures is difficult to discern. When looked at in a larger size, however,
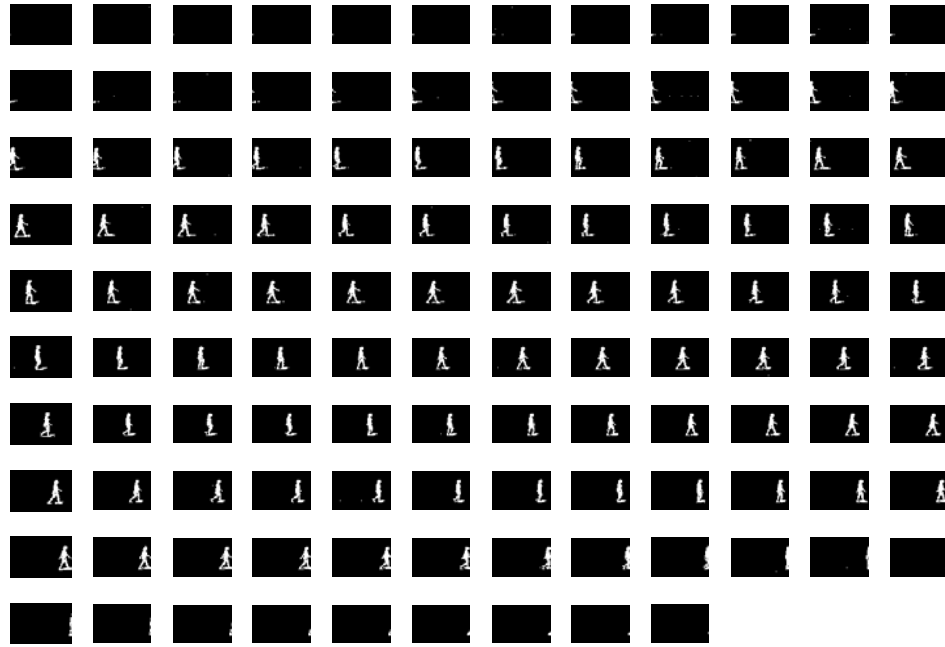

Figure 5. Silhouettes

there is random noise in both the silhouette of the individual and in the background. The removal of this noise is the next problem that needed to be undertaken. Larger pictures of that reworked data are included later in this chapter and the random noise in both the background and silhouette is more visible in them.

In order to investigate the removal of the random noise, the background removal process was reexamined. One frame of the data is 480 rows by 720 columns; both the background frames and the silhouette frames are the same size. In other words, there are 345,600 pixels per frame. Since they are both presented in black and white (zero or one), there is no RGB dimension. In this example there are 117 silhouette frames and twenty-seven background frames, each with 345,600 pixels for a total of only 60,480,000 pieces of information which greatly reduces our data set.

Figure 6 shows the average change in one background frame (frame three) minus the average background frame. The figure does not show all the points, however a small
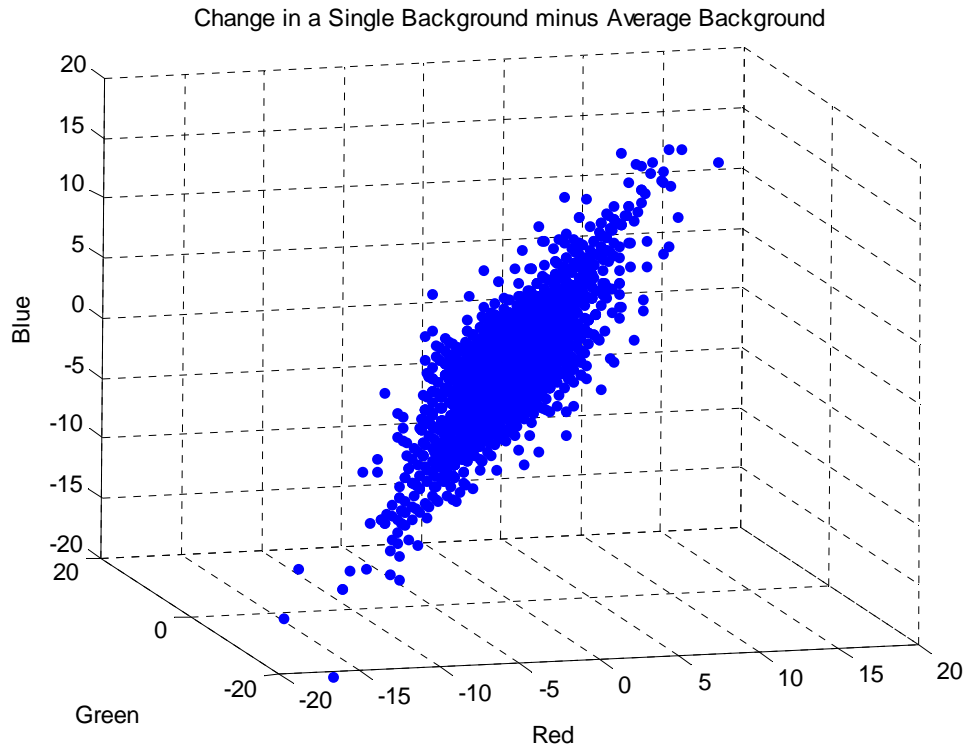

Figure 6. Data Points

random sample of the points, 5,000 of the 345,600 pixels, is graphed in the figure; one and forty-four hundredths percent of the data. By viewing the graph one can see there is a correlation in the RGB intensity.

Figure 7 shows the region equation (1) used to determine how the pixels are classified. The points are still our points from Figure 6 and the ellipsoidal surface is the
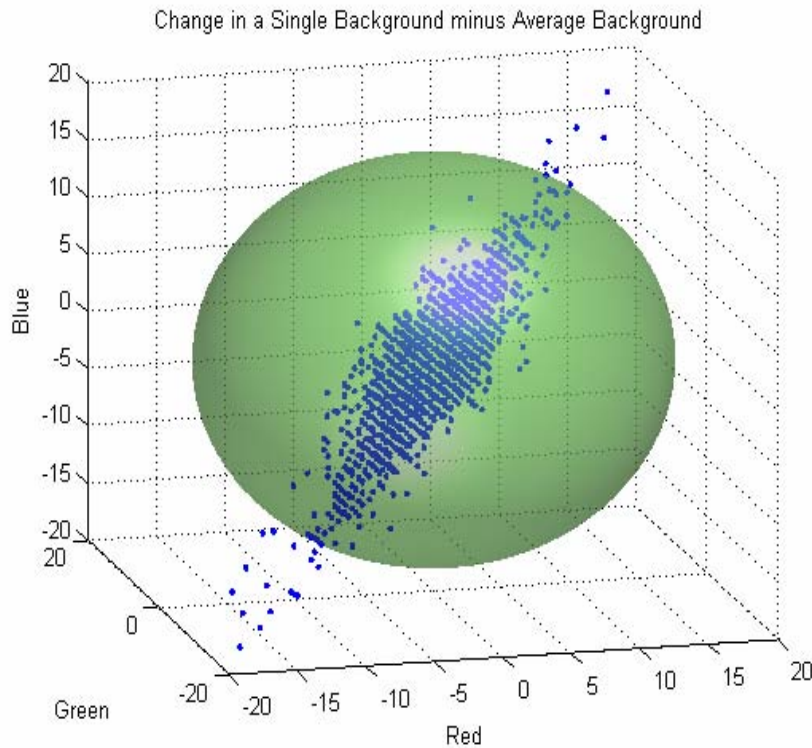

Figure 7. Elliptical Region

shape produced by using equation (1) to set a threshold on the data. The ellipsoid represents a six standard deviation (from the mean) contour. One can see there are points outside the ellipsoid that should be identified as a part of the background. Similarly, there are points inside the ellipsoid that will be considered as part of the background but should be image. Equation (1) introduces unwanted noise into both the background

frames and the silhouette frames. Setting the number of standard deviations away form the mean at six contributes to the misidentification.

Simply changing the standard deviation doesn't change the shape of the ellipsoid. If the rejection region is formed into a longer and narrower ellipsoidal shape, the rejection region can better fit the data presented. Since there is an obvious correlation between the RGB intensities in the background our next step is to use this correlation data to describe the ellipse. This correlation data is used to build the covariance matrix for the example data:

$$C = \begin{bmatrix} 8.8800 & 6.8880 & 7.4331 \\ 6.8880 & 8.0289 & 7.2211 \\ 7.4331 & 7.2211 & 9.2185 \end{bmatrix}$$

Each of the 345,600 pixels from image frame three was put into a 345,600x3 matrix consisting of red, green, blue pixel intensity columns. This matrix was multiplied by the covariance matrix according to equation (2) to account for the correlation between the various pixels. Each row in the matrix is a single pixel in the frame; therefore there are 345,600 rows in the matrix.

$$X = \begin{bmatrix} R & G & B \end{bmatrix}^T * \begin{bmatrix} 8.8800 & 6.8880 & 7.4331 \\ 6.8880 & 8.0289 & 7.2211 \\ 7.4331 & 7.2211 & 9.2185 \end{bmatrix}^{-1} * \begin{bmatrix} R & G & B \end{bmatrix} \qquad (2)$$

Just as in Figure 6, 5000 of the pixels were chosen and are plotted in Figure 8. The blue, more dispersed, points are the original pixels without being multiplied by the covariance matrix. The red, more compacted, elongated, points are the pixels after they have been operated on by the covariance matrix. The red set of 5,000 points is generated

31

by this equation (2) and not by equation (1). They are simply distinct points. Therefore,

a rejection region to determine the range of pixels that are to be kept or thrown out is not

determined. While the shape looks like the elongated ellipsoidal rejection region desired,

it is merely a collection of points. A rejection region for these compacted points is then
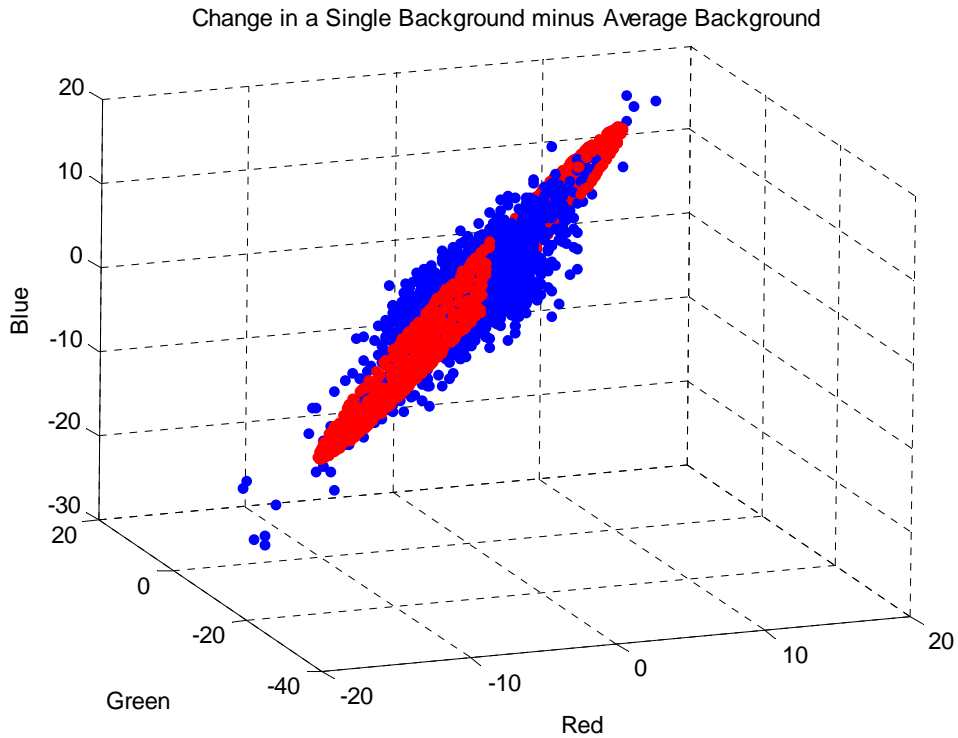
sought.

Figure 8. Manipulation of Correlation Matrix

To see the difference between the rejection region determined by equation (1) and

the set of corrected pixel points generated by equation (2), the two are superimposed on

the original pixel points in Figure 9. The red group of data points is much closer to the

background points. An equation using the red points could give a better equation for the

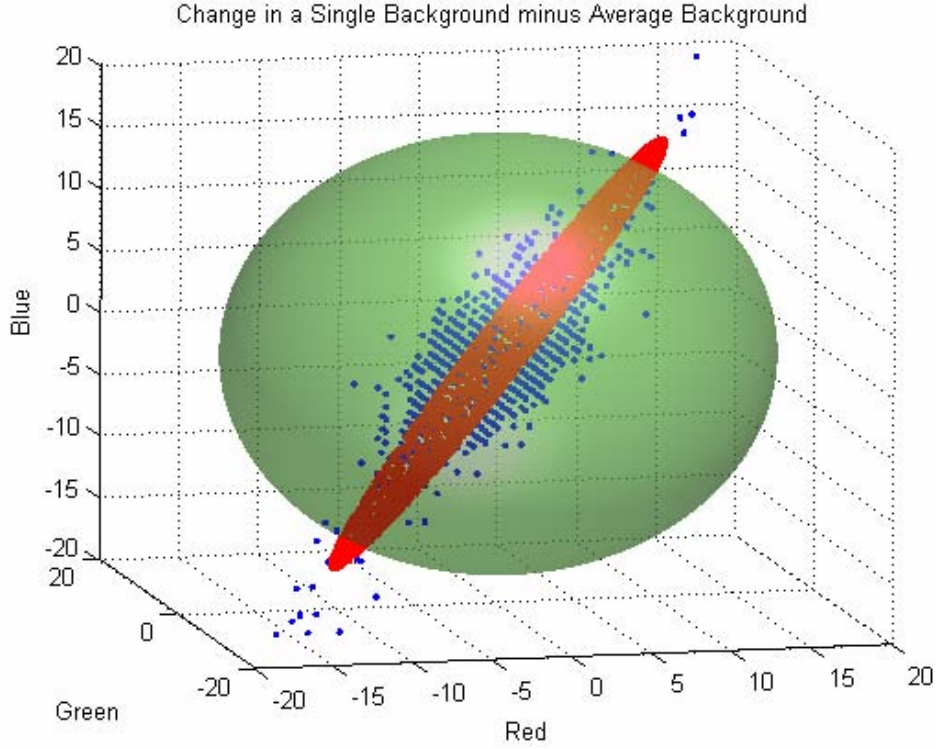region to determine the range of pixels that are kept or thrown out.

Figure 9. Difference of Regions

Another change may be made to the covariance matrix, C, in order to better visualize the process. The covariance matrix is symmetric, and therefore the Cholesky factorization of the matrix can be performed. This factorization uses only the diagonal and upper triangle of C and produces an upper triangular matrix A such that $A^T{}_*A = C$. Then instead of multiplying the data by the inverse of C, multiply the data by A. Since A is only an upper triangular it has less data (the bottom triangle is all zeros) further reducing computational load required. Equation (2) is thusly modified:

$$X = [A] * [R \quad G \quad B] \tag{3}$$

Equation (3) describes the pink region in Figure 10. The blue points are the same

original pixels used previously. This ellipsoidal, almost spherical, region used to assign

the pixel as background or image is capturing nearly all of the data. Only one point

among the 5,000 random sample points in Figure 10 appears to lie outside of the pink
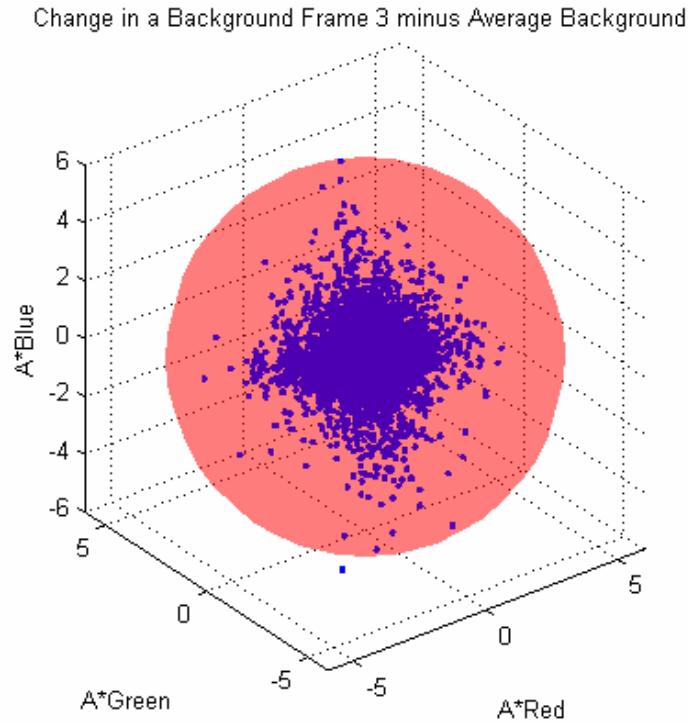
decision region.

Figure 10. Background Points

Now that the rejection region is better defined, the noise introduced into the

background and image frames by equation (1) can be reduced. Figure 11 shows the

background which was in frame 3 of the video. Note the presence of some small white

dots. These dots are assumed to be noise created in the video processing of the

background.

34

Figure 11. Unfiltered Background Frame 3

In Figure 12 using the MATLAB filter, explained below, these white spots have been eliminated. Since the background is going to be represented by black pixels and the figure by white pixels, it is necessary to remove as many of the white pixels as is possible from the average backgrounds. By comparing Figures 11 and 12, the effectiveness of the filter is apparent.
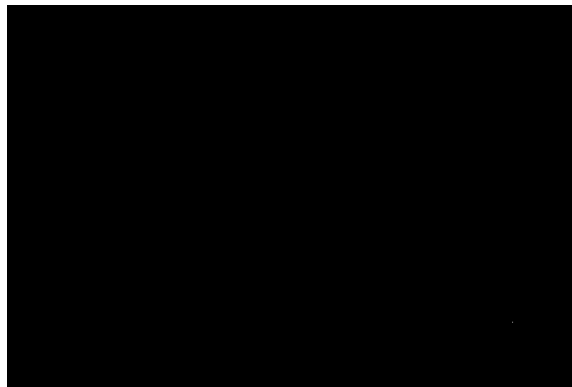

Figure 12. Filtered Background Frame 3

MATLAB has many different built in filters to help with problems like this. The filter determined to work best in this application is the medfilt2 filter. The medfilt2 filter performs a median filtering of the data in two dimensions. This works by assigning each

output pixel the median value in the given neighborhood around the corresponding pixel in the input image. The default of the medfilt2 filter performs a median filtering on the data in a three-by-three neighborhood with the pixel in question being the center. This means the filter will look at a pixel and one and a half values from this pixel in a two dimensional direction. It will then take the average value and assign it to the pixel. This gives the picture a smoothing effect to clean it up. The default filter was effective, but even more effective was using a four-by-four neighborhood.

Now that equation (3) defines well the region of the background pixels, it can be applied to the image frames, the frames that include the shadow or the image of the person walking. Figure 13 shows the random points with the background and the image
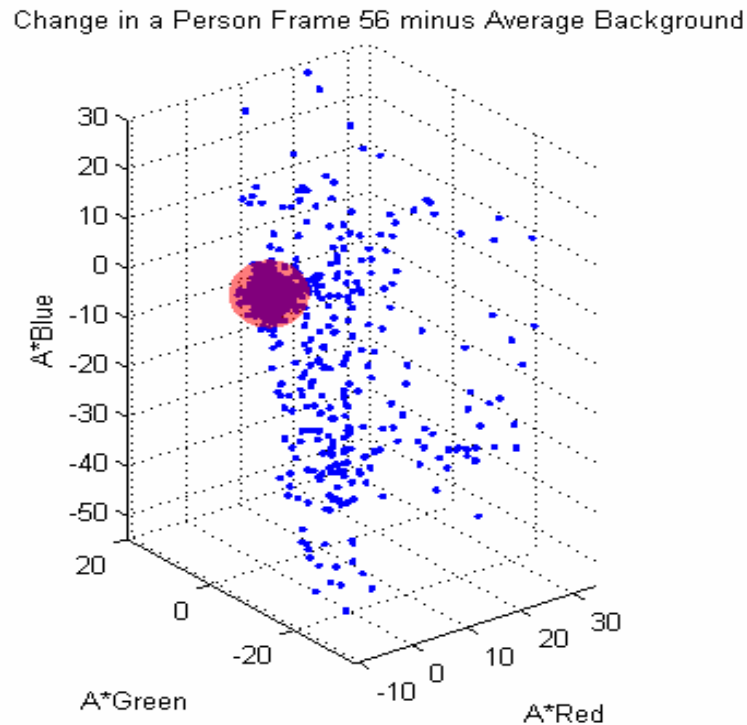

Figure 13. Points on Frame 56

36

of the person walking. The rejection region determining background pixels is also highlighted. The points in the background are grouped closer together than the points outside of the region. The points outside of the rejection region are the points determined to be associated with the image of the person. This will show how many of these points have been assigned wrong. The goal is to achieve no error in the background or the person. Thus the background should be solid black and the image of the silhouette solid white.

Figure 14 shows an actual frame of the image of the walking person's silhouette. As with the previous attempt, the number of standard deviations from the mean to include in the rejection region may be adjusted. When five standard derivations were used there

Figure 14. Unfiltered Image Frame 56

was too much noise. The level of noise was based on what was felt to be an acceptable level of noise. At six standard derivations the level of noise was much lower. An iterative process discovered that at 5.2 standard deviations much of the noise disappeared and the image did not improve appreciably as the number of standard deviations was increased to six. An acceptable level of noise was found using 5.2 standard deviations

from the mean as the threshold in determining whether a pixel is a background pixel or not. This is the same frame fifty-six that is shown as a set of points in Figure 13. Viewing the actual image of the silhouette instead of just the points, it can be seen that there is noise in both the background and in the silhouette.

The next step was to further to clean up the noise by using the MATLAB medfilt2 filter. Figure 15 presents the same frame as depicted in Figure 13 and figure 14 as a filtered image. Almost all of the noise in the background and in the image of the silhouette is eliminated. The number of noise pixels could be further reduced if the standard deviation was raised or a stronger filter was used. For this thesis the noise is deemed to be at an acceptable level.

Figure 15. Filtered Image Frame 56

From the two figures with the random points (Figures 10 and 13) and the unfiltered and filtered pictures of the person walking (Figures 14 and 15) it is apparent there is still error resulting from the procedure. To ascertain the magnitude of the error, the sum of the ones in a typical background frame was considered. Ideally, there should not be any ones because a one is assigned when the pixel is not a background pixel.

Therefore any ones would be an error in a frame with nothing but background pixels. To determine how much data is not removed from each frame, all the ones for each frame, background and silhouette frames, were summed.

This graph in Figure 16 shows the percentage of ones per frame for both the raw data and the filtered data. The first few and last few frames are background frames which explains near constant nature of the error. The slight slow rise in the beginning of the graph is explained by the shadow of the individual entering the frame. Where the graph is almost vertical is when the image of the individual appears on the frames. The thought-provoking aspect of the graph is in the middle of the graph where it begins to dips up and down.
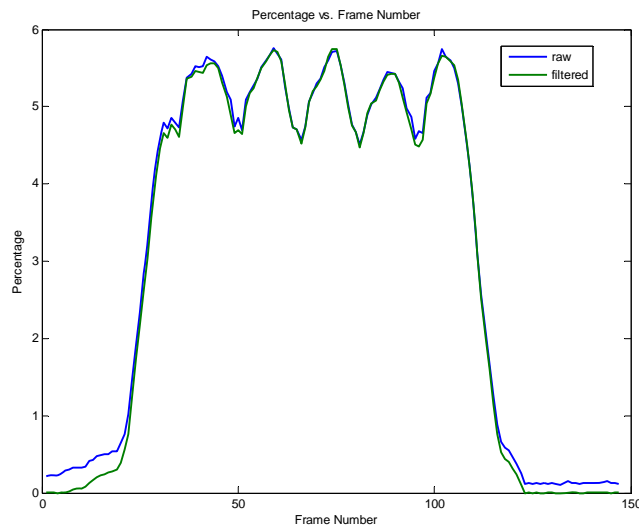

Figure 16. Percentage vs. Frame Number

It is surmised that the graph dips when an individual's gait is most compact i.e. the legs and arms are lined up with the body. The peaks are where the gait is spread as far apart as possible, i.e. the individual's legs and arms are wide spread. An example of

39

the gait when it is spread has already been displayed in Figures 13-15. Now the same figures in a frame where the individual's gait is more compact is analyzed.

This will be accomplished by looking at frame sixty-six of the same individual. In this frame the individual's gait is at its most compact. Since this frame is ten frames past frame fifty-six, this frame is only one third of a second after that frame. The actual picture of the image of the silhouette for this frame should be produced with the same acceptable level of noise as frame fifty-six. Again, the amount of noise will demonstrate how many of the points have been assigned incorrectly.

The interest in this image of the silhouette is that the gait is at its most compact. It is hypothesized that this is causing the dips in the graph of Figure 16. As before, Figure 17 uses a standard deviation rejection region determined at 5.2 standard deviations from the mean. Viewing the actual image of the silhouette can tell there is noise and where the noise is located. There is noise in both the background and the silhouette, just like in the prior frame.
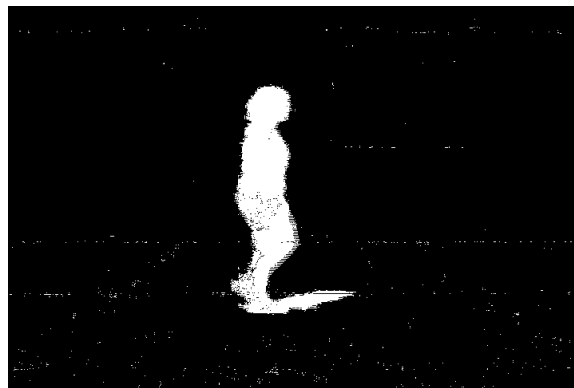

Figure 17. Unfiltered Image Frame 66

As before, the medfilt2 filter in MATLAB was employed to clean up the noise in the frame. Figure 18 presents the same frame as in Figure 17 but this time the MATLAB medfilt2 filtered frame is employed. Almost all of the noise in the background and in the image of the silhouette is gone. The noise is deemed to be at an acceptable level, there are only a couple of dark spaces in the white silhouette and almost no white spots in the black background.
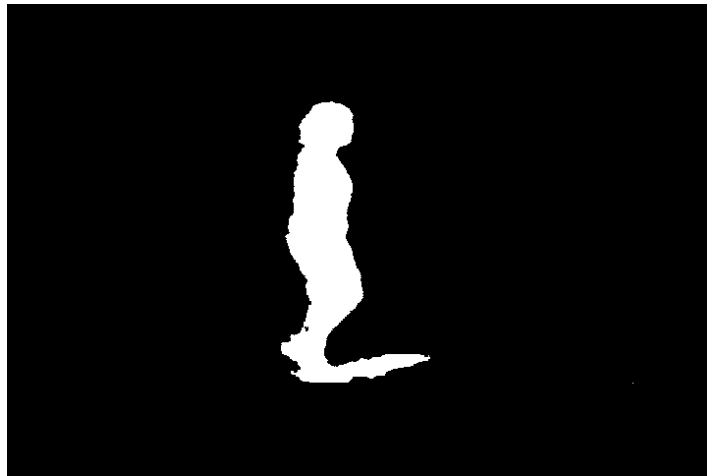

Figure 18. Filtered Image Frame 66

Figure 19 shows the random points with the background and the image of the person walking for frame sixty-six, the same frame shown as silhouettes in Figure 17 and Figure 18. The region determining the rejection region for background pixels is highlighted. Again, the points in the background are closer together than the points outside of the region. Upon a close examination the points outside of the background can be seen to be grouped closer together than before, in Figure 13.
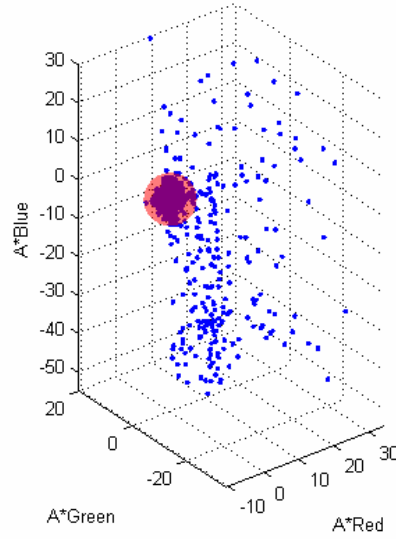
Figure 19. Points on Frame 66

Table 2 compares the percentage of ones in the frame for four selected frames of

the raw (unfiltered) data and the filtered data. This non-background data is either the

image of the individual or it is random noise. Frames fifty-six and sixty-six are the same

Table 2. Non-Background Data

Frame Number	Raw Data %	Filtered Data %
1	0.22	0.00
56	5.52	5.50
66	4.58	4.52
149	0.15	0.00

frames as Figure 14, Figure 15, Figure 17, and Figure 18, respectively. Since frame one

and frame 149 are background only frames, the percentage of non-background data for

those two frames should be zero and are shown to see how well the background scene is identified. It is beneficial to use the MATLAB filter because it appears to filter out the noise data and leave the individual.

Now that the method has been developed using the video stream from one person, the same method was tested on the video stream of another individual. This time it was used on data gathered the second day. Only the rejection region is based on the covariance matrix of the background data is utilized.

In Figure 20 the background pixels for the second set of data after transformation via Cholseky factorization are shown. Again, the ellipsoid shape region used to assign
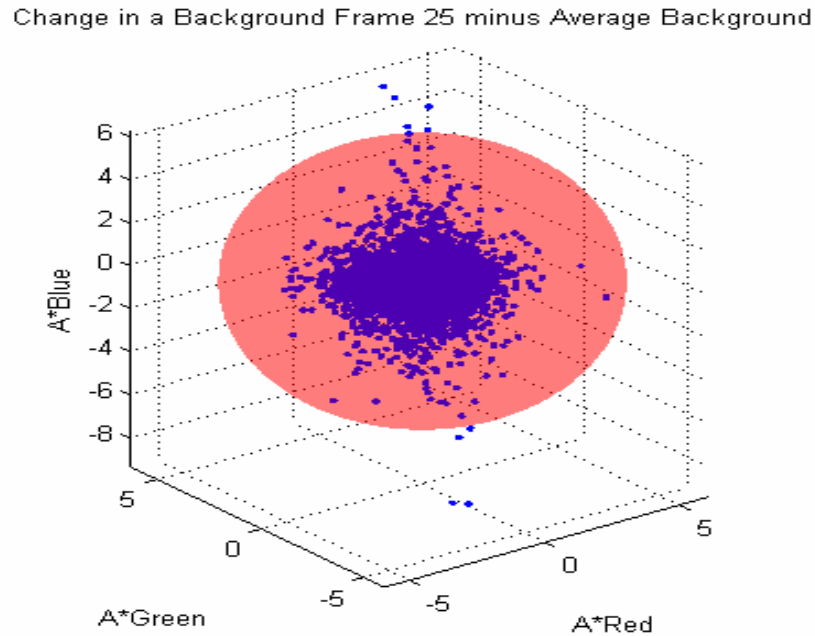

Figure 20. Background Points for Individual 2

the pixel as background or not is capturing almost all of the data. There are more points outside of this region than in the other example. This could be because the winds from

the approaching storm could have jarred the camera enough to make the background

move slightly in the frame. The resulting vibration in the camera would affect the

amount of noise in the pictures.

The unfiltered background which was in frame twenty-five of the video is shown

in Figure 21. Note the presence of some small white dots. These dots are most likely to

Figure 21. Unfiltered Background Frame 25

have been produced from noise created in the video processing of the background.

In Figure 22 using the MATLAB filter, these white spots have been eliminated.

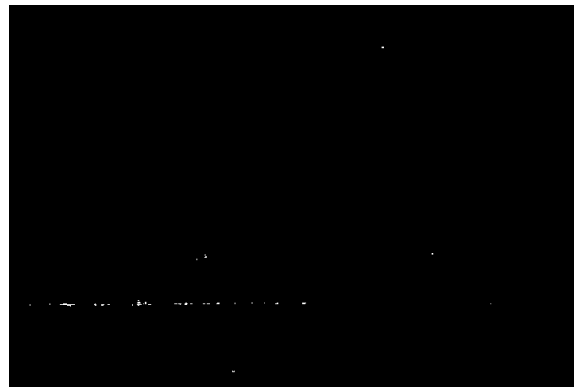(This is the same process that was used with the other example and shown in Figure 11

Figure 22. Filtered Background Frame 25

and Figure 12.) Since the background is going to be represented by black pixels and the

figure by white pixels, it is necessary to remove as many of the white pixels as is possible

from the average backgrounds. By comparing Figures 11 and 12 and Figures 21 and 22,

the effectiveness of the filter is apparent.

Using the ellipsoidal rejection region that accounts for correlation between red,

green, and blue, the pixels in a frame which includes the image of the person walking

were tested. This time a frame where the gait is more compact (frame 108) and a frame

with a wide spread in the gait (frame 114) are both presented to determine any dips in the

graph of the non-background data.

Figure 23 shows the random points with the background and the image of the

person walking for frame 108. In the picture of this frame the persons gait is compact.

The region determining the background is still highlighted by the sphere. The points in

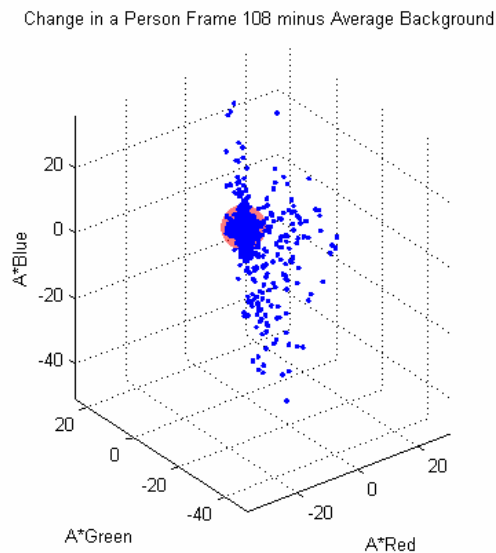the background are closer together than the points outside of the region.

Figure 23. Points on Frame 108

Here, in Figure 24, the image of the silhouette is shown as its gait is coming together. As before in the first example, Figure 24 uses a rejection region of 5.2 standard deviations from the mean. Viewing the unfiltered image of the silhouette there is a
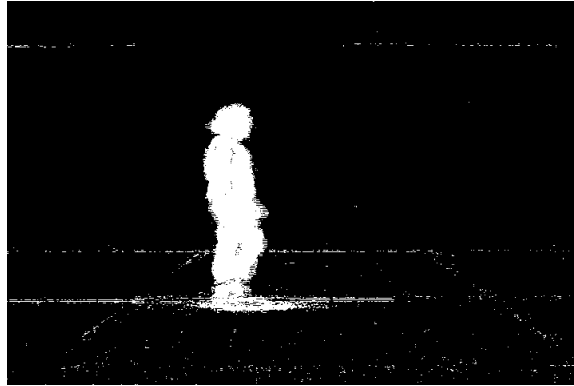

Figure 24. Unfiltered Image Frame 108

noticeable amount of noise. Noise is present in both the background and the silhouette. Clearly, there is more noise in this example than in the previous example, the video taken on the first day. Again, this could be due to the winds from the storm; perhaps the camera had been jarred vertically, producing the horizontal line of noise. A higher standard deviation might fix some of the noise. As before, the medfilt2 filter in MATLAB can be used to clean up the noise in the frame.

Figure 25 is the MATLAB medfilt2 filtered picture for frame 108. This is the same frame as Figure 24. There is still some noise in the background but almost all of the noise in the image of the silhouette is gone. As discussed above, if the rejection region were wider, more of the noise could be cleaned up, or a stronger medfilt2 filter could clear up some of the noise.

46

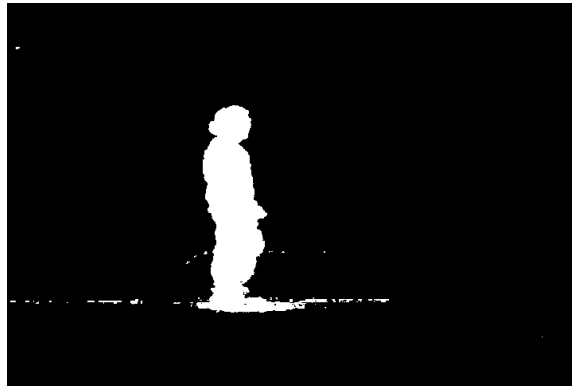**Figure 25. Filtered Image Frame 108**

Figure 26 shows the random points with the background and the image of the person walking for frame 114. In the picture of this frame the person's gait is widely spread. The region determining the background is still highlighted. The points in the background are closer together than the points outside of the region as expected.
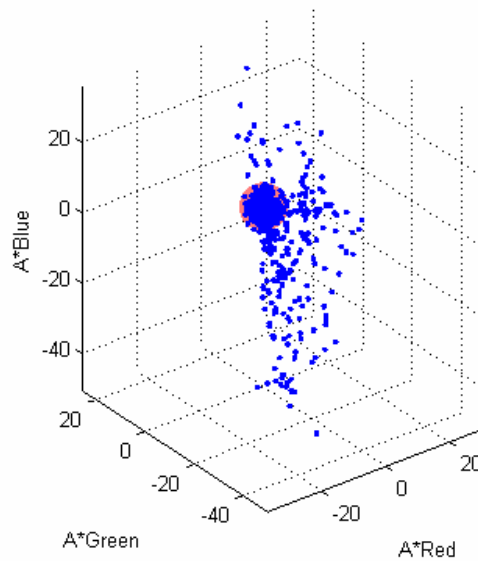

Figure 26. Points on Frame 114

Here is the image of the silhouette with its gait spread. Figure 27 uses the same rejection region as before. This is the same frame 114 that is used above in Figure 26.

Figure 27. Unfiltered Image Frame 114

There is still more noise in this example than in the other example. Again, this could be due to the winds from the storm. Using a higher number of standard deviations to create the rejection region might fix some of the noise. The medfilt2 filter in MATLAB can again be utilized to reduce more of the noise in the frame.

Figure 28 is the MATLAB medfilt2 filtered picture for frame 114. This is the same frame as Figure 26 and Figure 27. There is still more noise in this example than in

Figure 28. Filtered Image Frame 114

the previous example. There is still noise in the background but in the image of the silhouette almost all of the noise is gone. There are just a few spots left. Again, these spots could be taken care of if the standard deviation was raised or a stronger filter was used.

Figure 29 shows the graph of the percentage of ones per frame for both the raw data and the filtered data. The basic shape of the graph is still the same. (The two graphs can be viewed on the same page in Appendix A.) The first few and last few frames are background frames. The vibration of the camera can explain the jaggedness at the left and right of the graph instead of the smooth curve in the Figure 16. There is no slow rise in the beginning of the graph because it was an overcast day and the person did not produce a shadow entering the frame. So the graph is almost vertical when the image of the individual appears on the frames. The dips, representing the motion of the subject's extremities, are still in the middle part of the graph. The basic look of the graph is much
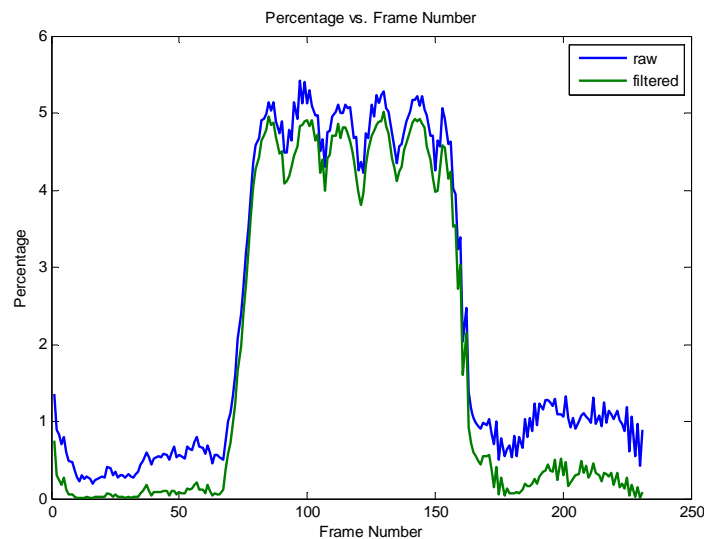

Figure 29. Percentage vs. Frame Number for Individual 2

49

the same as in the other example; however, there does appear to be some differences. These differences, if they exist, could be considered to differentiate one individual's gait from another's.

Table 3 displays the percentage of non-background data for four selected frames: frame one, frame 108, frame 114, and frame 231, the last frame. This non-background data is either the individual or the random noise. Frames 108 and 114 are the same frames as Figure 24, Figure 25, Figure 27, and Figure 28. The percentage for frame one and 231 are shown to see how well the background scene is identified. The background frames contain a higher percentage of ones in this video than in that taken on the first day. It is speculated that is at least partially due to the move severe weather conditions. Again, it is beneficial to use the MATLAB filter because it appears to filter out the noise data and leave the individual.

Table 3. Non-Background Data

Frame Number	Raw Data %	Filtered Data %
1	1.36	0.75
108	4.75	4.41
114	5.00	4.81
231	0.90	0.09

Research Questions Answered

From chapter one, the overall research question for this study is: If the human gait is unique to every individual, can a person be identified by their gait? This thesis considered the research question:

Can the background scene of a video be effectively removed from the movement of the individual in the video?

This question was answered in the process of the research in this thesis. The background was successfully removed from the movement of the individual. This was demonstrated by the ability to create a video of a white silhouette of the image of the individual walking on a black background. The background can be successfully removed just by using the developed MATLAB code, although, an image of the silhouette with less noise can be achieved by using the code together with the MATLAB medfilt2 filter.

Summary

This chapter explains the analysis and results yielded by the research. This thesis does not differentiate different individuals by their gait but begins the process through background removal and the notion of an individual's gait being unique. The chapter gives the detailed results of the research. A detailed explanation of how the background removal is achieved and what was discovered when the background was removed is offered. The automated background removal process was modeled with the use of the video stream of one individual from the first day of recording. Several figures of the actual image and background along with associated graphs of the background rejection region and associated noise reduction techniques are presented. After the model was

developed and refined, it was tested on the video stream of another individual from the second day of recording. The model that was developed using the first individual did an acceptable job of removing the background and associated noise on the second individual even though the recording conditions of the second day were considerably more than severe than on the first day of recording. The chapter ends with the answer to the research question posed: the automated methodology was able to acceptably remove the background from the image of a walking individual.

V. Conclusions and Recommendations

Chapter Overview

This chapter provides the conclusions and recommendations produced by the research. Again, the principal goal is to identify individuals by their gait, and this thesis begins the process of reaching this goal. This chapter first gives conclusions of the research and then the significance of the research. From there a recommendation for future research is offered.

Conclusions of Research

The thesis has an overarching goal of identifying individuals by their gait. This is an ambitious goal which is actually only begun in this thesis. The specific goal for the thesis was to achieve the first step of the process, removing the background from a video containing data of an individual walking. Using MATLAB as a computational tool, in a stream of video a moving individual is separated from the background. A successful sequence of frames with a white silhouette of the image on a black background is produced through implementation of our method. This thesis demonstrates the removal of the background behind a walking individual by means of an automated process is possible. This allows a researcher to concentrate on just the silhouette of the individual without the background noise.

After the background is removed, researchers will be able to study whether it is feasible to identify an individual based on their gait. The background removal is

achieved in a nearly automatic way, making it an easy transition from a video of one individual's gait to a video of another individual's gait.

Significance of Research

This thesis makes a step in the direction of being able to identify an individual based on their gait. The thesis also presents a methodology which at least begins automating the process. Once the background only files are identified, the computer and MATLAB programs do all the calculations to separate the background from the individual. The identification of the individual in the picture as distinct from the background is a fully objective and automated process.

There are many different views on what the important features are in the human gait. Researchers agree on one thing: the human gait is an important advancement in biometric for recognizing people from a distance. It is generally agreed gait recognition research needs to be continued and the technology developed further. Since this methodology only distinguishes between background and individual it has application to both model-based and model-free gait recognition research. Either class of research will definitely benefit from distinguishing the individual from the background. A step in the direction of an automatic background removal is clearly highly desirable.

Additionally, in today's highly charged political climate homeland security is of paramount importance. An unobtrusive biometric identification technique is highly desirable. Entry portals could be equipped cameras and computers programmed to record and compare gaits of individuals. While more work needs to be done the possibilities do seem encouraging.

Recommendations for Future Research

While this research does present a significant advancement in the methodology of background removal, other steps in gait recognition still need to be completed before being a viable option for identifying individuals in any circumstance.

In this research the beginning and ending files containing only background had to be identified prior to running the computer programs. Additionally each computer program had to be independently initiated. Both of these step need to be fully automated to provide a truly automated background removal process. The second of these steps would appear to be fairly straightforward by writing a master computer program which would automatically start a sequence of the separate computer routines.

Another question concerning the removal of the background is the composition of the background itself. In this study a fairly solid colored building wall was utilized, creating a good contrast with the walking figure. Additional work could consider how the process would work with a multicolored background or a changing background or even a moving platform. For instance, if the program is utilized in a Homeland Security setting to provide the identification of known or suspected terrorists, the background could consist of other walking individuals. The removal of such a background is certainly a different question.

From the sequence of frames with a white silhouette of the image on a black background, can a skeletonization based on anatomical placement of the human body be produced? In this step, scaling of the silhouette of the image could be used to move the image to the same line and the same scale in every frame of the sequence. This way the

images are examined in the same place on the frame and read on the same scale, so metrics like height could be determined. Once the image is scaled, the development of a skeletonization of the silhouette can be completed. The silhouette of the image is developed into the skeletonization based on anatomical placement of the human body. Then, as was done with the silhouette, the skeletalized image can be placed in motion. Upon adding motion, statistical measures could be developed to compare gaits.

The data for this study involved all the individuals walking a straight line horizontal to the camera wearing gym clothing, shorts and tee shirts. There are many simple variations of this scenario that could be changed to generate a wide range of significant research on gait recognition. In addition to removing just the background from the individual, it would be beneficial to also remove the shadow produced by the individual. Additional questions could arise if the individual were made to walk in a different pattern, for instance towards the camera, away from the camera, or diagonal to the camera. It is also possible that the speed of the individual could affect the recognition of the gait. An interesting question is how pants, skirts or other clothing could conceal the gait. Obviously, there is still research that can be done.

Summary

This chapter provides research conclusions and recommendations. The primary goal is to identify individuals by their gait. In this thesis the first step in achieving this goal was successfully addressed. The research demonstrates the removal of the background behind a walking individual by means of an automated process is possible. The identification of the individual in the picture as distinct from the background is a

fully objective and automated process. After background removal, it does seem to be

plausible to identify an individual based on their gait. This methodology yields an

important step towards an automated gait recognition and identification program. After

the background is removed, researchers will be able to study whether it is feasible to

identify an individual based on their gait. There is still a considerable amount of research

that needs to be conducted before a truly automated gait recognition system is viable.

This chapter identifies a number of such possibilities for future research.

Appendix A

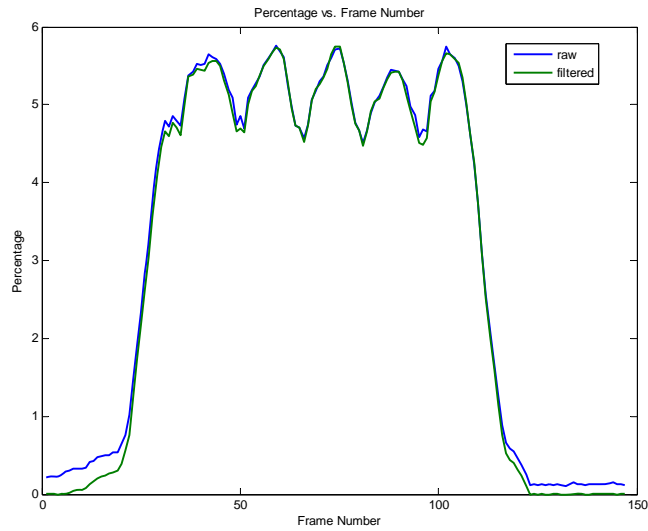Comparison page for the Percentage of Non-Background data in both examples
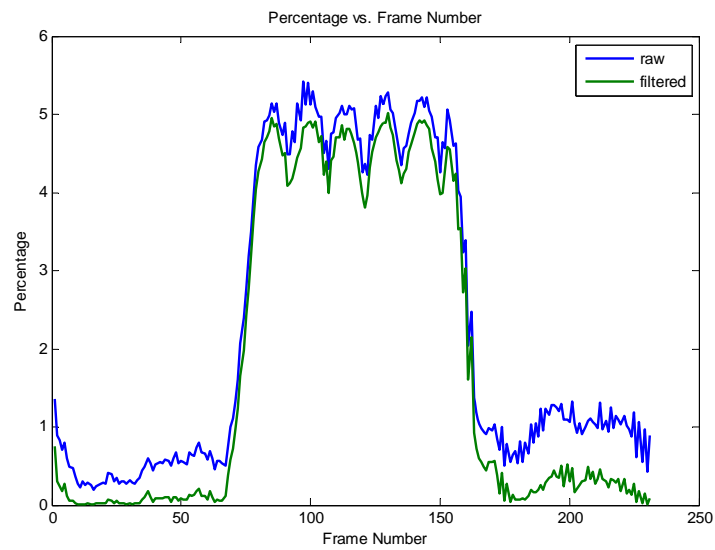

Figure 30. Percentage vs. Frame Number for Individual 1

Figure 31. Percentage vs. Frame Number for Individual 2

Appendix B

MATLAB Programs

```matlab
function [A,photoavg] = RGBBackground(start,stop);

% Reads jpegs into matlab, then finds an avg RGB intensity for
% background, gives% back an average background picture. Also
% gives the background picture in subplots of the Red, Green
% and Blue intensities.

close all

for i=start(1):stop(1),
  A(i-start(1)+1,:,:,:)=imread(['I:\Research\First Pass\Person1\Pass#1\'
  sprintf('frame%i.jpeg',i)]);
end

for i=start(2):stop(2),
  A(stop(1)+i-start(2)+1,:,:,:)=imread(['I:\Research\First
  Pass\Person1\Pass#1\'          sprintf('frame%i.jpeg',i)]);
end

S = size(A,4);
photoavg=mean(A);
photoavg = squeeze(photoavg);
image(uint8(photoavg));

figure
colormap(gray(256))
subplot(3,1,1);image(photoavg(:,:,1))
subplot(3,1,2);image(photoavg(:,:,2))
subplot(3,1,3);image(photoavg(:,:,3))
```

```
function [C,v] = diffavg(A,photoavg,s,n);
% Pulls in the avg background and finds diffavg, the difference
% for all background frames minus avg background. Produces a 3-D
% scatter plot of the difference of the RGB intensity in selected
% frame and covariance.

diffavg1=zeros(size(A),'int16');

for i=1:size(A,1);
    diffavg1(i,:,:,:) = squeeze(int16(A(i,:,:,:))) - int16(photoavg);
end

Asize = prod(size(diffavg1(:,:,:,1)));
for i=1:3
    v(i)= var(reshape(diffavg1(:,:,:,i),Asize,1,1,1));
end

figure
R = squeeze(int16(A(s,:,:,1))) - int16(photoavg(:,:,1));
G = squeeze(int16(A(s,:,:,2))) - int16(photoavg(:,:,2));
B = squeeze(int16(A(s,:,:,3))) - int16(photoavg(:,:,3));
L = size(A,2)*size(A,3);
reorderedindex = randperm(L);
shortindex = reorderedindex([1:n]);

scatter3(R(shortindex),G(shortindex),B(shortindex),'.')
xlabel('Red')
ylabel('Green')
zlabel('Blue')
title('Change in a Single Background minus Average Background')

X = [R(:),G(:),B(:)];
X = double(X);
C = cov(X);
```

60

```matlab
function P = personpict(start,stop);
% Reads into Matlab all the jpegs with images of the individual.
% Puts out a subplot of all pictures of the person and a movie,
% avi file, of the pictures in motion.

for i=start:stop,
  P(:,:,:,i-start+1)=imread(['I:\Research\First Pass\Person1\Pass#1\'
  sprintf('frame%i.jpeg',i)]);
end

figure
for i=1:size(P,4);
    subplot(10,12,i)
    imshow(P(:,:,:,i))
end

fig=figure
mov = avifile('orgdata.avi','compression','Cinepak')
map = colormap(gray(256));
for i=1:size(P,4);
    temp = uint8(P(:,:,:,i));
    image(temp);
    imwrite(temp,map,sprintf('movie/movie%03i.png',i));
    F = getframe(gca);
    mov = addframe(mov,F);
end
mov = close(mov)
```

```matlab
function I = Backsub(A,P,photoavg,k,C,vars,s,n);
% Takes the avg background picture and subtracts the image of the
% person.  Displays an image of the person in binary form for all
% frames in P. Also displays a movie, avi file, of the binary
% pictures in motion.

for i=1:size(P,4);
    I2=int16(photoavg)-int16(P(:,:,:,i));

I(:,:,:,i)=(I2(:,:,1).^2/(k*vars(1))+I2(:,:,2).^2/(k*vars(2))+I2(:,:,3).
^2/(k*vars(3))>1);
    I(:,:,:,i) = medfilt2(I(:,:,:,i));
end

figure
for i=1:size(P,4);
    subplot(10,12,i)
    imshow(I(:,:,:,i))
end

fig=figure
mov = avifile('bwdata.avi','compression','Cinepak')
map = colormap(gray(256));
for i=1:size(P,4);
    temp = double(I(:,:,:,i));
    temp = 255*(temp)+ 1;
    image(temp);
    imwrite(temp,map,sprintf('movie/movie%03i.png',i));
    F = getframe(gca);
    mov = addframe(mov,F);
end
mov = close(mov)

figure
R = squeeze(int16(A(s,:,:,1))) - int16(photoavg(:,:,1));
G = squeeze(int16(A(s,:,:,2))) - int16(photoavg(:,:,2));
B = squeeze(int16(A(s,:,:,3))) - int16(photoavg(:,:,3));
L = size(A,2)*size(A,3);
reorderedindex = randperm(L);
shortindex = reorderedindex([1:n]);

[X,Y,Z] = meshgrid(-20:.5:20);
U = X.^2/(k*C(1,1))+Y.^2/(k*C(2,2))+Z.^2/(k*C(3,3))+
(X.*Y)/(k*2*C(1,2))+(X.*Z)/(k*2*C(1,3))+(Y.*Z)/(k*2*C(2,3));
isosurface(X,Y,Z,U,1)
alpha(.5)
hold on

scatter3(R(shortindex),G(shortindex),B(shortindex),'.')
xlabel('Red')
ylabel('Green')
zlabel('Blue')
title('Change in a Single Background minus Average Background')
```

```
function I = Backsub_C(A,P,photoavg,k,C,vars,s,n);
% Takes the avg background picture and subtracts the image of the
% person. Displays an image of the person in binary form for all
% frames in P. Also displays a movie of the pictures in motion with
% the correlation.

IC = inv(C);

R = squeeze(double(P(:,:,1,s))) - double(photoavg(:,:,1));
G = squeeze(double(P(:,:,2,s))) - double(photoavg(:,:,2));
B = squeeze(double(P(:,:,3,s))) - double(photoavg(:,:,3));
L = size(P,1)*size(P,2);

for i=1:size(P,4);
    I2=abs(int16(photoavg)-int16(P(:,:,:,i)));
    BOOL   =
int16((I2(:,:,1).^2.*IC(1,1))+(I2(:,:,2).^2.*IC(2,2))+(I2(:,:,3).^2.*IC(
3,3))+(2*(I2(:,:,1).*I2(:,:,2).*IC(1,2)))+(2*(I2(:,:,1).*I2(:,:,3).*IC(1
,3)))+(2*(I2(:,:,2).*I2(:,:,3).*IC(2,3)))>36);
    P(:,:,1,i) = medfilt2(BOOL);
end

figure
for i=1:size(P,4);
    subplot(10,12,i)
    imshow(P(:,:,:,i))
end

fig=figure
mov = avifile('bwdata3k.avi','compression','Cinepak')
map = colormap(gray(256));
maxP = double(max(max(max(max(P)))))
for i=1:size(P,4);
    temp = uint8(P(:,:,1,i)*256);
    image(temp); colormap(gray(256));
    imwrite(temp,map,sprintf('movie/movie%03i.png',i));
    F = getframe(gca);
    mov = addframe(mov,F);
end
mov = close(mov)

figure
reorderedindex = randperm(L);
shortindex = reorderedindex([1:n]);

[X,Y,Z] = sphere(20);
X = 6*X;
Y = 6*Y;
Z = 6*Z;
surf(X,Y,Z,'EdgeColor','none','FaceColor','r');alpha(.3);
hold on

IC1 = chol(IC);
R1 = IC1(1,1)*R + IC1(1,2)*G + IC1(1,3)*B;
G1 =             IC1(2,2)*G + IC1(2,3)*B;
B1 =                         IC1(3,3)*B;

scatter3(R1(shortindex),G1(shortindex),B1(shortindex),'.')

title('Change in a Person Frame minus Average Background')
axis equal
axis vis3d
```

```
function [bool_raw,bool_filtered] =
Backsub1(A,P,photoavg,k,C,vars,s,s1,n);
% Takes the avg background picture and subtracts the image of the
% person. Displays an image of the person in for all frames in P
% also displays a movie of the pictures in motion. Uses the
% correlation equation.

IC = inv(C);

R = squeeze(double(P(:,:,1,s))) - double(photoavg(:,:,1));
G = squeeze(double(P(:,:,2,s))) - double(photoavg(:,:,2));
B = squeeze(double(P(:,:,3,s))) - double(photoavg(:,:,3));
L = size(P,1)*size(P,2);

for i=1:size(P,4);

    I2=abs(int16(photoavg)-int16(P(:,:,:,i)));
   BOOL   =
int16((I2(:,:,1).^2.*IC(1,1))+(I2(:,:,2).^2.*IC(2,2))+(I2(:,:,3).^2.*IC(
3,3))+(2*(I2(:,:,1).*I2(:,:,2).*IC(1,2)))+(2*(I2(:,:,1).*I2(:,:,3).*IC(1
,3)))+(2*(I2(:,:,2).*I2(:,:,3).*IC(2,3)))>k^2);
  % P(:,:,1,i) = BOOL;   % To run without filter
    P(:,:,1,i) = medfilt2(BOOL,[4 4]);

  bool_raw(i) = sum(sum(BOOL))/prod(size(BOOL));
  bool_filtered(i) = sum(sum(P(:,:,1,i)))/prod(size(BOOL));

end

figure
for i=1:size(P,4);
    subplot(10,12,i)
    imshow(P(:,:,:,i))
end

fig=figure
mov = avifile('2bwdata5_5filt.avi','compression','Cinepak')
map = colormap(gray(256));
maxP = double(max(max(max(max(P)))))
for i=1:size(P,4);
    temp = uint8(P(:,:,1,i)*256);
    image(temp); colormap(gray(256));
    imwrite(temp,map,sprintf('movie/movie%03i.png',i));
    F = getframe(gca);
    mov = addframe(mov,F);
end
mov = close(mov)
figure
reorderedindex = randperm(L);
shortindex = reorderedindex([1:n]);
[X,Y,Z] = sphere(50);
X = 6*X;
Y = 6*Y;
Z = 6*Z;
surf(X,Y,Z,'EdgeColor','none','FaceColor','r');alpha(.3);
hold on
IC1 = chol(IC);
R1 = IC1(1,1)*R + IC1(1,2)*G + IC1(1,3)*B;
G1 =               IC1(2,2)*G + IC1(2,3)*B;
B1 =                           IC1(3,3)*B;
scatter3(R1(shortindex),G1(shortindex),B1(shortindex),'.')
title('Change in a Person Frame minus Average Background')
xlabel('A*Red')
```

```
ylabel('A*Green')
zlabel('A*Blue')
axis equal
axis vis3d

figure
R2 = squeeze(double(A(s1,:,:,1))) - double(photoavg(:,:,1));
G2 = squeeze(double(A(s1,:,:,2))) - double(photoavg(:,:,2));
B2 = squeeze(double(A(s1,:,:,3))) - double(photoavg(:,:,3));
L = size(A,2)*size(A,3);
reorderedindex = randperm(L);
shortindex = reorderedindex([1:n]);
[X,Y,Z] = sphere(50);
X = 6*X;
Y = 6*Y;
Z = 6*Z;
surf(X,Y,Z,'EdgeColor','none','FaceColor','r');alpha(.3);
hold on
IC1 = chol(IC);
R1 = IC1(1,1)*R2 + IC1(1,2)*G2 + IC1(1,3)*B2;
G1 =                IC1(2,2)*G2 + IC1(2,3)*B2;
B1 =                               IC1(3,3)*B2;
scatter3(R1(shortindex),G1(shortindex),B1(shortindex),'.')
title('Change in a Background Frame minus Average Background')
xlabel('A*Red')
ylabel('A*Green')
zlabel('A*Blue')
axis equal
axis vis3d
```

Bibliography

Begg, R. and J. Kamruzzaman. "A Machine Learning Approach for Automated Recognition of Movement Patterns Using Basic, Kinetic and Kinematic Gait Data," *Journal of Biomechanics*, 38: 401-408 (2005).

Boyd, Jeffrey E. and James J. Little. "Phase Models in Gait Analysis," *Exemplars versus Models Workshop Computer Vision and Pattern Recognition*, Kauai, HI (December 2001).

Fujiyoshi, Hironobu, Member, Alan J. Lipton, Nonmember and Takeo Kanade, Member. "Real-Time Human Motion Analysis by Image Skeletonization," *Institute of Electronics, Information and Communication Engineers Transactions Information and Systems*, E87-D, 1 (January 2004).

Hayfron-Acquah, James B., Mark S. Nixon and John N. Carter. "Human Identification by Spatio-Temporal Symmetry," *Image, Speech and Intelligent Systems Group, 16th International Conference on Pattern Recognition*, 1: 10632 (2002).

Kale, A., N. Cuntoor, B. Yegnanarayana, A. N. Rajagopalan and R. Chellappa. "Gait Analysis for Human Identification," 4th Annual Conference Audio and Video-Based Biometric Person Authentication 2003, *Lecture Notes in Computer Science*, 2688: 706-14 (2003).

Krzanowski, W. J. *Principles of Multivariate Analysis*. New York: Oxford University Press, 1988.

Kshirsagar, Anant M. *Multivariate Analysis*. New York: Marcel Dekker, Inc.: 1972.

Lie, Agus Santoso, Ryo Shimomoto, Shohei Sakaguchi, Toshiyuki Ishimura, Shuichi Enokida, Tomohito Wada and Toshiaki Ejima, T. Kanade, A. Jain and N. K. Ratha (eds.). "Gait Recognition Using Spectral Features of Foot Motion," Audio and Video-Based Biometric Person Authentication: 5th International Conference, Hilton Rye Town, NY, USA, July 20-22, 2005. *Lecture Notes in Computer Science*, 3546: 767 (2005).

Ning, Huazhong, Liang Wang, Weiming Hu and Tieniu Tan. "Model-based Tracking of Human Walking in Monocular Image Sequences," *National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Science, Beijing, P. R. China*, 100080. 15 March 2006 http://www.ifp.uiuc.edu/~hning2/papers/TENCON2_ning&wang.pdf.

Nixon, M. S., J. N. Carter, D. Cunado, P. S. Huang and S. V. Stevenage. "Automatic Gait Recognition," Motion Analysis and Tracking (Ref. No. 1999/103), *Industrial Electrical Engineers Colloquium*, 3: 1-6 (1999).

Nixon, M. S., J. N. Carter, J. M. Nash, P. S. Huang, D. Cunado and S. V. Stevenage. "Automatic Gait Recognition," *Department of Electronics and Computer Science Department of Psychology University of Southampton, Southampton*, 23 April 2006 http://eprints.ecs.soton.ac.uk/641/02/ieegait.pdf.

Tingley, Maureen, Carla Wilson, E. Biden and W. R. Knight "An Index to Quantify Normality of Gait in Young Children" *Gait and Posture*, 16: 149-158 (2002).

Wang, Liang, Tieniu Tan, Senior Member, IEEE, Weiming Hu and Huazhong Ning. "Automatic Gait Recognition Based on Statistical Shape Analysis," *Institute of Electrical and Electronics Engineers Transactions on Image Processing*, 12: 1120-1130, (September 2003).

"WP-11: INTEGRATION: State of the Art in Automatic Human Detection, Motion and Behavior Analysis in Multimedia Data." 15 March 2006 http://www.ercim.org/pub/bscw.cgi/d13655/D11.1.pdf.

Yoo, Jang-Hee, Mark S. Nixon and Chris J. Harris. "Extracting Gait Signatures based on Anatomical Knowledge," *University of Southampton, Southampton*, 23 April 2006 http://www.bmva.ac.uk/meetings/02/6March02/soton2.pdf.

Yu, Shiqi, Liang Wang, Weiming Hu and Tieniu Tan, Senior Member, IEEE, "Gait Analysis for Human Identification in Frequency Domain," *Third International Conference on Image and Graphics*, 282-285, (2004).

Vita

Captain Jennifer J. Samler was born in Springville, New York. She graduated from Robert E. Lee High School, Montgomery, Alabama in May 1995. From there she went to Harrisburg, Pennsylvania to attend The Pennsylvania State University Harrisburg campus where she graduated with a Bachelor of Science in Mathematical Science in May 2001.

Captain Samler entered active duty in July 2001 when she attended the Air Force Officer Training School where she received her Commission in September 2001. Her first assignment was to Eglin Air Force Base, Air Combat Command, 36th Electronic Warfare Squadron where she served as the Low Observable Operations Analyst. Upon completion of this assignment Capt Samler began her graduate studies at The Air Force Institute of Technology (AFIT), pursuing a Master of Science in Mathematics.

Upon completing AFIT, Capt Samler was assigned to Kirkland Air Force Base, Air Force Martial Command, Office of Aerospace Studies.

1. REPORT DATE (DD-MM-YYYY) 06-09-2006	2. REPORT TYPE Master's Thesis	3. DATES COVERED (From – To) Jan 2006 – Sept 2006

4. TITLE AND SUBTITLE

Statistical Approach to Background Subtraction for Production of High-Quality Silhouettes for Human Gait Recognition

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Samler, Jennifer J., Captain, USAF

5d. PROJECT NUMBER
If funded, enter ENR #

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GAM/ENC/06-04

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This thesis uses a background subtraction to produce high-quality silhouettes for use in human identification by human gait recognition, an identification method which does not require contact with an individual and which can be done from a distance. A statistical method which reduces the noise level is employed resulting in cleaner silhouettes which facilitate identification.

The thesis starts with gathering video data of individuals walking normally across a background scene. From there the video is converted into a sequence of images that are stored as joint photographic experts group (jpeg) files. The background is subtracted from each image using a developed automatic computer code. In those codes, pixels in all the background frames are compared and averaged to produce an average background picture. The average background picture is then subtracted from pictures with a moving individual. If differenced pixels are determined to lie within a specified region, the pixel is colored black, otherwise it is colored white. The outline of the human figure is produced as a black and white silhouette. This inverse silhouette is then put into motion by recombining the individual frames into a video.

15. SUBJECT TERMS
Gait Recognition, Background Subtraction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Samuel A. Wright, Maj, USAF
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	77	19b. TELEPHONE NUMBER (Include area code) (937) 255-3636, ext 4549 (samuel.wright@afit.edu)

Standard Form 298 (Rev. 8-98)