2017

# Gait Analysis from Wearable Devices using Image and Signal Processing

Bradley A. Schneider
*Wright State University*

## Repository Citation

Schneider, Bradley A., "Gait Analysis from Wearable Devices using Image and Signal Processing" (2017). *Browse all Theses and Dissertations*. 1879.
https://corescholar.libraries.wright.edu/etd_all/1879

# GAIT ANALYSIS FROM WEARABLE DEVICES USING IMAGE AND SIGNAL PROCESSING

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

BY

BRADLEY A. SCHNEIDER

B.S., Morehead State University, 2012

2017

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

November 27, 2017

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Bradley A. Schneider ENTITLED Gait Analysis from Wearable Devices using Image and Signal Processing Techniques BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Tanvi Banerjee, Ph.D.
Thesis Director

Mateen M. Rizki, Ph.D.
Chair, Department of Computer
Science and Engineering

Committee on
Final Examination

Tanvi Banerjee, Ph.D.

Yong Pei, Ph.D.

Thomas Wischgoll, Ph.D.

Barry Milligan, Ph.D.
Interim Dean of the Graduate School

# ABSTRACT

Schneider, Bradley A. M.S. Department of Computer Science and Engineering, Wright State University, 2017. Gait Analysis from Wearable Devices using Image and Signal Processing.

We present the results of analyzing gait motion in-person video taken from a commercially available wearable camera embedded in a pair of glasses. The video is analyzed with three different computer vision methods to extract motion vectors from different gait sequences from four individuals for comparison against a manually annotated ground truth dataset. Using a combination of signal processing and computer vision techniques, gait features are extracted to identify the walking pace of the individual wearing the camera and are validated using the ground truth dataset. We perform an additional data collection with both the camera and a body-worn accelerometer to understand the correlation between our vision-based data and a more traditional set of accelerometer data. Our results indicate that the extraction of activity from the video in a controlled setting shows strong promise of being utilized in different activity monitoring applications such as in the eldercare environment, as well as for monitoring chronic healthcare conditions.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1 INTRODUCTION

As our current society ages, the need for elderly care and the early diagnosis of age-related diseases will sharply increase. While many diagnosis techniques rely on self-reporting of symptoms, it is acknowledged that self-reporting suffers from inaccuracy and bias as it is difficult to enforce a standard scale across patients. There is too much subjectivity in the description of symptoms and their severity for reliable and accurate diagnosis. What feels painful to one individual may seem completely tolerable to another. In addition to this subjectivity, diagnoses following the observations of symptoms may occur after a catastrophic event such as a dangerous fall has already occurred. Such diagnoses are useful for avoiding repeated issues, but are insufficient for protecting aging patients from the initial problem. It is extremely desirable to be able to anticipate these issues and take preventative action against them before they occur.

Existing research has shown that Activities of Daily Living (ADL) are a good indicator of elderly health, and monitoring such health indicators shows promise in the early diagnosis of age-related disease. ADL's consist of various activities that are performed on a routine basis in one's daily life. Examples of ADL's range from concise, well-defined activities such as handwashing and brushing teeth, to physical activities such as standing up or walking, to longer complex activities such as cooking or doing laundry. Observation of ADL's such as these can provide metrics related to quality and frequency of the activity. Most ADL's are performed in an expected pattern or with some degree of regularity, but when these behaviors change in timing or quality, or cease to occur at all, this is usually an indication of a new or worsening health condition. Thus, being able to monitor these activities provides rich input to the early recognition of medical conditions.

In addition to potentially aiding early diagnosis of medical conditions, monitoring of ADL’s can also provide a rich means of quantifying rehabilitation progress following an injury. When a patient suffers an injury that impacts his or her ability to continue their daily routine, the goal is to rehabilitate the injury as quickly as possible so their routine can be resumed. Since ADL’s are, by definition, activities that are performed every day by almost everyone, they are an excellent measure of basic physical ability. Automated monitoring of ADL’s could be used to remove error associated with self-reporting of progress, or the hassle of making doctor’s office visits for periodic examination. The monitoring of parameters indicating the quality of ADL’s during the rehabilitation process can give a quantifiable measure of progress to physicians.

In this paper we focus specifically on monitoring gait (walking) activities and extracting information on how the locomotion is occurring. We aim to provide a non-invasive, lightweight, and wearable system for extracting gait information from a subject. Our system aims to be operable in the home to maximize convenience and usability. Existing methods of in-home monitoring require extensive setup of the home environment with external sensors. It may be costly or inconvenient to install these sensors, and the area in which they are useful is typically limited to a single room. Video-based monitoring is a common approach, but may also come with privacy concerns and does not solve the limitation of only providing information about a single environment in which it is installed.

We address the limitations of existing methods by using wearable sensors. Because the sensors are worn on the subject's body, they are effective in nearly any environment in which the subject may be located. Additionally, the cost of installing sensors in a pre-determined environment, such as the home, is avoided.

A single camera mounted in a pair of eye glasses provides the primary input to the system. Gait analysis performed with computer vision on wearable devices is a challenging task due to the limited first-person perspective of the device. Consequently, there may be a lack of information available to describe the movement of the subject. However, it has the benefit of requiring little configuration and is not prone to accumulated error over time, as wearable accelerometers or positioning devices may be [1, 2]. Approaches of this type rely on the predictability of a subject's movement during locomotion to recognize gait. Patterns of movement in the captured video are recognized and lead to the recognition of the gait activity and extraction of gait parameters.

To better understand the data that we collect from the video source, we also include the use of an accelerometer in some data collection activities. The accelerometer is worn simultaneously with the camera by the subject, but measures acceleration at the torso instead of the head. The accelerometer provides a physical measure of movement of the subject, albeit in a slightly different location. While the gait parameters extracted from the video are verified using techniques that to not rely on the accelerometer data, this data provides a more traditional measure of movement as a comparison for the movement described by the video techniques. We seek to correlate the parameters of the two data sources in order to verify our results.

The rest of this paper is organized as follows: Chapter 1 contains an overview of related work; Chapter 1 describes the sensors, data collection process, and methods used in this work for data collection and processing; Chapter 1 contains the results of the experiments conducted described in chapter 1; Chapter 1 contains a discussion and analysis of the obtained results. We finally conclude with the Conclusions chapter (Chapter 6) and the Future Work (Chapter 7).

# 2 RELATED WORK

This section contains an overview of the current body of work related to our study of gait through wearable and video-based sensors.

## 2.1 Non-wearable Video-based methods

A dominant challenge in the domain of ADL monitoring is finding a method of monitoring which does not create too large of a burden on the patient being assessed. Monitoring ADL’s involves potentially invasive data collection in the home, and privacy concerns of the user must be addressed. Activity detection through video analysis has been an active field of research due to the non-invasiveness of the approach. Methods based on other (non-video) sensors typically require the subject to be instrumented with cumbersome gadgets such as accelerometers that may hinder the normal behavior pattern of users. If the subject is not able to perform the activity in the way they normally would, the data collection will not allow for measuring the activity as it is usually performed. With traditional video-based techniques, stationary cameras are deployed in the environment, making for a more practical alternative. The cameras do not interfere with the subject as the activity is being performed.

Banerjee, et. al describe a method for identifying ADL’s in order to study behavior patterns of the elderly to detect health changes using a Microsoft Kinect depth camera [3]. Because the camera is placed with an external view of the environment, information on the movement of the subject as well as the context of labeled items in the scene are used to detect activities using a hidden Markov model. While the third-person camera approach often comes with privacy concerns, the use of only the depth channel of the sensor provides in implicit amount of privacy.

In [4], Anderson, et. al use linguistic summarizations of temporal fuzzy inference curves to represent the state of a three-dimensional object called a voxel person to perform automated fall detection for elderly residents of a nursing home. Multiple cameras are required to construct the three-dimensional voxel representation, and to address privacy concerns, only the human silhouette is extracted from the frame, and raw video is discarded. Fuzzy set theory is used to classify the state of the voxel person as upright, in-between, or on the ground.

These works clearly indicate that video-based approaches can be successfully developed for activity recognition. However, since the environment must be instrumented with one or more cameras, these methods are constrained to operate within that closed space or the camera's field of view, and cannot be used elsewhere without the installation of additional cameras. While this is acceptable for certain activities that always occur in a predictable place, such as handwashing or cooking, it is severely limiting for the analysis of activities such as walking, which require a larger, more open environment. Additionally, these methods focus on activity classification only, and do not describe the actual manner in which the activity was performed.

## 2.2 Wearable Video-based Methods

Wearable vision sensors can provide a convenient solution to the complexity problems from which other techniques suffer. However, many published methods of gait analysis that use wearable technology focus more on instrumenting the subject with accelerometers or other inertial-based sensors which can more directly give an accurate indication of movement during locomotion. For this reason, there are fewer methods successfully using computer vision at this point in time.

Cho et al. describe an approach combining multiple types of sensors in an attempt to take advantage of both the accuracy of accelerometers and the convenience of computer vision [5]. A multi-resolutional, grid-based optical flow method is applied to video collected from a wearable vision sensor. By examining different regions of the video for flows in different directions, activities such as walking forward or backward, turning, and sitting down are able to be detected from the video. Activities detected by the video component of the system are limited to those describing movement in the environment, and two similar activities may not be distinguishable. This video-based approach relies solely on the perceived direction of movement of the vision sensor through the environment, and the video is not used to estimate the pose of the wearer. Because of the lack of consideration for the physical effects of locomotion on the movement of the camera, this approach is equally suited to estimating vehicular or robotic movement. That is, the camera is used only to perceive location, and does not provide any parameters on the quality or form of the gate. To improve the activity detection accuracy, accelerometers were also placed on the subject, contributing information about the estimated pose of the wearer.

Taking advantage of published information on human joint movement during locomotion to estimate the pose of the subject, rather than just the movement of the subject relative to the environment, can lead to the extraction of more detailed information about the gait. For this purpose, it is necessary to utilize information on human behaviors to construct a model of human motion. Hirasaki et al. document the effect of locomotion on the head and body, demonstrating the translation and rotation during the gait [6]. Unuma et al. modeled human locomotion using Fourier expansions due to the cyclic nature of the activity [7]. These methods and other related works give a basis for building accurate models of human gait for analysis in egocentric video, enabling approaches based entirely on video.

In such an approach based entirely on video, Watanabe et al. take advantage of the predictability of limb motion during locomotion to analyze movement and classify gait from a single video source [2]. A downward-facing camera is attached to the leg so that it captures the motion of the environment as the locomotion occurs. A calibration step is performed to document the movement of the camera on the leg during the walking activity. Walking samples are taken by a motion-capture system and models of various states, including a slow walk, walk, and run. A model is formed based on the waist position, traveling speed, and angular speed at a given time. The action is described by up to a fifth order Fourier transform, which represents the cyclic motion described in the human motion studies [6, 7]. State prediction for the walking state occurs at each sample in time and is based on a likelihood estimation from the captured parameters, and includes an error calculation based on the difference in the expected location of known landmarks in the environment and the actual observed location.

While these works successfully predict walking state based on the video input, we are interested in being able to also capture the parameters of the gait in order to describe quantitatively the way in which the activity is being performed.

# 3 METHODS

In this section, we discuss the devices and methods used to collect data. We also discuss the methods and algorithms used to process the collected data.

## 3.1 Hardware

The intent of our study is to design a system with a sensor that can be worn without imposing a significant physical burden to user. To accomplish this task, we select a video camera embedded within a pair of glasses to collect first-person video segments. Multiple commercial solutions that fit this constraint currently exist or are in development. We chose to use the Pivothead SMART Architect Edition glasses [8] for our video capture, which are pictured in Figure 1. This device appears nearly identical in physical form to a common pair of eyeglasses, but provides a high-definition video sensor in the bridge of the glasses over the nose. The earpieces also provide a pluggable platform into which an extended battery and Wi-Fi module may be inserted, as well as a micro-SD storage card.

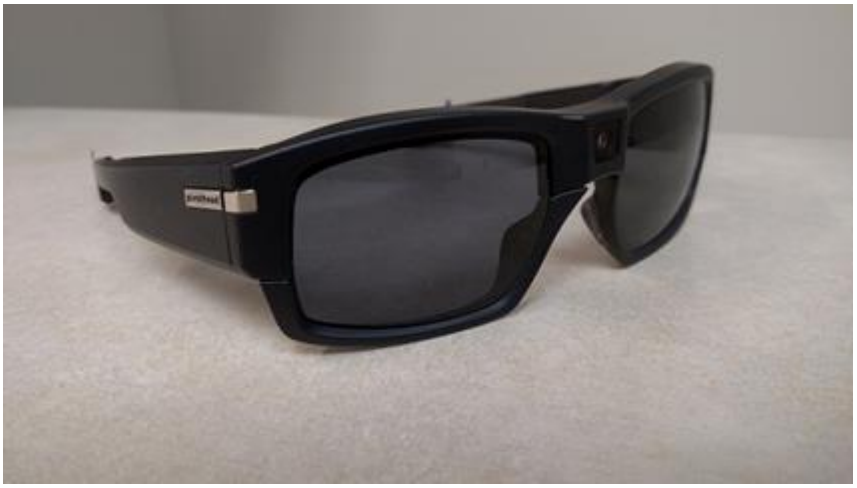

*Figure 1.  Pivothead SMART Architect Edition glasses*

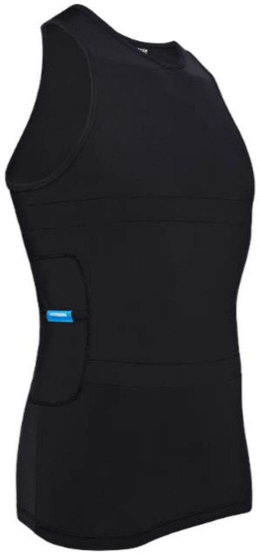

*Figure 2. Hexoskin Wearable Body Metrics Vest*

Classic approaches to the problem of activity recognition and analysis with wearable devices involve the use of accelerometers [9, 10, 11]. To provide a second data set to which the data from the Pivothead glasses may be compared, we also selected a commercial fitness vest which collects accelerometer data over the course of the activity. The chosen device is the Hexoskin Smart Shirt, which is shown in Figure 2 [12]. While the vest is more cumbersome to take on and off, it provides a combination of a more traditional sensing modality and a relatively convenient and user-friendly device. The Hexoskin shirt also collects pulse and respiration data which are not intended to be used for the purpose of our study.

## 3.2 Data Collection

The following sections describe the two data collection events that occurred during this study to collect data from subjects performing gait activities.

### 3.2.1 Video-only Data Collection

In the initial validation of the automated method against manually annotated truth data, video was collected from four subjects as they performed a walking activity while wearing the Pivothead glasses. In order to control as many aspects of the environment as possible, the activity was performed indoors on a treadmill set at a constant speed. This provides a static background for the video and eliminates the possibility of capturing movement in the video which was not a consequence of the activity being performed. For example, a video recorded during a walk in the outdoors may include images of other pedestrians. These pedestrians will have an independent trajectory through the frame which is not consequential to the movement of the camera. Therefore, our method would not want to consider this motion.

### 3.2.2 Video and Accelerometer Data Collection

Once the initial validation was complete, additional data was collected in the same environment from five subjects. Each subject in this activity wore both the glasses and Hexoskin accelerometer vest. The video technique was assumed to be valid, so rather than comparing against annotated truth data, the second collection event had the goal of correlating the outputs of the video analysis with the outputs from the accelerometer, hence the addition of the second sensor. As with the first data collection event, this collection was conducted in a static environment using a treadmill to eliminate extra movement in the view of the camera. Each participant walked at five different speeds to form five sets of video/acceleration data per subject. The speeds range from 2 to 3.5 miles per hour. This was performed so that we could not only test the generalizability of the correlation across participants, but also identify whether walking at a slower or faster pace impacted the results of the correlation.

The data collection method was designed to ease the synchronization of the data collected from the camera and the accelerometer. The camera device was turned on first in each collection. Following this, the accelerometer was initialized in view of the camera. In this way, a rough approximation of the difference of start times of the two devices may be computed.

The approximation of the starting times gives some idea of the initial offset of the two data recordings, but does not provide enough precision. To get a more precise alignment, we perform an activity to embed a physical marker in the data, with the goal of producing a unique waveform in both the accelerometer and the video motion. Prior to beginning a gait activity, the subject would complete a brief aperiodic movement, followed by at least three seconds of no movement.

The beginning of the gait activity is marked by the transition between very low movement (expressed in the sensor as very low amplitude of accelerometer readings) and an apparent return to periodic behavior. An alignment of the data was chosen based on this transition which was initially found to be too imprecise. To refine the data alignment, we further perform a cross-correlation of the signals from each device (i.e. the camera and the vest). The cross-correlation computes a correlation score of the waveforms with one dataset fixed in time and the other moved temporally by a single frame. Assuming that the two waveforms do in fact have some correlation, a perfectly chosen alignment will produce the highest correlation at zero (given that they have already been approximately well-aligned). In reality, the correlation from the rough marker position tended to be somewhere near to zero, but not at zero. This shows that our rough marker was not perfectly chosen, so we then shifted the alignment of the data to maximize that correlation.

## 3.3 Motion Extraction Method

Our initial study aims to validate the use of a body-worn camera as a sensing device for collecting parametric gait information using the video collected in 3.2.1. The first step in the process of converting raw video to parametric gait information is to extract motion vectors from the collected video. It is our observation that the major component of motion in first-person video is caused by the motion of the camera itself, rather than by the independent movement of objects within view of the camera. We take advantage of this observation to determine the motion of the subjects' head while wearing the fixed camera. We evaluated three methods of automatically extracting motion from video.

### 3.3.1 Dense Optical Flow

Dense optical flow is an algorithm that takes two consecutive frames of video as input and provides as output a motion vector for each pixel in the frame. Since there is an attempt to describe the movement of each and every pixel between the two frames, the algorithm is computationally expensive, but may also provide a more complete view of the motion in different regions of the video frame since each pixel is considered. Figure 3 illustrates the vector output of performing Dense Optical Flow on a single frame of the captured video.

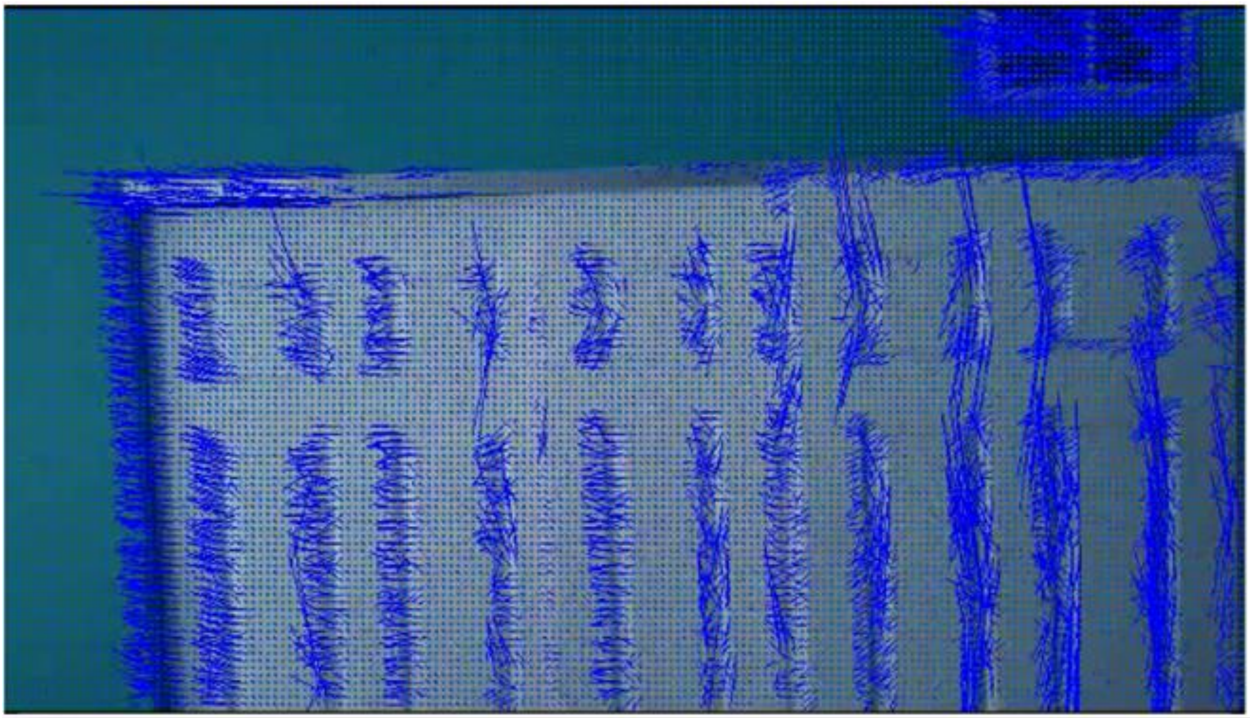

*Figure 3. Sample output from a single frame with the Dense Optical Flow technique. Each pixel is evaluated and every fifth motion vector is plotted here with a multiplied magnitude for the purpose of visualization.*

We employ the algorithm described by Gunnar Farneback, which bases the calculation of motion on the displacement estimation for each pixel neighborhood. The displacement estimation for the pixel neighborhood is approximated by a polynomial expansion [13]. The general idea of the algorithm is to approximate local neighborhoods of a pixel with a polynomial of the form given in equation (1) .

$$f_1(x) = x^T A_1 x + b_1^T x + c_1 \tag{1}$$

where $A_1$ is a symmetric matrix, $b_1$ is a vector, and $c_1$ is a scalar.

If a new function (image) is constructed by adding a displacement $d$, this yields equation (2) below.

$$f_2\,(x) = f_1(x - d) = (x - d)^T A_1(x - d) + b_1^T(x - d) + c_1 \qquad (2)$$

$$= x^T A_2 x + b_2^T x + c_2$$

where $A_2$ is a new symmetric matrix, $b_2$ a vector, and $c_2$ a scalar.

Since $f_2$ is a translation of $f_1$, we can set the coefficients of equations (1) and (2) equal to each other, and generate the following system of equations:

$$A_2 = A_1, \qquad (3)$$

$$b_2 = b_1 - 2A_1 d, \qquad (4)$$

$$c_2 = d^T A_1 d - b_1^T d + c_1 \qquad (5)$$

The displacement $d$ can be solved for using equation (4). This is the basis by which Dense Optical Flow provides a motion estimation. The displacement between two consecutive frames describes the amount of motion between them for the given pixel.

### 3.3.2 Sparse Optical Flow

Sparse optical flow is another algorithm used to provide an estimation of motion for video features across frames. However, unlike Farneback’s method, motion vectors are calculated only for regions of the frame that are deemed to be robust for making such motion detections, rather than for every pixel. This requires a two-step process of (1) image feature extraction and (2) optical flow calculation across the set of identified features. Since motion is only computed for a subset of each frame (i.e. the identified features), the algorithm is more efficient.

The feature extraction method that was used was the Scale Invariant Feature Transform (SIFT) [14]. SIFT identifies features which are scale-invariant. Such features are desirable for motion detection across frames of video. Because they are scale-invariant, the features should be easily identified even as they move around the frame, and even if they move nearer to or further from the camera. A reliable identification of the same feature in different positions between frames will likely lead to a more accurate estimate of motion.

The features extracted by SIFT are passed as input to a sparse optical flow implementation derived from the work of Lucas and Kanade [15]. The Lucas-Kanade optical flow method requires as preconditions that the time increment (and by extension, the distance that a feature moves) between frames is very small, and that the intensity values change across frame images smoothly. This is because the optical flow method was originally intended to accomplish *image registration*, or finding the same sub-image in two different images. As the time step between two frames goes to zero, the location of a given sub-image, or image feature in the case of optical flow, in the second image is nearer to its original location in the first image. With this assumption, the method is able to restrict its search for the displacement to the neighborhood of the feature's original location, similarly to Dense Optical Flow. It is assumed that the camera frame rate of thirty frames per second is a sufficiently high framerate for gait-related activities, and that naturally occurring scenes tend to have a smooth intensity gradient.

For two images $F(x)$ and $G(x)$, the image registration solution requires finding a disparity vector $h$ that minimizes the difference in an image $F(x+h)$ and $G(x)$. To illustrate their algorithm, Lucas and Kanade use the single-dimensional case. In a single dimension, and for small enough $h$, $h$ can be approximated with the following:

$$F'(x) \approx \frac{F(x+h) - F(x)}{h} = \frac{G(x) - F(x)}{h} \tag{6}$$

$$h \approx \frac{G(x) - F(x)}{F'(x)} \tag{7}$$

Since the value of $h$ depends on $x$, it is suggested that a single score for the difference of the two images is found using the average:

$$h \approx \frac{\Sigma_x \frac{w(x)[G(x) - F(x)]}{F'(x)}}{\Sigma_x w(x)} \tag{8}$$

where $w(x)$ is a weight applied at each $x$ that is inversely proportional to the difference in the rate of change of $G$ and $F$ at $x$.Unfortunately, the equation given in (8) is undefined when $F'(x) = 0$. To fix this, Lucas and Kanade replace (7) with the following approximation:

$$F(x+h) \approx F(x) + hF'(x) \tag{9}$$

Using this new approximation, a replacement for (8) that generalizes to multiple dimensions and avoids division by zero can be given by:

$$h \approx \frac{\Sigma_x F'(x)[G(x) - F(x)]}{\Sigma_x F'(x)^2} \tag{10}$$

In iterative form, which provides a sequence of $h_i$ that converges to the best $h$, and with a similar weighting function to (8):

$$h_0 = 0, \tag{11}$$

$$h_{k+1} = h_k + \frac{\Sigma_x w(x) F'(x + h_k)[G(x) - F(x + h_k)]}{\Sigma_x w(x) F'(x + h_k)^2}$$

In multiple dimensions, the technique is similar, but uses the gradient instead of the derivative:

$$h \approx \frac{\Sigma_x \left(\frac{\delta F}{\delta x}\right)^T [G(x) - F(x)]}{\Sigma_x \left(\frac{\delta F}{\delta x}\right)^T \left(\frac{\delta F}{\delta x}\right)} \tag{12}$$

### 3.3.3 Speeded Up Robust Features Matching

Speeded Up Robust Features (SURF) matching is another algorithm which provides an estimation of motion between frames for features of video. Instead of estimating a flow vector for each feature, as the optical flow methods do, SURF matching attempts to identify the same features in each input frame independently [16]. When the position of the same feature is known in the two consecutive frames, it is then possible to build the motion vector by taking the difference of the two positions.

## 3.4 Motion Vector Consolidation

Through the use of the optical flow and SURF matching algorithms, we are seeking to describe the movement of the camera worn by the subject over time. Because the camera is stationary and attached to the subject, the movement of the camera directly describes the movement of the subject. However, each of the chosen algorithms provide numerous differing motion vectors per frame of video. Depending on the number of features chosen in each frame and how accurately the motion is determined by the algorithms, it is possible to have disagreement between the vectors in a single frame, which necessitates a function that can give a single overall result that best represents the larger-scale motion of the frame as a whole.

To achieve motion vector consolidation and produce a single resulting motion vector per frame of video from multiple feature vectors, we evaluate four common statistical measures – $min$, $max$, $median$, and $mean$. These simple measures were chosen to handle potential tendencies of the motion algorithms to under- or over-estimate the motion vectors. Because the motion vectors have two components, $< u, v >$ (derived from either (4) or (12)), we calculate the statistical measures based on the magnitude of the candidate vectors rather than on the individual components.

## 3.5 Waveform Parameter Extraction

While the investigation into optical flow algorithms advances the goal of describing the motion of the camera by examining the flow vectors extracted from the video, it does not directly help to describe the activity that was being performed. To help view the changes in activity over time, the optical flow vectors are split into $u$ and $v$ components and plotted against time, constructing two waveforms – one describing the amount of horizontal motion over time, and one describing the vertical motion over time.

The expectation of each waveform is that it exhibits periodic behavior. It is known from physiological research that the head moves in cyclic patterns during a walking activity. Since the camera is attached to the head, we also expect the camera to move in a cyclic motion, and therefore expect to see a cyclic pattern in the motion vectors over time. To extract parametric information from the waveforms, we look to signal processing techniques to give information from the frequency domain. By capturing parameters that describe the waveform activity over time, we are directly capturing parameters describing the motion of the subject.

### 3.5.1 Periodograms

The periodogram of a signal gives an estimation of the spectral density, revealing the frequency components that comprise the signal. It describes the amount of power present in the signal at a certain frequency. Therefore, a single distinct peak in the periodogram identifies a strong sinusoidal component at that frequency. We use this spectral density analysis to find the periodicity of the motion represented by our extracted waveforms.

### 3.5.2 Coherence

The magnitude-squared coherence (or just coherence) is a statistic used to discover relationships between two signals with regard to their spectral content. For two waveforms $x(t)$ and $y(t)$, the coherence is defined as

$$C_{xy}(f) = \frac{\left|C_{xy}(f)\right|^2}{C_{xx}(f) * C_{yy}(f)} \tag{13}$$

where $C_{xy}(f)$ is the cross-spectral density of $x$ and $y$, and $C_{xx}(f)$ and $C_{yy}(f)$ are the auto-spectral densities of $x$ and $y$ respectively. Since the coherence measure is based on the spectral densities of the signals, it gives a more formal comparison of the frequency content than the periodogram. We can see from the equation that the coherence approaches 1 as the spectral densities of $x$ and $y$ are more similar, and approaches 0 as they differ. Similar to the periodogram analysis, we expect that the coherence measure of signals representing an activity with the same period will have a value near to 1 at least at the matching frequency.

## 3.6 Signal Smoothing

The collection of data from both the camera and accelerometer sensors is affected by various kinds of noise and error, depending on the sensor. In the case of the camera sensor, image noise, poor exposure, and blur are all sources of error. Each of these provides visible artifacts in the recorded images and negatively impacts the motion extraction process. In the case of the accelerometer, data is affected by inherent inaccuracies as well as physical noise caused by undesired movement of the sensor within the pocket of the vest. Before attempting to correlate the data from these sensors, it is desirable to smooth the signals and eliminate local variance due to one of the mentioned sources of error.

A simple moving average is used to smooth the data. For a signal $f$ and a given window size $w$, we define the moving average with the following equation:

$$f(n) = \frac{1}{w}[f(n) + f(n-1) + \cdots + f\big(n-(w-1)\big)] \tag{14}$$

## 3.7 Correlating Walk Frequency to Speed

An important goal of our study is to examine and understand the correlation of data between the two sensors (the camera and accelerometer), and also the correlation of the data from each sensor individually to the actual pace at which the recorded activity was performed. Correlation of raw signal data to other raw signal data often gives poor results due to noise in the two signals. Even after smoothing or otherwise reducing noise in the data, the correlation may not perform well and is especially sensitive to the sampling rate and alignment of the data in the time domain. For this reason, we avoid performing a linear correlation between the waveforms directly. Instead, we perform correlation on features derived from the signals. In this case, we correlate the features presented in Table 1. These features were collected during the data collection described in section 3.2.2. Figure 8 shows a visual depiction of the axes/channels used in the dataset relative to the subject.

*Table 1 - Description of features in correlation data set*

| Feature | Description |
|---|---|
| $s_T$ | The speed of the treadmill |
| $s_U$ | the peak frequency in the U (horizontal) channel from the glasses times stride length |
| $s_V$ | the peak frequency in the V (vertical) channel from the glasses times stride length |
| $s_X$ | the peak frequency in the X (forward/backward) channel from the accelerometer times stride length |
| $s_Y$ | the peak frequency in the Y (vertical) channel from the accelerometer times stride length |
| $s_Z$ | the peak frequency in the Z (horizontal) channel from the accelerometer times stride length |

Each of the computed features involves a factor of the estimated stride length of the subject. The estimated stride length is computed based on the height of the subject using equation (15) [17]. This is necessary because subjects of differing heights (stride lengths) will take steps at different frequencies when walking at the same speed. Adjusting the features by a factor of stride length corrects for this difference when correlating across subjects.

$$length = \begin{cases} 2 * 0.415 * height\ \ for\ male\ subjects \\ 2 * 0.413 * height\ for\ female\ subjects \end{cases} \quad (15)$$

The six features are computed per trial per subject for a total of 25 recorded values per feature. First, we linearly correlate the data from each channel within each sensor. That is, we correlate the $U$ and $V$ channels from the camera with each other, and we correlate the $X$ and $Y$, $X$ and $Z$, and $Y$ and $Z$ channels from the accelerometer. Following this, we correlate across sensors, correlating the $U$ with each of $X$, $Y$, and $Z$, and $V$ with each of $X$, $Y$, and $Z$. This builds an understanding of how the data from each sensor is providing related information. Finally, all five channels are correlated to the walking speed of the subject.

# 4 RESULTS

In this section we discuss the results from each of the methods in Chapter 1 that were applied to the collected data during our study.

## 4.1 Motion Extraction Method

The Sum of Squared Errors (SSE) measure was used to evaluate the performance of the optical flow methods and vector consolidation functions for extracting the video frame motion. The SSE was computed for the output of each of the three optical flow methods on the same input video to determine which method gave the result most consistent with the annotated truth data. The SSE was calculated four times per optical flow method (once for each statistical aggregation method) to determine the effect of the various aggregation methods on the output of the algorithms. The results of this calculation are presented in Table 2. While the SSE measure itself is unit-less, the relative ranking of the SSE values in ascending order is indicative of the method that most effectively matched the truth data.

For Dense Optical Flow method, the lowest SSE produced was in conjunction with the mean aggregation method, which resulted in an SSE of 47.253. The Sparse Optical Flow produced SSE measures of 6.421 and 7.080 with the median and mean aggregation methods, respectively. The best result with SURF matching was given using the minimum statistical method and resulted in an SSE of 51.334. Based on the SSE measure, when using the median, mean, and max statistical methods, Sparse Optical Flow produced the best result, followed by Dense Optical Flow, and then SURF matching produced the worst results.

*Table 2. Computed SSE Values by Algorithm and Aggregation Method, Normalized by Total Frame Count*

| | Dense Optical Flow | Sparse Optical Flow | SURF Matching |
|---|---|---|---|
| min | 53.846 | 45.212 | 51.334 |
| median | 53.689 | **6.421** | 1023.548 |
| mean | 47.253 | **7.080** | 178.964 |
| max | 4097.331 | 877.551 | 489795.918 |

## 4.2 Waveform Parameter Extraction

Two waveforms are constructed from the output of the optical flow technique – one describing the horizontal component of motion, and the other describing the vertical component of the motion. Mapping the individual components of the output vectors over time produces two waveforms. Similar waveforms were also manually constructed from the annotated truth data for the same video inputs, providing a baseline against which the automated methods may be evaluated. We are interested in measuring the similarity between the waveforms from the optical flow technique and the manually collected truth data to determine whether the methods are appropriate.

### 4.2.1 Cross-Correlation

The cross-correlation of two signals measures the similarity between them at different lag intervals. We compute the cross-correlation of the manually constructed truth waveforms and the waveforms output by optical flow. In this case, a high correlation at lag time $t = 0$ will indicate that the two signals have similar content. A sample of the cross-correlation output is shown in Figure 4. As expected, a peak at $t = 0$ confirms the similarity between the signals. The shape of the cross-correlation result also confirms the periodicity of the signal. As the signals come in and out of phase with a consistent frequency, the cross-correlation plot contains equidistant peaks.

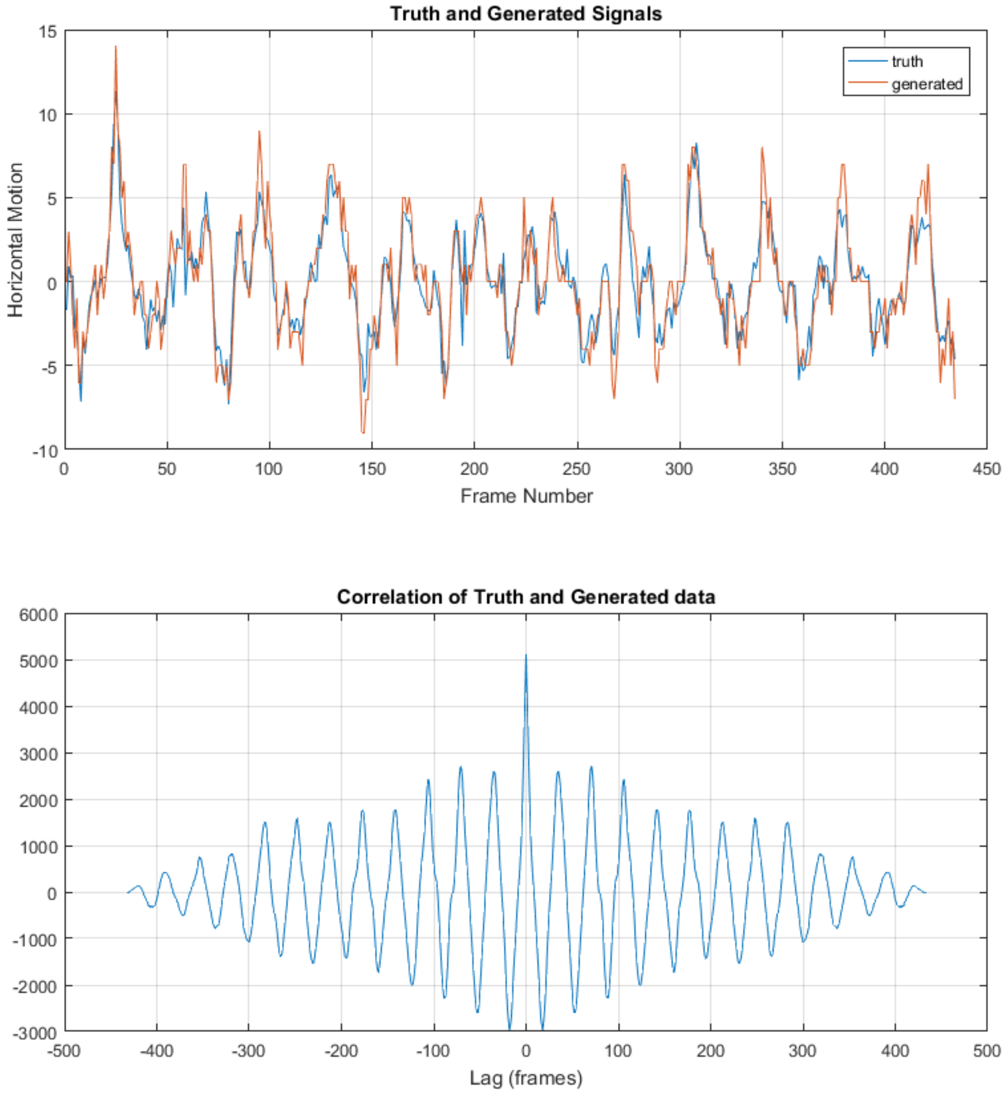

*Figure 4. Plot of generated waveform and truth waveform for horizontal component of motion (top) and plot of cross-correlation of the two signals (bottom)*

### 4.2.2 Periodograms

After extracting the motion vectors using the outlined algorithms, we begin extracting parametric information from the generated waveforms in the $u$ and $v$ dimensions. This process was performed for the data collection described in section 3.2.1. A periodogram is constructed for the truth data and also for each waveform. Peaks were identified in each periodogram to indicate the largest frequency components.

The periodograms for the $u$ dimension of motion for two of the four subjects at the two speeds are presented in Figure 5. At 2.3 miles per hour, subjects A and B had an identified walking pace of .793 Hz, or one step every 0.63 seconds. At 3.9 mph, the identified rate for both increased to 1.02 Hz, or one step every 0.49 seconds. Similarly, at 2.3 miles per hour, subject D also had a calculated gait pace of .793 Hz, while subject C had a pace of .963 Hz, or one step every 0.52 seconds.

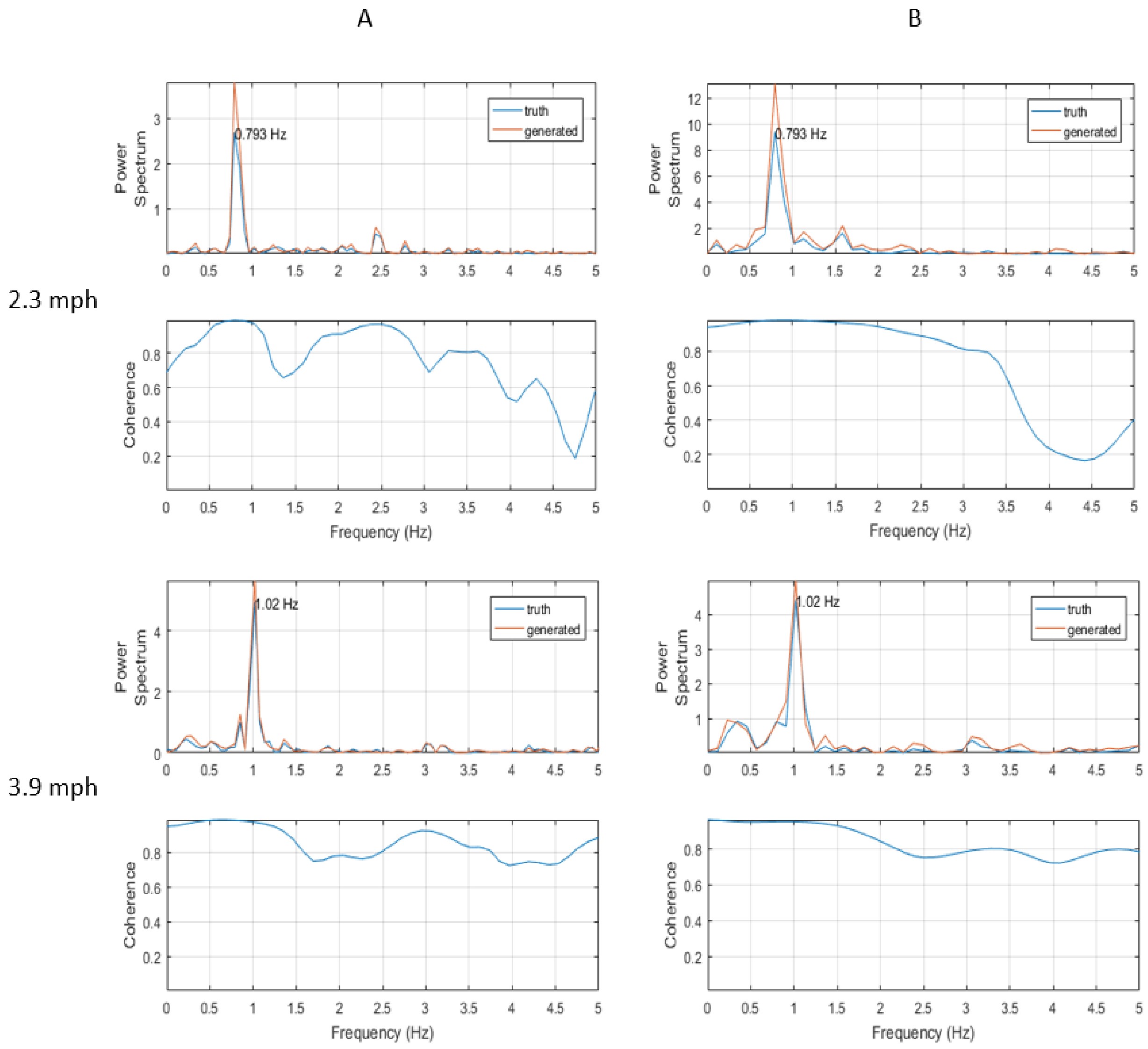

*Figure 5. Plots of periodograms (calculated and truth) and coherence of calculated to truth waveforms for subjects A and B at 2.3 mph and 3.9 mph*

### 4.2.3 Coherence

The result of computing the coherence between the generated waveforms and the manually annotated truth waveforms is also shown in Figure 5 for two of the subjects at two speeds. We notice that the coherence measure for the truth and generated waveforms is 1 at each of the peak frequencies in the periodograms. At other frequencies, the coherence measure is often less than 1, but these frequencies do not represent the activity that was being performed so we do not expect to find a strong coherence.

## 4.3 Signal Smoothing

We applied the moving average technique to the data recorded from each sensor with the goal of eliminating undesirable noise and erroneous values from the data. Figure 6 shows four periodograms built from the same recorded data using the accelerometer sensor with different sizes of moving average filters applied to the time series data. The upper left was built from data with no averaging, the upper right with a window of four frames, the lower left with a window of 8 frames, and the lower right with a window of 16 frames. The periodograms built from the unsmoothed data and the data with a window size of $w = 4$ contain peak frequency content at just below 10 Hz, which is much quicker than the walking activity was taking place. The periodograms built with window sizes $w = 8$ and $w = 16$ contain peak frequencies much nearer to 1 Hz, which is much more representative of the activity that was being performed. As the size of the averaging filter increases, the higher frequency components are removed from the data.

We know by the Nyquist-Shannon sampling theorem that the sampling rate must be at least twice the maximum frequency identified in the data [18]. While smoothing the signal does not strictly modify the sampling rate, it does have the effect of eliminating higher frequencies from the data. Sharp changes in the signal within the window are eliminated by the average operation. Thus we exercise caution in choosing too large of a window for the moving average as we do not wish to lose important frequency information from the signal. We do not perform averaging with a window larger than one fourth of the sampling rate in order to preserve the integrity of the data.

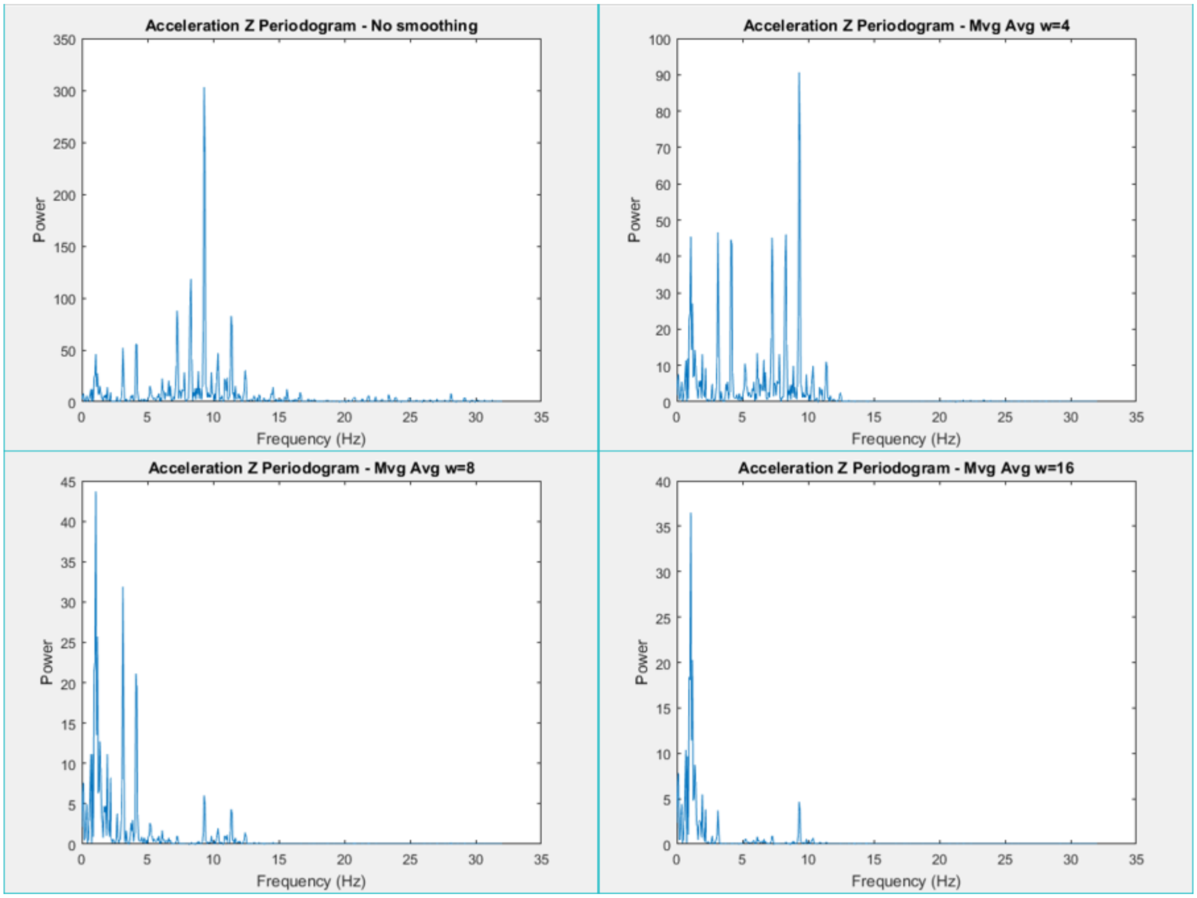

*Figure 6. Periodogram of accelerometer data (z-axis) after applying moving average with 4 different window size ($w = 0, 4, 8, 16$)*

Fortunately, the activities we wish to examine do not occur at these higher frequencies. That is, we can logically assume that gait does not occur at 10 Hz (e.g. 20 steps per second), so smoothing the values and filtering out these frequencies does not affect the frequency data that is considered significant to our findings.

## 4.4 Correlating Walk Frequency to Speed

The computation of the features in Table 1 require the identification of peak frequencies in the periodograms of each subject wearing the camera and accelerometer sensors at each of the five different speeds. Unfortunately, even after attempting to remove noise with the techniques described in section 3.6, some data remained too noisy to give a single clear peak. In these cases, the periodogram may have multiple competing peaks that are similar in height, complicating the task of choosing a single representative frequency. When a visual inspection of the periodogram revealed such a situation, the data were discarded from the correlation computations as a bad collection, regardless of whether it may cause the correlation to improve or worsen. Figure 7 shows an example of a histogram that led to discarded data.

Here we see that our algorithm has identified a single maximum peak in the signal at a frequency of 2.227 Hz. However, several other local maxima exist in the signal. There is another peak of nearly equal power near 0.5 Hz, and multiple other significant peaks between the two largest. Such a number of large peaks in the periodogram indicate a very noisy signal resulting from inaccurate results from the optical flow method that was used.

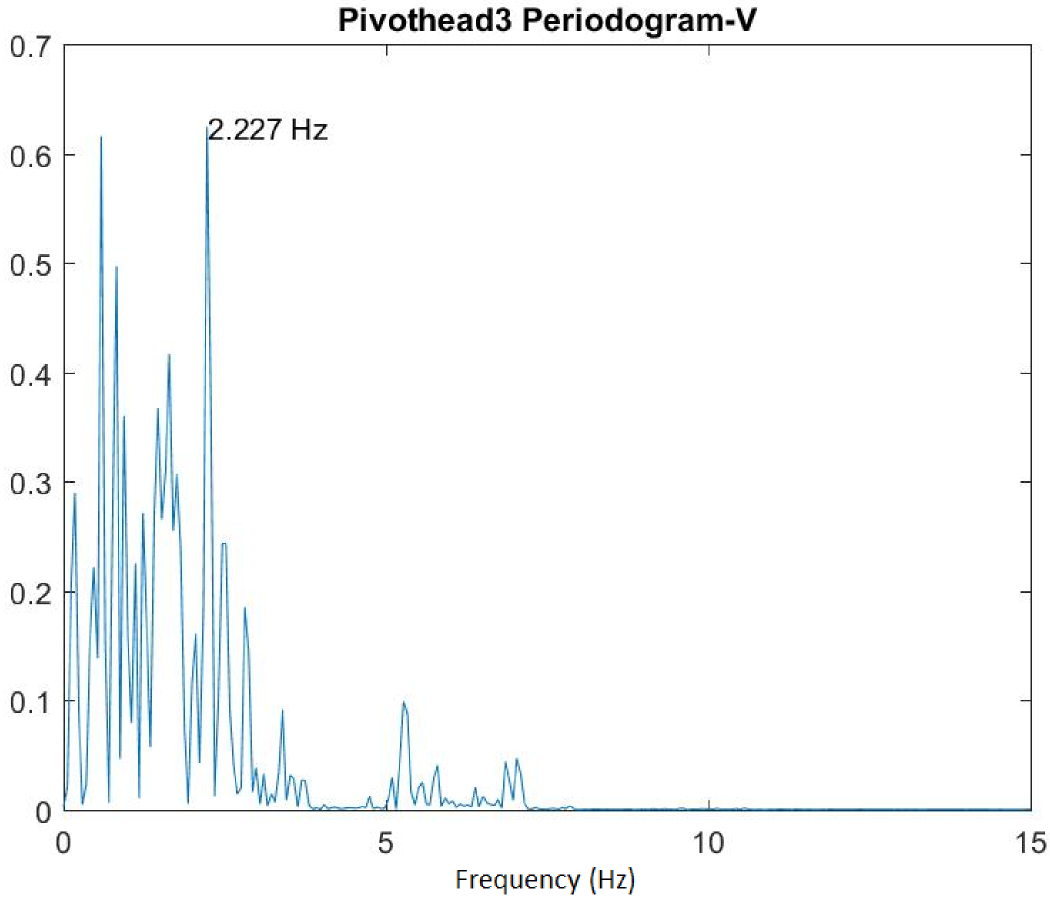

*Figure 7. Example of noisy periodogram that resulted from a bad collection*

All computed pairwise correlation results are presented in Table 3, and the definition of each discussed feature below is in Table 1. Figure 8 illustrates the axes of each device. The axes of the same color between the sensors correspond to the same physical direction of travel, though the sensors are worn on different parts of the body.

The first set of correlations computed were between channels within each sensor. For the camera, we find that the $s_U$ and $s_V$ features have an extremely strong correlation with each other (with significance at $\alpha$ =0.01). The features from the $Y$ and $Z$ channels of the accelerometer vest, $s_Y$ and $s_Z$ respectively, also have a strong correlation with each other, and are significant at the $\alpha = 0.01$ level. There is also a significant correlation at the $\alpha = 0.05$ level between the $s_X$ and $s_Y$ features. The feature computed from the $X$ channel of the accelerometer is not expected to be significantly correlated to the other features, since it represents the front/back acceleration of the subject, which is minimal when walking on a treadmill. This would not be the case for data collected from a gait activity where the subject was not on a treadmill. However, since the alignment of the axes in the accelerometer vest is not precisely measured, we cannot guarantee that the X-axis of the accelerometer is pointing exactly forward in the direction of travel. That is, there may still be some component of motion in the U- and V-axes of the camera data that correlates to the X-axis of the accelerometer, so we retain that feature in the dataset.

After finding several significant within-group correlations, we begin examining between-group (i.e. between-sensor) correlations. The results of correlating features between the two sensors indicate that there is significant correlation between $s_U$ and $s_Y$ at the $\alpha = 0.01$ level, and weaker but significant correlations between $s_V$ and $s_Y$, which are significant at the $\alpha = 0.05$ level. There also exists correlations between $s_U$ and $s_Z$, and $s_U$ and $s_X$, significant at the $\alpha = 0.05$ and $\alpha = 0.01$ levels respectively. Refer to Figure 88 and Table 1 for descriptions of these features.

Finally, we turn to examine correlations between each feature from each sensor and the “ground truth” feature – the $s_T$ derived from the recorded speed of the treadmill. A very strong correlation of 0.954 was found between $s_Y$ and $s_T$ as well as $s_Z$ and $s_T$, each significant at the $\alpha = 0.01$ level. A correlation also exists between $s_X$ and $s_T$ which is significant at the $\alpha = 0.05$ level, implying that all channels of the accelerometer have frequencies that are relatively well-correlated to the true speed of the gait. From the camera data, we find a correlation significant at $\alpha = 0.05$ between $s_U$ and $s_T$.

We find no significant correlation between the $s_V$ and $s_T$ features, so the frequency of the vertical motion in the camera data is not significantly correlated to the true speed of gait. We find that the features computed from the V channel tend to be noisier than the features from other channels, which hurts the linear correlation. A more robust outlier detection or signal filtering method may be able to remove such outliers and improve the strength of this correlation, since we do expect that the V channel feature would be correlated to the gait speed as the other features are (as shown in Table 3).

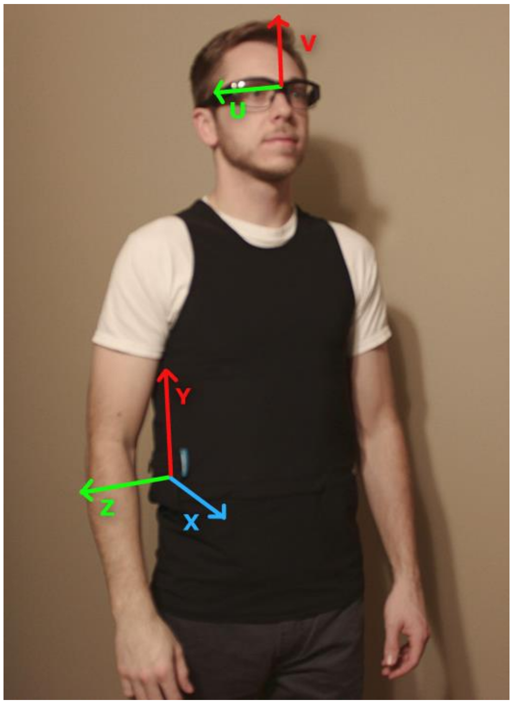

*Figure 8. Subject wearing Hexoskin vest and Pivothead camera glasses with axes overlaid*

*Table 3 – Correlations between computed features*

| | $s_T$ | $s_U$ | $s_V$ | $s_X$ | $s_Y$ | $s_Z$ |
|---|---|---|---|---|---|---|
| $s_T$ | 1 | | | | | |
| $s_U$ | 0.492424** | 1 | | | | |
| $s_V$ | 0.38651 | 0.970096*** | 1 | | | |
| $s_X$ | 0.415744** | 0.387766* | 0.340153 | 1 | | |
| $s_Y$ | 0.954*** | 0.554251*** | 0.49274** | 0.397917** | 1 | |
| $s_Z$ | 0.797043*** | 0.441839** | 0.373742 | 0.279735 | 0.846169*** | 1 |

* - significant at $\alpha = 0.1$<br>
** - significant at $\alpha = 0.05$<br>
*** - significant at $\alpha = 0.01$

# 5 DISCUSSION

In this section we discuss the findings, complications encountered, and the implications of the results documented in section 3.7.

## 5.1 Motion Extraction

When considering the performance of the three different motion extraction algorithms, it is clear that the Sparse Optical Flow produces the best results. In each of the cases, regardless of the aggregation method that was applied, the Sparse Optical Flow algorithm produced data with a lower SSE than each of the other methods. We attribute the superior performance to the targeted selection of features that provide a more accurate estimate of motion within the frame. Features are identified based on their ability to be found in consecutive frames, providing motion vector estimates with less variance. We recall that the assumption has been made that the motion extracted from the video is resultant from the motion of the subject wearing the camera device. The environment being recorded has been limited to otherwise static objects. Thus an ideal motion extraction will produce vectors for each feature that are identical. Due to noise and error in the detection methods, we expect some variance among the extracted vectors, but a method producing vectors with lower variance indicates a preferable result.

Optical Flow-based algorithms do not perform well in regions of video with uniform brightness, such as a blank wall. In these areas, features are difficult to detect from frame to frame, so the algorithm produces vectors that are very small or zero. For this reason, these points are avoided as features in Sparse Optical Flow. However, Dense Optical Flow includes every pixel, including these undesirable ones, in the set of features. When examining the motion vectors produced by the Dense Optical Flow method, it is apparent that the method suffers from finding too many of these near-zero vectors. This influences the distribution of the vectors and damages the results produced by all of the statistical aggregation methods with the Dense Optical Flow algorithm.

SURF matching provided the least consistent SSE across the aggregation methods. The best result from SURF matching was produced with the $min$ aggregation method, where it out-performed Dense Optical Flow, but still resulted in greater error than Sparse Optical Flow. However, the SSE values from SURF matching with the other aggregation methods had a much greater variance than the SSE values from the other motion detectors. While SURF matching does not suffer from the problem of too many features that degraded the performance of the Dense Optical Flow, the high amount of error when using the other aggregation methods reveals that the SURF matching has a tendency to over-estimate the amount of motion in the frame in comparison to the truth data. A visual inspection of the data supports this conclusion. Such large overestimations for certain image features in the video frame skewed the motion estimates produced by this method. The minimum vectors produced per-frame were the only vectors that were reasonably near to the truth values since the others were so large.

Since the Sparse Optical Flow method outperformed the other methods by such a large margin, it was used in all subsequent activities as the motion extraction method. The mean was used as the vector aggregation method. Even though the median aggregation method produced a lower SSE, it was only marginally better, and at the expense of computational complexity.

## 5.2 Waveform Parameter Extraction

We verify the similarity between the extracted motion vectors and the manually annotated truth data using several methods. A visual inspection of the data presented in Figure 4 shows that the optical flow produces a time series of motion components that is nearly identical to the annotated vectors. However, it is desirable to prove this relationship using quantitative methods.

### 5.2.1 Cross Correlation

We use cross-correlation to successfully prove that the two signals have similar content. The cross-correlation calculation is maximized at lag time $t = 0$, indicating that the two signals are most nearly correlated when they are aligned at the start. We identify that the data has a cyclic nature from the shape of the cross-correlation plot as well. The correlation varies from a positive local maximum to a negative local minimum in a periodic fashion. These phenomena are equally spaced across frames, supporting the claim that the two signals are periodic. The points corresponding to local maxima are points where the two signals are aligned and in phase, and the local minima are where the signals are misaligned and out of phase. Since the two signals match so nearly, the resulting chart is nearly symmetric about the line $\mathrm{t} = 0$, as seen in Figure 4.

### 5.2.2 Periodograms

Periodograms are regularly used to extract frequency information from signals. In our data collection routines, we fixed the speed of the participant to a constant. This means that we expect to identify one major frequency on which the signal is focused. While there is an expectation that other frequencies may have some content associated with them due to noise (perhaps due to the camera sensor), we expect this to be the minority case. A visual inspection of the periodograms shows agreement in each of the eight collected video segments (see Figure 5). In fact, in all cases, we confirm that the peak in the truth data occurs at the same frequency as the peak in the extracted motion vector data. We further confirm by manual analysis of the collected video that the identified frequencies represent the true frequency at which the gait occurred.

During the data collection described in section 3.2.1, four subjects walked at two different speeds. As seen in Figure 5, at 2.3 miles per hour, subjects A, B, and D had an identified walking pace of .793 Hz, which equates to one step every 0.63 seconds. At 3.9 mph, both subjects had an extracted stride frequency of 1.02 Hz, or one step every 0.49 seconds. At 2.3 miles per hour, subject C had a pace of 0.963 Hz, or one step every 0.52 second. The more frequent steps at the same speed are expected, possibly due to differences in the height of the subjects.

### 5.2.3 Coherence

While the identified peaks prove that the majority of frequency content in the computed and truth signals are similar, we attempt to quantify the degree to which the frequency data matches using the magnitude squared coherence (Figure 5). As noted earlier, this confirmed our expectations that the signals have matching frequency content with regard to the true frequency at which the gait was occurring.

The single frequency at which we expect the waveforms to match – the frequency at which the gait activity occurred – has a perfect coherence of 1 for each participant. We notice that some other frequencies also produce a high measure of coherence between the two waveforms. For example, the coherence plot for subject A (Figure 5) at 2.3 miles per hour shows a significant amount of coherence between the truth and generated data at just less than 2.5 Hz. A similar peak in coherence occurs just above 3 Hz for subject A at 3.9 mph. These are much quicker than the activity that was taking place, so these frequencies do not directly describe the rate at which the gait occurred. However, it may represent another cyclic movement that takes place within a single stride (since it was more frequent than the gait). More analysis would be required to determine the cause and whether the pattern may be generalized to multiple subjects.

## 5.3 Comparing Video to Accelerometer

The comparison methods in section 5.2 provide evidence that the motion vectors extracted by the optical flow algorithm produce similar content to vectors manually annotated by hand. Once this procedure was validated, we adopted the assumption that the results produced from subsequent data collections were also going to be valid. Rather than continue to verify the vectors produced by optical flow against truth data, we seek to correlate them to data from a more traditional accelerometer sensor. As referenced in section 1, the current body of work contains accelerometer-based approaches to measuring gait parameters with wearable devices.

### 5.3.1 Data Alignment

One of the most difficult challenges in comparing the video- and accelerometer-based data is aligning the data such that the samples from the devices are appropriately correlated in time. The two commercial devices used must be powered on separately, meaning that they are highly unlikely to begin recording at exactly the same moment. The video data conveniently provides the context of the surrounding seen in the video to temporally locate the time of the recording. The accelerometer, however, contains no information regarding the environment, which poses a problem.

#### 5.3.1.1 Temporal Alignment

The data collection process was designed in a way that aided the synchronization. The camera device was turned on first in each collection. Following this, the accelerometer was initialized in view of the camera. In this way, a rough approximation of the difference of start times of the two devices may be computed.

The approximation gives some idea of the initial offset of the two data recordings, but does not provide enough precision. To get a more precise alignment, we perform a second physical marker in the data to produce a unique waveform in both the accelerometer and the video motion. Prior to beginning a gait activity, the subject would complete a brief aperiodic movement, followed by at least three seconds of no movement. The gait activity then began immediately following the period of no movement. In the extracted waveforms, this created a unique signature of high amplitude followed by nearly zero amplitude, followed by periodic movement of gait.

#### 5.3.1.2 Accounting for Sampling Rate

Another difficulty with aligning the data is that the two devices have different sampling rates. The frame rate of the video is 30 frames per second, and the sampling rate of the accelerometer is 64 Hz. This leaves the option of down-sampling the accelerometer or up-sampling the video. We decided to up-sample the video in order to get matching frame rates which is necessary for the frequency-based calculations of correlation and coherence. Down-sampling was not used on the accelerometer because the final rate would have been less than half of the original, resulting in loss of information.

### 5.3.2 Frequency Content

A comparison of frequency content between the camera and accelerometer data reveals that the data from each device produced a waveform with identical frequencies. It is clear from the periodograms in Figure 6 that the measurements from the two sensors have distinct errors and the data must be smoothed before comparing between the camera and accelerometer. The unsmoothed accelerometer data does not appear to contain the frequency component of the walking activity as its primary feature. Instead, a much more frequent pattern is prominent in the periodogram. However, as the data is smoothed with a large enough window, the higher frequencies are eliminated from the signal, leaving only the truly significant frequency from the given activity. In this case, the higher frequency noise is likely due to the sensitivity of the accelerometer device to very slight movements inside the pocket of the vest. A window size of $w = 16$ (one fourth of the sampling rate of the device) appears to give the best results for isolating the desired frequency and was used in subsequent comparisons of the accelerometer data to the camera data.

## 5.4 Correlating Walk Frequency to Speed

The correlation of computed features provides insight into the data being collected by each sensor. The strong within-group correlations (correlations between channels from the same sensor) are expected, but confirm that there is a strong relationship between the correlated axes of motion during the activity. That is, there is a strong linear relationship between the movement captured by each sensor in the horizontal and vertical planes during the walking activity.

The between-group relationships (correlations between channels from different sensors) were not as strong as the within-group relationships. This should also be expected since the sensors are worn on different parts of the body. However, there is still a moderate correlation between data from the two sensors. This result supports the conclusion that the camera sensor and optical flow method can produce similar (correlated) data to the accelerometer for the gait activity. Since related work often involves the use of accelerometer sensors, finding that camera data is correlated provides support for applying existing techniques across sensor types.

The correlation between computed features and the actual true walking speed of the subject is another important result. We find that the most prominent frequencies detected in all channels of data being collected are correlated to the true walking speed, except for the $V$-channel from the camera glasses. As discussed in Section 4.4, this result could be improved with additional outlier detection or signal filtering. For example, some of the outlier frequencies fall into ranges that are not realistic for the activity being performed, such as a stride frequency of less than 0.25 Hz (one stride every 4 seconds). These frequencies could easily be eliminated with a bandpass filter on appropriate frequencies determined by the activity under investigation such as walking or running.

Our strong correlation results  provide validation that the features being computed (and the methods used to compute them) are providing useful information for the task of estimating the walking speed. That is, we confirm that the estimated frequency of the activity and the estimation of stride length are in fact providing information related to the true gait parameter of speed.

Through this research, we are seeking to determine whether the data between the sensors is significantly correlated to allow for one sensor to take the place of the other. We find similar elimination of sensors based on correlation with promising results in [19]. Minimizing the number of sensors in our approach provides a simpler and less expensive solution, and improves the usability for applications with the targeted demographics, which include the elder-care and rehabilitation communities. In the case of our devices, there is much more existing research on accelerometer-based methods, making the accelerometer a reliable choice. However, the camera sensor is most convenient for a subject to put on and use, so it would be desirable to eliminate the accelerometer vest from our approach. We expect from our findings that we should receive acceptable results from applying existing methods to the camera data since the camera data is significantly correlated to the accelerometer data.

In the general case, our conclusion is that the data between the sensors is strongly correlated enough to allow for one sensor to replace the other. A system which retains both the accelerometer and camera sensor, may provide richer data than a system with just one of the sensors. Since the sensing modalities differ, they are prone to differing types of noise under different circumstances. In the event that one sensor provides a bad dataset, the other sensor may be used to fill in a temporal gap in the data, making the system more robust than a system with a single sensor or even with multiple sensors of the same type.

# 6 CONCLUSION

In this work we had the goal of identifying parameters that describe a gait activity using a wearable camera and accelerometer sensors. Our instrumentation is non-invasive and lightweight, providing convenience to the subject. We performed two data collection events in controlled environments – one with subjects wearing only the camera, and one with subjects wearing the camera and accelerometer vest simultaneously.

Using sparse optical flow methods, statistical aggregation, and signal processing techniques, we are able to identify frequencies of motion that occur during a subject's gait from video data alone. We also have success in correlating the frequencies of motion in the horizontal plane (perpendicular to the ground) to the true frequency at which the subject was walking.

Building upon these results, we perform similar signal processing techniques to identify frequencies in data collected from the accelerometer vests. We find significant correlations between the data derived from 1) channels within each sensor, 2) channels between the two sensors, and 3) channels from each sensor and the true walking speed.

Our results indicate that it is possible to perform parametric gait analysis via a commercial wearable camera and vest which are very simple to use. Our method is currently limited to controlled environments, but the devices also provide an easily portable hardware configuration.

# 7 FUTURE WORK

The following sections detail additional work that could be conducted in the future to advance or improve upon the results found and described herein.

## 7.1 Environmental Variability

Each of the described data collection events were conducted on a treadmill in a room with a static background and consistent lighting conditions to minimize errors in motion detection. However, the downside to this decision is that it effectively removes a dimension of movement by keeping the subject stationary. Therefore, removing this restriction will improve the portability of the system and potentially improve the results of the method.

To remove the environmental restrictions requires having a method of removing unwanted objects from consideration when computing motion vectors. The current method is based on the assumption that the majority of motion on the frame may be attributed to the motion of the subject. If other objects in view have motion independent of the subject’s motion, this will invalidate the assumption. To correct for this, objects that are not part of the background need to be detected and removed. We save the exercise of detecting these objects and ignoring them during motion analysis for a future project.

## 7.2 Kinematic Model of Gait

The work described herein extracts gait parameters from collected data and compares the extracted parameters to truth data. While the methods produce the desired result, the collected data does not provide a complete model of gait. A limitation of the camera and accelerometer is that they do not provide such truth data against which a comparison may be formed. The data provides little or no external view of the subject. For example, in the collected data, we would expect to find a strong correlation between subject height and the frequency of steps at a fixed speed (having shorter legs requires more frequent steps than having longer legs), but unfortunately did not collect this data from the participants and no such measure can be drawn from the data. We were also unable to explain the weaker frequencies detected in the video (discussed in section 4.2.3). This would require richer data on the pose of the subject during the activity, as well as understanding the noise sources of the video and accelerometer data.

A future data collection is scheduled to provide a full gait model via a motion capture system that collects video from 21 time-synchronized cameras. This system will provide a detailed model of major joint movement during gait, and a much richer truth dataset against which parameters may be extracted.

# 8 REFERENCES

[1] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *IEEE Conference on Computer Vision and Pattern Recognition,* 2012.

[2] Y. Watanabe, T. Hatanaka, T. Komuro and M. Ishikawa, "Human gait estimation using a wearable camera," in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2011.

[3] T. Banerjee, J. M. Keller, M. Popescu and M. Skubic, "Recognizing complex instrumental activities of daily living using scene information and fuzzy logic," *Computer Vision and Image Understanding,* vol. 140, pp. 68-82, 2015.

[4] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz and M. Aud, "Linguistic summarization of video for fall detectio nusing voxel person and fuzzy logic," *Computer Vision and Image Understanding,* vol. 113, no. 1, pp. 80-89, 2009.

[5] Y. Cho, Y. Nam, Y. J. Choi and W. D. Cho, "Smartbuckle: human activity recognition using a 3-ais accelerometer and a wearable camera," in *Proceedings*

*of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, 2008.

[6] E. Hirasaki, S. T. Moore, T. Raphan and B. Cohen, "Effects of walking velocity on vertical head and body movements during locomotion," *Experimental Brain Research,* vol. 127, no. 2, pp. 117-130, 1999.

[7] M. Unuma, K. Anjyo and R. Takeuchi, "Fourier principles for emotion-based human figure animation," in *Proceedings of the 22nd annual conference on computer graphics and interactive techniques,* 1995.

[8] Pivothead, [Online]. Available: http://www.pivothead.com. [Accessed 25 September 2017].

[9] T. Banerjee, M. Peterson, Q. Oliver, A. Froehle and L. Lawhorne, "Validating a Commercial Device for Continuous Activity Measurement in the Older Adult Population," *Smart Health*, accepted.

[10] E. Martin, V. Shia and R. Bajcsy, "Determination of a patient's speed and stride length minimizing hardware requirements," in *Proceedings of 2011 International Conference on Body Sensor Networks*, 2011.

[11] S. Chen and J. Lach, "Nonlinear Feature for Gait Speed Estimation using Inertial Sensors," in *Proceedings of the 8th international Conference on Body Area Networks*, 2013.

[12] "Hexoskin Wearable Body Metrics," [Online]. Available: https://www.hexoskin.com/. [Accessed 9 November 2017].

[13] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the 13th Scandinavian conference on image analysis,* 2003.

[14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[15] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAID),* 1981.

[16] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "SURF: Speeded up robust features," *Computer Vision and Image Understanding (CVIU),* vol. 110, no. 3, pp. 346-359, 2008.

[17] A. R. Pratama, Widyawan and R. Hidayat, "Smartphone-based Pedestrian Dead Reckoning as an Indoor Position System," *International Conference on System Engineering and Technology,* pp. 1-6, 2012.

[18] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE,* vol. 37, no. 1, pp. 10-21, 1949.

[19] N. Jalloul, F. Poree, G. Viardot, P. L'Hosts and G. Carrault, "Activity Recognition using Complex Network Analysis," *IEEE Journal of Biomedical and Health Informatics,* vol. PP, no. 99.