Evaluation of an Inexpensive Depth Camera for In-Home Gait Assessment

Erik Stone*, and Marjorie Skubic
Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA

Abstract. An investigation of a new, inexpensive depth camera device, the Microsoft Kinect, for passive gait assessment in home environments is presented. In order to allow older adults to safely continue living in independent settings as they age, the ability to assess their risk of falling, along with detecting the early onset of illness and functional decline, is essential. Daily measurements of temporal and spatial gait parameters would greatly facilitate such an assessment. Ideally, these measurements would be obtained passively, in normal daily activity, without the need for wearable devices or expensive equipment. In this work, the use of the inexpensive Microsoft Kinect for obtaining measurements of temporal and spatial gait parameters is evaluated against an existing web-camera based system, along with a Vicon marker-based motion capture system for ground truth. Techniques for extracting gait parameters from the Kinect data are described, as well as the potential advantages of the Kinect over the web-camera system for passive, in-home gait assessment.

Keywords: fall risk, gait, Kinect, smart environments, eldercare

1. Introduction

To allow older adults to continue living longer in independent settings, and thus reduce the need for expensive care facilities, low-cost systems are needed to detect not only adverse events such as falls but to assess the risk of such events, in addition to the early onset of illness and functional decline. Continuous, ongoing assessments of physical function would help older adults live more safely in independent settings, while also facilitating targeted medical interventions when needed. Ideally, such measurements would be obtained passively, in the course of normal daily activity [7]. This work focuses on developing a robust, low-cost, vision based monitoring system for measuring gait in a home environment, with the broader goals of assessing fall risk and detecting the early onset of illness and functional decline.

Research has shown the importance of measuring a person's gait [16] and that the parameters which describe locomotion are indispensible in the diagnosis of frailty and fall risk [29]. Studies have identified stride-to-stride variability as a predictor of falls [14,22], suggested that gait velocity may be predictive of adverse events in well functioning older adults [25], and have shown gait velocity slows prior to cognitive impairment [5].

Vision-based monitoring systems have the resolution needed to yield the detailed measurements of physical function used for typical fall risk assessment protocols, early illness detection, etc., passively, in the home environment on a continuous basis. Furthermore, research has shown that the privacy concerns of older adults to video-based monitoring systems can be alleviated through appropriate handling and processing of the video data, e.g., in the form of silhouettes [8].

Recently, Microsoft has released a new, inexpensive sensor device, called the Kinect, to allow controller free game play on their Xbox system. The device uses a pattern of actively emitted infrared light to produce a depth image. That is, the value of each pixel in the image depends on the distance from the device of what is being viewed. Furthermore, the depth image is invariant to visible lighting. This technology allows for a three-dimensional (3D) representation using a single, relatively inexpensive Kinect sensor. Additionally, this depth camera tech-

* Corresponding author. E-mail: ees6c6@mizzou.edu

nology offers a number of potential benefits for low-cost, vision based monitoring systems.

This paper presents a preliminary investigation, first described in [32], along with a more detailed evaluation of the Kinect as a sensor for passive, in-home gait measurement. First, a brief discussion of related work is presented. Next, techniques for acquiring spatial and temporal gait parameters from the depth data of the Kinect are described, along with a comparison of the measurements obtained from the Kinect to those obtained from an existing web-camera based system, and a Vicon marker-based motion capture system. Potential advantages of the Kinect over traditional camera systems are also presented. Finally, the major points of this paper are discussed along with future work.

2. Related Work

Recent research in activity monitoring of older adults has focused on the use of passive infrared (PIR) motion sensor suites in the home [15,34]. These sensor suites yield information about the daily activity levels of monitored subjects, and arrays of such sensors have been used to obtain velocity measurements on a continuous basis in home settings [2,11]. While such systems do not raise privacy concerns among older adults, they typically do not, due to the coarse nature of the PIR sensors, produce measurements of the detail necessary for the assessment of fall risk, specifically, spatial and temporal gait parameters beyond walking speed (e.g., step time, step length, gait symmetry), timed up and go (TUG) time, sit to stand time, etc [10]. Existing systems for capturing such measurements are typically wearable, accelerometer-based devices, expensive gait or marker-based motion capture systems, or direct assessment by a health care professional (typically, with a stop watch) [16].

Wearable accelerometer-based devices for obtaining detailed measurements of physical activity, specifically gait parameters, is an area that has been the focus of much research [9]. Efforts have even included utilizing accelerometers in existing smart devices, which individuals may already own and potentially carry with them. However, many elderly adults are reluctant to use wearable devices because they consider them to be invasive or inconvenient, especially during times when they are not feeling well [7]. Furthermore, wearable devices generally require active involvement on the part of the user for putting the device on, taking it off, charging batteries, etc. Although wearable devices offer the distinct advantage of measurements outside the home, they may not be as reliable for daily monitoring and assessment as passive, environmentally mounted sensors in the home.

Human motion analysis using vision technology is another widely researched area that has been applied to gait assessment, with two basic approaches: marker and marker-less. Marker-based systems detect markers attached to a subject's body in multiple camera views. Given the location of the markers in different camera views, the 3D position of the marker can be obtained. The use of markers helps to yield highly accurate and robust measurements of a person's motion. Obviously, however, marker-based motion capture systems are not practical for in-home, continuous monitoring.

A large amount of work has been done regarding marker-less video-based motion capture systems. Marker-less video based motion capture systems generally work by extracting the silhouette of the subject in multiple, calibrated camera views, projecting the silhouette from each of the views into a discretized volume space, and fitting a skeletal model to the intersection formed by the projection of the silhouettes in the discretized volume space [4,24]. Such systems have been shown to yield excellent results. However, they are typically expensive, computationally intensive, require a controlled environment, and require high quality and/or a large number of cameras, attributes which limit their suitability for in-home activity monitoring.

A number of researchers have looked at using systems composed of a single or multiple cameras for monitoring purposes in non-laboratory settings [13,17,19,20,23,27,35,36]. For example, in [13], a single camera system is developed that can identify and track people in outdoor environments. The system can estimate body posture (e.g., standing, sitting, lying, etc.) as well as track six primary body parts: head, torso, feet, and hands. In [36], a single camera system is used to estimate the location of people in an indoor environment, along with moving speed. This information is then converted into Activity of Daily Living (ADL) statistics for meaningful summarizations of activity. Lastly, in [17], researchers make use of a time-of-flight (TOF) based depth camera to identify and track people in a living room setting. The height of the person is used to classify their activity as standing, sitting, or lying. None of these systems specifically looked at obtaining gait parameters in the home.

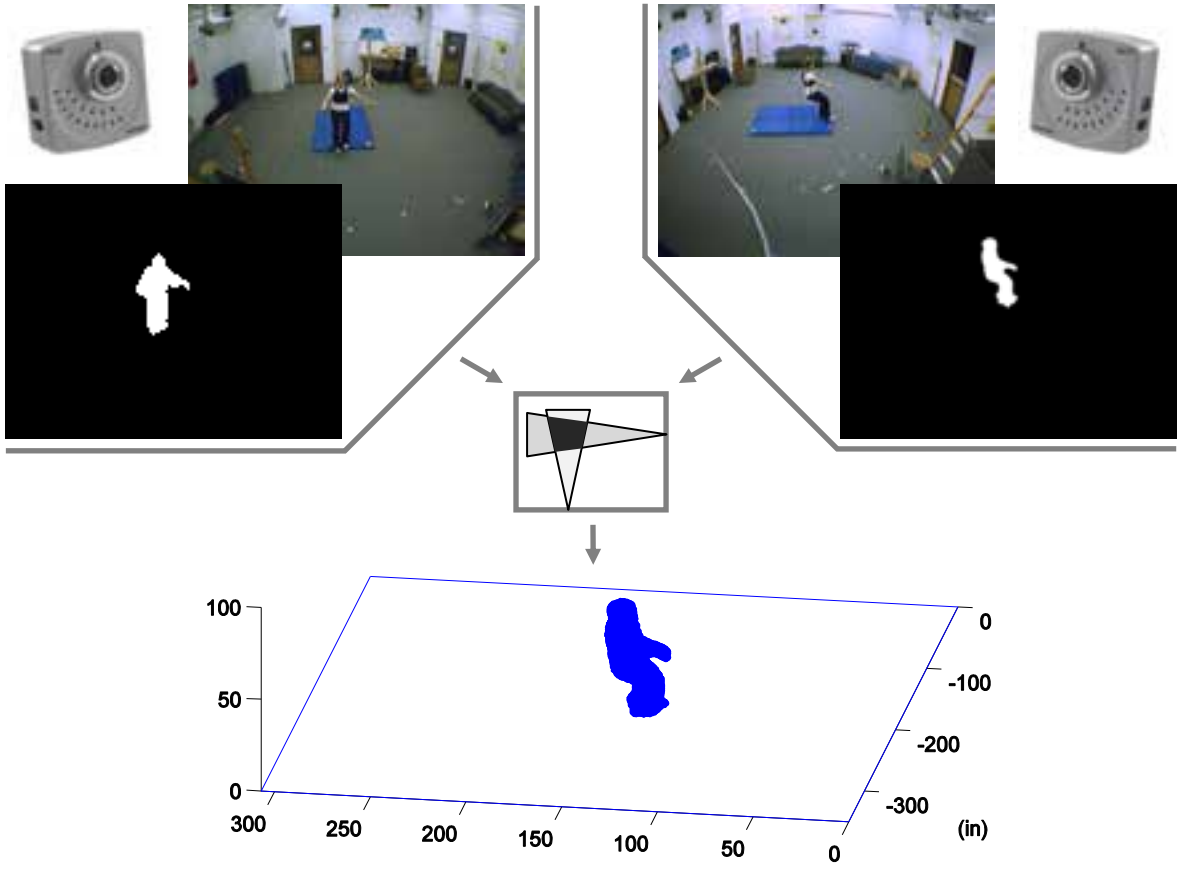

Fig. 1. Web-camera system. Two calibrated cameras positioned orthogonally in the environment capture images of the scene. Silhouettes are extracted from the captured images. The silhouettes are projected into a discretized volume space. The intersection of the silhouettes in volume space forms a three-dimensional (3D) representation of the person for tracking and analysis.

3. Systems

3.1. Vicon

The Vicon system is a highly accurate marker based motion capture system used in a variety of animation, life sciences, and engineering applications [26]. The system outputs 3D coordinates of detected markers at 100 frames per second. For this work, it serves to provide ground truth data for comparison purposes.

3.2. Web-Camera

The web-camera based system, outlined in Figure 1, consists of two inexpensive web-cameras, positioned roughly orthogonal, monitoring the environment. Silhouettes are extracted from captured images using a background subtraction technique based on color and texture features. Details of the background subtraction algorithm for extracting silhouettes are described in [21], while the procedure for updating the background models and dealing with other issues encountered in dynamic, noisy environments is described in [31]. The updating procedure makes use of both pixel (2D) and voxel (3D) data.

Briefly, the system operates as follows. First, intrinsic and extrinsic calibration parameters for both cameras are obtained *a priori*, allowing for a three-dimensional representation to be formed in a discretized volume space as the intersection of the projection of the silhouettes from each camera. Typically, the space is discretized into one inch (2.54 cm) non-overlapping, cubic elements (voxels). The system runs real-time at five to fifteen frames per second depending on available computing resources. For the

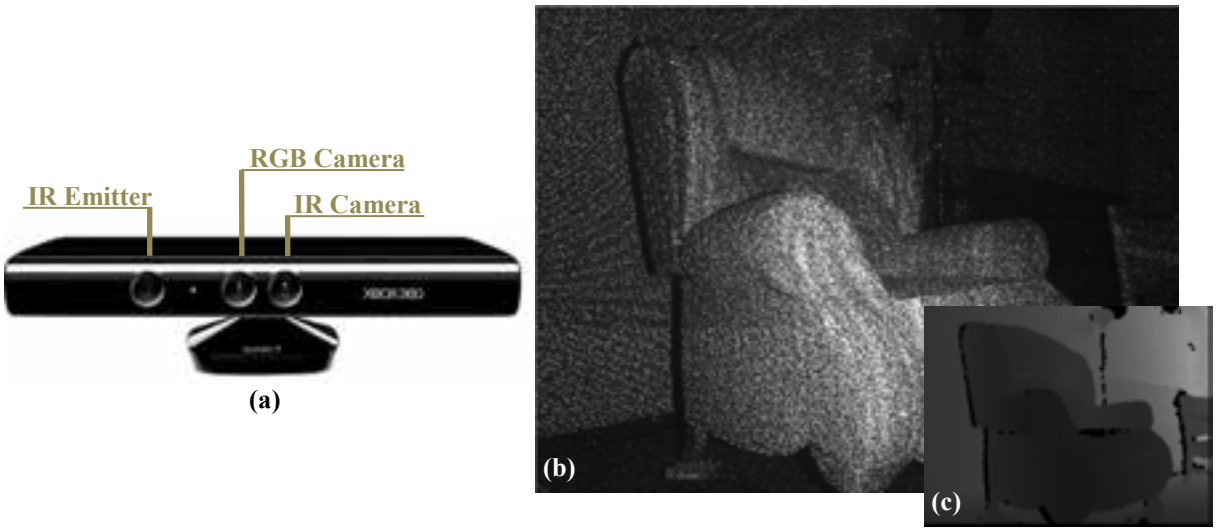

Fig. 2. (a) Microsoft Kinect sensor. (b) Raw image from infrared (IR) camera of Kinect showing actively emitted IR pattern as projected on a recliner chair. (c) Resulting depth image. Darker pixels are closer to the device, lighter pixels are further away. Black pixels indicate no depth value was returned.

experiments collected in this work, the system was run at five frames per second.

The web-camera system has been evaluated for fall detection, gait measurement, body sway measurement, and sit-to-stand measurement [1,3,30,33] with good results. The goal is to develop a passive, in-home, low cost, activity monitoring system for elderly adults that is capable of both detecting falls and helping to assess fall risk through the extraction of various physical parameters, including gait.

Currently, the system has been deployed in five apartments of an elderly independent living facility with plans to deploy five additional systems. The deployed systems are actively extracting walking sequences, automatically, in real-time. The systems will be evaluated over a two year period.

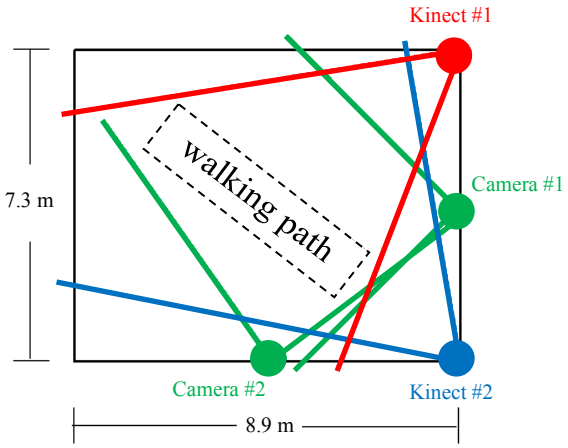

Fig. 3. Positioning of web-cameras, Kinects, and walking path in test environment. Walking path is approximately 17 ft. (5.2 m) in length. Lines show field of view for each device.

3.3. Kinect

The Kinect [28], Figure 2 (a), released by Microsoft in North America on November 4, 2010, was designed to allow controller-free game play on the Microsoft Xbox. The device makes use of technology developed by the Israeli company PrimeSense, and contains both an RGB camera, and an infrared (IR) sensitive camera, from which a depth image can be produced based on a pattern of projected infrared light. This pattern is shown in Figure 2 (b) projected onto a recliner chair.

The depth data returned from the device (at 30 frames per second) is an 11-bit 640x480 image which is invariant to visible lighting. The precision of the distance measurment for each pixel is dependent on the distance from the Kinect, with the precision decreasing from approximately one centimeter at two meters to approximately ten centimeters at six meters. The minimum range of the device is approximately one meter. For use with the Microsoft Xbox, it is recommended that the user be approximately two meters from the device.

When used with the Xbox, the Kinect allows 3D motion tracking using a skeletal model, gesture recognition, facial recognition, and voice recognition. Following its release, Linux and windows drivers were developed, and the Kinect has been used for a variety of purposes from entertainment to robotics [28].

3.4. Layout

For the experiments conducted in this work two web-cameras (forming the web-camera based system) and two Kinects (operating individually) were positioned in a laboratory environment, alongside a Vicon motion capture system. Figure 3 shows the approximate placement of the web-cameras, Kinects, and the location of the walking path in the test environment. The cameras and Kinects were positioned approximately eight feet (2.4m) off the ground, and the Kinects were angled downward approximately 20 degrees from horizontal. One Kinect was placed inline with the walking path while the other was placed orthogonal in order to evaluate the impact of the positioning on the accuracy of the extracted gait parameters.

4. Methodology

In this section, the techniques used to extract the gait parameters of walking speed, stride time, and stride length from the 3D depth data returned by a single Kinect are described. For the purpose of evaluation, the assumption has been made that there is only one person in the scene at a time, and that the environment is stationary. Thus, the 3D point cloud obtained from the Kinect is for a single person. For use in a real world, dynamic environment, a high-level tracking algorithm and model updating procedure would be necessary to achieve this, analogous to what has been done with the web-camera system [31]. Finally, the extrinsic parameters of the Kinect with respect to the room have been computed, thus allowing accurate estimation of height with respect to the floor.

For the web-camera based system, an existing algorithm was used to extract gait parameters [30]. In short, the algorithm attempts to identify footfall locations for a walking sequence by finding the positions at which the feet are stationary. Given the locations of the footfalls, and the times at which they occurred, gait parameters can then be computed.

4.1. Kinect – Calibration

The first step of obtaining accurate spatial parameters from the Kinect is calibration. This consists of two steps. First, as with traditional cameras, intrinsic, distortion, and stereo parameters for the IR and RGB cameras on the Kinect are estimated using a standard checkerboard calibration pattern and supplemental IR backlighting.

Second, calibration of the depth values returned from the Kinect is performed. The depth data returned from the Kinect must be transformed to obtain usable and accurate distances. For this work, the following equations, based on [18], were used to transform a raw Kinect depth value, D, an integer value typically in the range [660, 1065], for a given pixel, (x, y), to a distance, d:

$$d = \frac{b}{f - D'} \tag{1}$$

$$D' = D(1 + k_1 r + k_2 r^2) + k_3 x' + k_4 y' \tag{2}$$

$$r = \sqrt{(x')^2 + (y')^2} \tag{3}$$

where x' and y' are the normalized pixel coordinates computed using the intrinsic and distortion parameters of the IR camera. The parameters b, f, k_1, k_2, k_3, and k_4 are optimized over a large (~3,000) set of training points and the equation attempts to adjust for distortion effects. The training points are obtained by placing a large checkerboard calibration pattern in the environment, while moving the Kinect over a large range of distances and viewing angles with respect to the pattern. Using the known intrinsic parameters of the IR camera, the position of the calibration pattern with respect to the camera in each frame can be estimated. Simultaneously, the values associated with the pattern in the depth image can be recorded. Following collection of the training data, a global optimization is performed using the CMA-ES algorithm [12]. Example values for the parameters $\{b, f, k_1, k_2, k_3, k_4\}$ used to transform the raw depth values to inches are {14145.6, 1100.1, 0.027, -0.014, 1.161, 3.719}.

As stated in Section 3.3, the precision of the distance measurements decreases as a function of distance. For example, using the above parameters and

equations (1-3), a change in raw depth value from 900 to 901 for the center image pixel ($x' = y' = 0$) corresponds to a change in distance from 70.69 to 71.04 inches (179.55 to 180.44 cm). However, a change in depth value from 1060 to 1061 corresponds to a change in distance from 352.76 to 361.78 inches (896.01 to 918.92 cm).

4.2. Kinect – Foreground Extraction

Foreground extraction is performed on the raw depth images from the Kinect using a simple background subtraction algorithm. Specifically, a set of background training images is captured over which the minimum and maximum depth values for each pixel are stored to form a background model.

For a new frame, each pixel is compared against the range stored for that pixel in the background model and those pixels whose raw depth value lies outside the range by greater than a threshold, T, are considered foreground. For this work, T was adjusted as follows based on the number of available background training frames, B:

$$T = \begin{cases} 1 & if\ B \geq 600 \\ 2 & if\ 600 > B \geq 300 \\ 3 & else \end{cases} \quad (4)$$

As the unit of T is that of the raw depth values from the Kinect, the actual distance the threshold corresponds to will vary based on the distance from the Kinect. For example, given the values at the end of Section 4.1, T=1 corresponds to a distance of approximately 0.35 inches (0.89 cm) at two meters, but a distance of approximately 9.0 inches (22.9 cm) at nine meters. Additionally, as the number of available background training frames decreases, the accuracy of the background model generally decreases as well. Therefore, the threshold is raised as the number of background training frames decreases to help suppress erroneous foreground classifications. The threshold levels shown in (4) were chosen based upon experimentation.

Following this initial pixel classification, a block based filtering algorithm is run to reduce noise, and smoothing is applied to the depth values identified as foreground. Example foreground extractions are shown in Figure 4. Given the foreground extraction for a frame, and the intrinsic, extrinsic, and depth

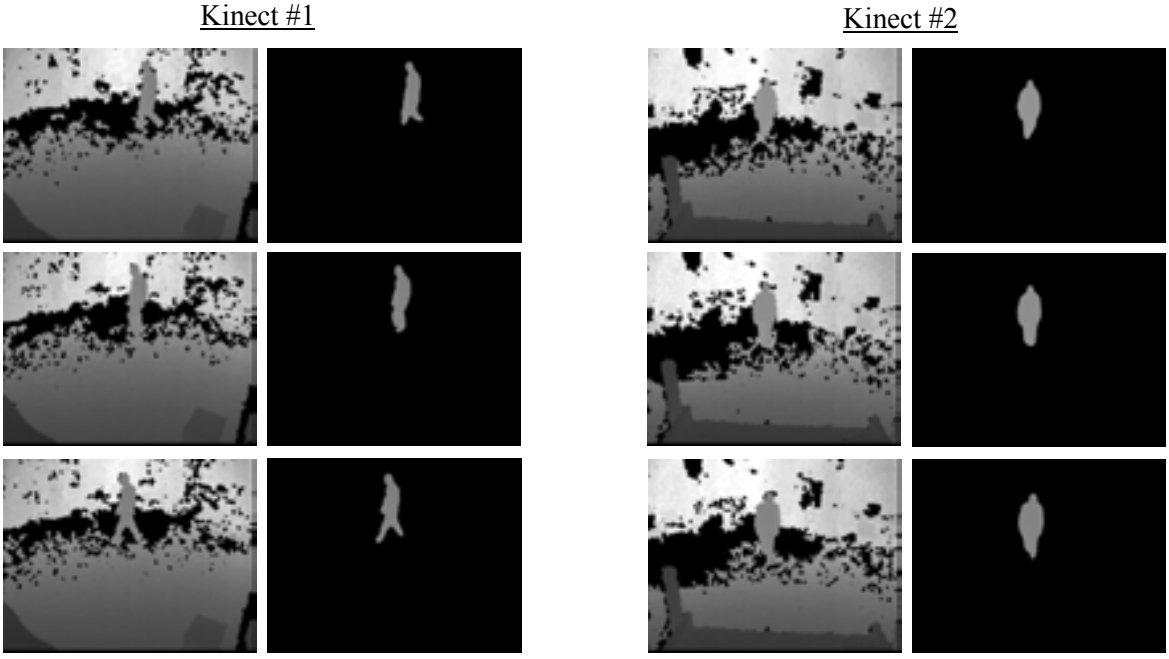

Fig. 4. Sample foreground extraction results from Kinect #1 and #2. ***Left*****:** Depth image from the Kinect. Darker pixels are closer to the device, lighter pixels are further away. Black pixels indicate no depth value was returned. ***Right*****:** extracted foreground.

calibration parameters for the Kinect, a 3D point cloud representation of the person can be obtained, as shown in Figure 5 (a).

This simple foreground extraction technique has proven to be quite robust and easily runs at the 30 frames per second rate with which depth data is received from the Kinect. In a dynamic, real world setting, background adaptation to handle non-stationary environments would need to be addressed. However, the invariance of the depth image to changes in ambient lighting addresses one of the significant issues affecting foreground extraction on color imagery. Furthermore, the computation required for foreground extraction on the depth data is minimal compared to that required for robust foreground extraction from color imagery, where a combination of texture and color features must be used, along with a graphics processing unit (GPU) in order to operate in real-time [21].

4.3. Kinect – Gait Parameters

Much of the work with the Kinect has focused on human body tracking using high degree of freedom skeletal models. Though such techniques are quite powerful, and may be essential to extracting some physical parameters, they often suffer from problems of instability, especially with noisy data. As with our web-camera based system, to facilitate capture in unstructured, noisy environments with no special requirements of the resident, we have opted to use techniques not based on skeletal models for extracting gait parameters.

For the results presented in Section 5 of this work, walking speed was estimated for a walking sequence using the following procedure:

1. project the centroids of the 3D point clouds for each frame onto the ground plane (e.g., dropping the vertical component of the computed centroid locations);
2. apply a moving average filter to the time series of the centroids (to smooth the data points

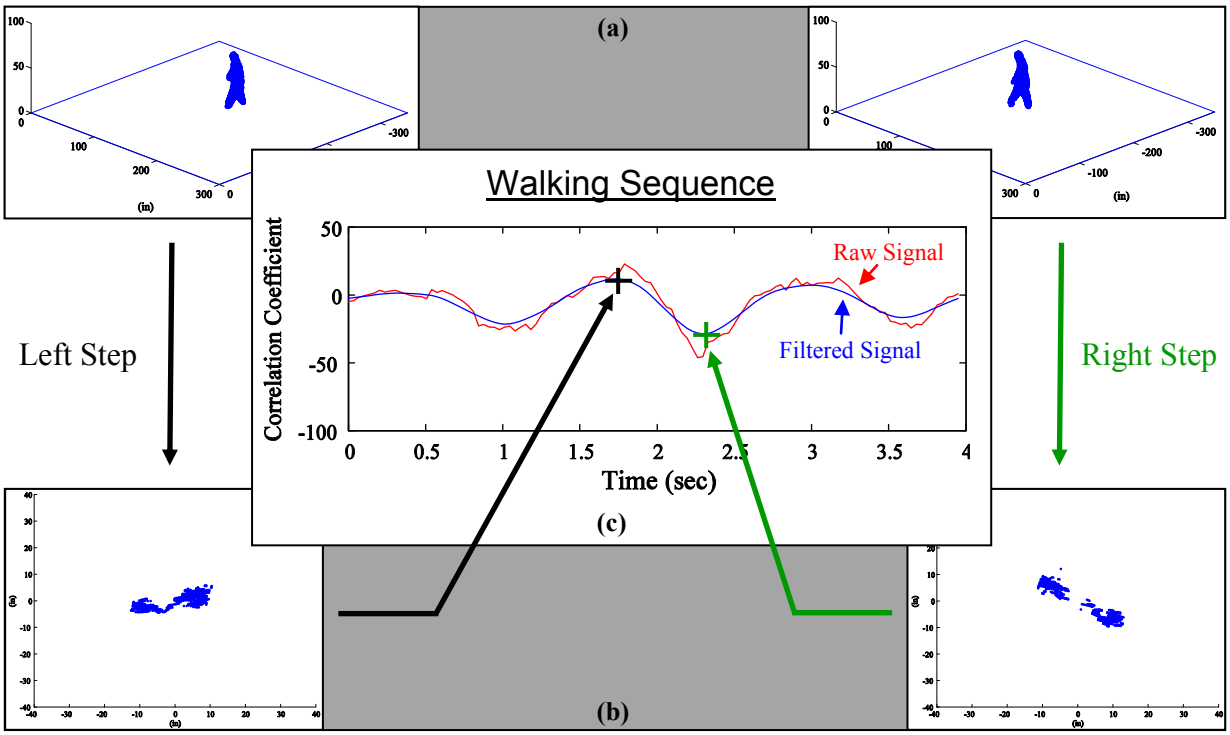

Fig. 5. Kinect gait extraction. (a) Three dimensional point cloud of a person for a single frame during a left step (left), and right step (right). (b) Normalized ground plane projections of points below 20 inches (50 cm) for the frames. (c) Left and right steps are detected as local maxima and minima, respectively, in the correlation coefficient time series of the walking sequence.

and reduce noise);
3. sum the frame-to-frame changes in position to calculate distance traveled (computed using the filtered centroid time series);
4. divide distance traveled by elapsed time.

For the results presented in Section 6 of this work, walking speed was estimated slightly differently to better match the definition in [6]. Specifically, the following procedure was used:

1. project the centroids of the 3D point clouds for each frame onto the ground plane;
2. apply a median and a moving average filter to the time series of the centroids (to smooth the data points and reduce noise);
3. calculate a best fit line to represent the direction of travel for the walking sequence;
4. compute the distance traveled for the walking sequence as the change in position along the direction of travel from beginning to end;
5. divide distance traveled by elapsed time.

This change is method, purely for agreement with the definition in [6], should have a minimal impact on the resulting accuracy as compared to the marker-based Vicon motion capture system due to the fact that the same methodology used for the Kinect data is applied to the data from the Vicon. Furthermore, the introduction of the median filter in step 2 is simply to further reduce the effect of any noise present in the data. The impact of this should be minimal given the controlled nature of the dataset.

For both Section 5 and 6, the number of steps and temporal gait parameters are estimated using only those 3D points with a height below 20 inches (50 cm). In previous algorithms developed for the web-camera based system, only those voxel elements with a height of 4 inches (10 cm) or less were used for computing these parameters. However, due to the fact that the foreground extraction algorithm operates on a depth image, points from the person that are close to the ground (and thus quite similar to the background model) are not extracted as foreground. Moreover, the distance at which a point is considered close to the ground depends on the distance from the Kinect, as the measurement precision decreases as the distance from the Kinect increases. Therefore, as participants are approximately 26.6 ft (8.1 m) from a Kinect at some point during each walking sequence, points significantly higher than the ground must be used in order to obtain any information about the lower extremities of the body.

First, points below 20 inches (50 cm) are projected onto the ground plane. Second, the projection is normalized by subtracting the mean, and rotating based on the localized walking direction. The localized walking direction, which is different from the direction of travel for the entire walking sequence, is computed for each frame in the sequence based on the change in position of the centroid over a window centered on the frame.

Given the normalized projection, containing N points, the following correlation coefficient is computed:

$$p = \frac{\sum_{n=1}^{N} x_n y_n}{N} \tag{5}$$

where x_n and y_n correspond to the X and Y coordinates of the n^{th} point in the projection.

The number of right and left steps for a walking sequence is obtained from the time series of the correlation coefficient. First, the correlation coefficient time series is filtered using a median and a moving average filter, both with a window size given by:

$$w = \frac{f * k}{v} \tag{6}$$

where v is walking speed, f is frame rate (for Kinect, 30 fps), and k is a constant parameter (16.6 was used), although it could be adapted based on the estimated height of the person. Finally, the signal is filtered using a moving average filter with a small, constant window size to remove any remaining minor local extrema. From the filtered signal, right steps are detected as local minima, while left steps are detected as local maxima.

Following footstep extraction, a series of simple heuristic rules are used to verify that the extracted left and right footsteps occur in the correct temporal order. If not, it is assumed the footstep extraction failed for the walking sequence.

Figure 5 shows example normalized projections, 5(b), along with a plot of the raw and filtered correlation coefficient time series, 5(c), for one walking sequence. The correlation coefficient of the normalized ground plane projection of 3D points below 20 in. (50 cm) has proven to be quite robust, even at large distances from the Kinect. As previously stated, at the extreme end of the walking path shown in Fig-

ure 3, the distance from the participants to Kinect #2 is approximately 26.6 ft (8.1 m).

Given the locations of the local minima and maxima (right and left steps) in the correlation coefficient time series, the temporal gait parameter of stride time (time between successive footfalls of the same foot) can be computed. In addition, the spatial gait parameter of stride length (distance between successive footfalls of the same foot) can be approximated as the distance moved over the period corresponding to the stride time. (For the results in Section 5, this distance is calculated using the sum of frame-to-frame changes in the position of the filtered centroid time series. For the results in Section 6, this distance is computed as the distance traveled along the walking direction.) Although this approximation of the stride length may yield inaccurate measurements given large, abrupt changes in stride, it should still capture the stride-to-stride variation which studies have shown to be predictive of falls.

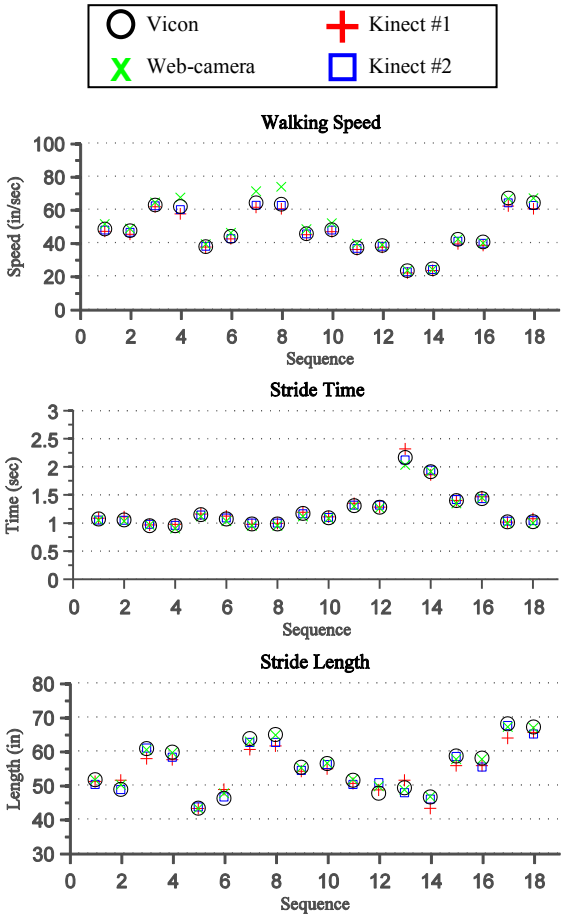

Fig. 6. Preliminary evaluation results comparing gait measurements from Kinect, web-camera, and Vicon systems.

5. Preliminary Evaluation

As a preliminary evaluation, a set of 18 walking sequences was collected and gait parameters were extracted from the three different systems. Three participants, all members of the eldercare research team at the University of Missouri, were asked to walk at slow, normal, and fast speeds, and two walks were collected for each speed for each participant. The walking path, as shown in Figure 3, was approximately 17 feet long and the number of steps per walking sequence varied from five to nine. In half of the walks the subject was moving towards Kinect #2, and in the other half the subject was moving away. (Refer to Figure 3 for a placement diagram of the different sensor systems.)

5.1. Results

Figure 6 shows plots of walking speed, average stride time, and average stride length for each of the walking sequences as computed for each of the sys-

TABLE I
PRELIMINARY EVALUATION RESULTS
ERROR DISTRIBUTION – $N(\mu,\sigma^2)$

% DIFFERENCE IN WALKING SPEED COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean % Diff	-4.1	-1.9	4.4
Std. Deviation.	1.9	1.2	4.8

% DIFFERENCE IN STRIDE TIME COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean % Diff	1.9	0.7	-2.3
Std. Deviation.	2.5	1.3	2.1

% DIFFERENCE IN STRIDE LENGTH COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean % Diff	-1.9	-1.1	0.2
Std. Deviation.	3.9	2.5	1.6

(a) (b) (c)

Fig. 7. Sample cases showing lack of returned values in Kinect depth image for certain articles of clothing: (b) and (c). Problem free case is shown in (a). ***Top*****:** RGB image. ***Middle*****:** Depth image. ***Bottom*****:** Extracted foreground. In this work, all such items of clothing were found to contain some amount of spandex.

tems. Table I provides a summary of the calculated percentage difference (mean and standard deviation) between each system and the Vicon for walking speed, average stride time, and average stride length.

As the results in Table I show, the individual Kinects yielded good performance compared to the Vicon and web-camera system in terms of the average percentage difference. The maximum absolute percentage difference recorded for Kinect #1 compared to the Vicon was 7.2%, 7.1%, and 7.2% for walking speed, stride time, and stride length respectively. The maximum absolute percentage differences recorded for Kinect #2 were 4.3%, 2.7%, and 6.5%.

The results also indicate that Kinect #2 slightly outperformed Kinect #1 in terms of accuracy in all three measurements on this dataset. However, given all of the steps involved in computing the gait parameters (calibration, foreground extraction, and gait extraction) it is not clear how much of this difference in performance is due to positioning. Ultimately, a more controlled experiment is likely required. That said, these results do indicate that accurate gait measurements are obtainable from both positions. This is significant, as the ultimate goal is to measure gait in dynamic, unstructured environments, where walking sequences may be at any arbitrary orientation with respect to the Kinect.

Interestingly, when it comes to measuring the temporal gait parameter of stride time, it would seem the web-camera system should be at a significant disadvantage compared to the Kinects due to its frame rate being six times slower. However, the results in Table I indicate the decreased frame rate does not cause a huge performance loss with respect to the Kinects. This is most likely a result of the differences in the algorithms used to extract gait information from the two systems; specifically, the limited ability to segment the feet of the participants in the Kinect depth imagery.

Finally, on the measure of stride length, the results show the web-camera system outperforming the Kinects. Here again, the ability of the web-camera system to explicitly extract the footfall locations used for the computation of stride length, as opposed to the approximation based on distance traveled of the centroid during the period corresponding to the stride time, as used for the Kinects, is the most likely factor.

5.2. Discussion

Although this preliminary investigation of the Kinect for in home gait measurement showed both the accuracy of gait measurements made using the device and many potential benefits for fall risk assessment and in-home monitoring systems, there are issues with the Kinect, some examples of which are shown in Figure 7, that need further consideration.

First, certain types of clothing fail to reflect enough of the emitted IR pattern back to the device to allow an estimate of depth at those pixels to be made [37]. Furthermore, the issue of subjects blending into the background when they are close to walls, or, in the case of fall detection, on the ground, is a concern in using the depth imagery alone for foreground segmentation and tracking. Potentially, a smart fusion of

depth and color foreground segmentation could address some of these issues. Finally, another potential drawback of the Kinect is the limited field of view, approximately 60 degrees. This restriction may require the use of multiple devices in many environments.

6. Human Subject Experiment

An additional dataset was collected, with approval of the Institutional Review Board (IRB) at the University of Missouri, to further evaluate the accuracy and methods for extracting gait parameters from the Kinect depth data. Thirteen participants were asked to perform eight walking sequences using the same setup as for the preliminary evaluation, shown in Figure 3. Due to technical difficulties, data was not collected from the Kinects for the entire duration of two of these walking sequences and they were discarded. Thus, the dataset contains a total of 102 walking sequences. Seven of the participants were male and six were female, while ages spanned a range from late 20's to late 60's. Gait information from the Kinects is presented here, along with the Vicon system for ground truth comparison.

6.1. Results

Of the thirteen participants, three of the female participants happened to be wearing pants that poorly reflected the IR pattern emitted by the Kinects (a problem illustrated in Figure 7). That is, beyond approximately 12 to 15 ft. from the Kinect, no measurements were returned from the lower half of the body in the Kinect depth imagery. All three of the garments were black in color. However, other black clothing worn by a number of participants did not

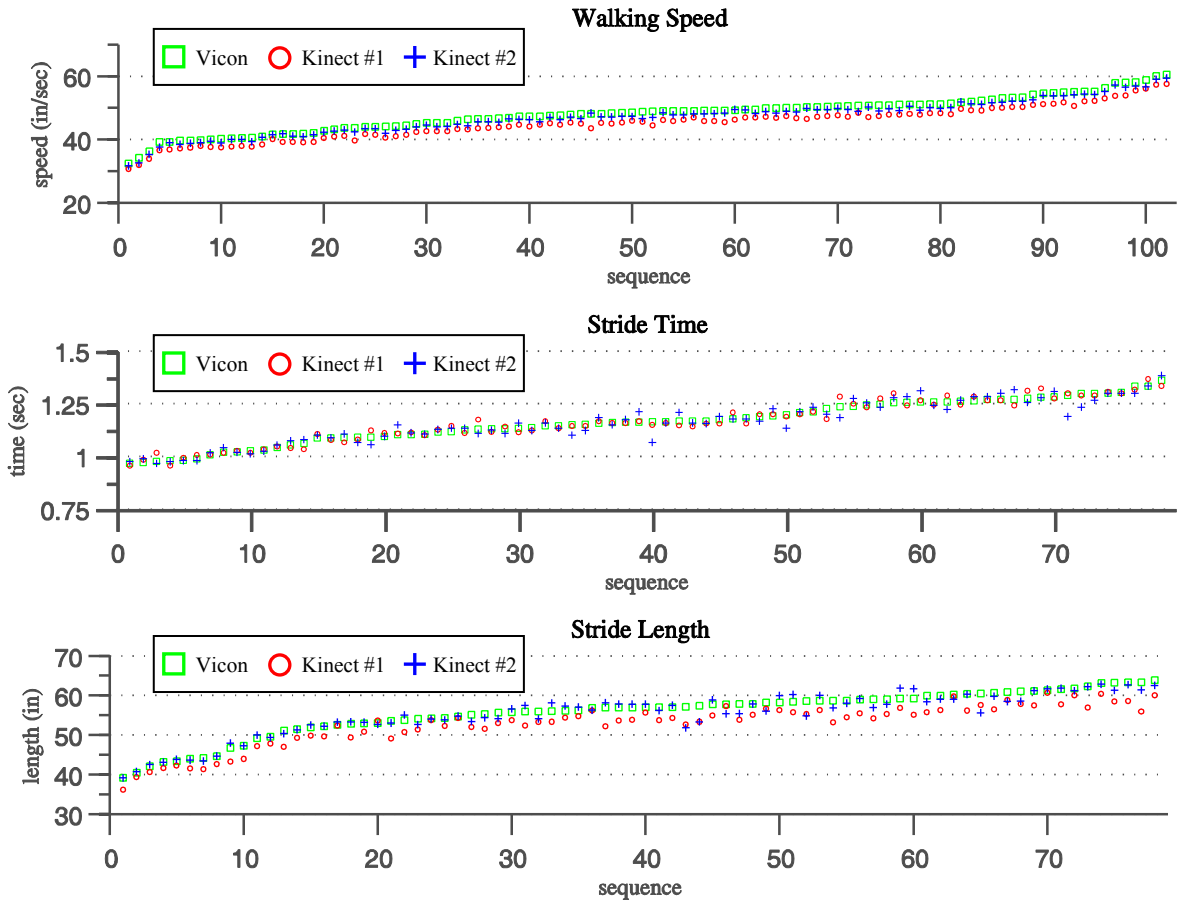

Fig. 8. Human subject experiment results comparing gait measurements from Kinect and Vicon systems. For each graph, the sequences have been sorted in ascending order by the measure.

TABLE II
HUMAN SUBJECT RESULTS
ERROR DISTRIBUTION – $N(\mu,\sigma^2)$

% DIFFERENCE IN WALKING SPEED COMPARED TO VICON

	Kinect #1	Kinect #2
Mean % Diff	-6.1	-2.1
Std. Deviation.	1.0	1.1

% DIFFERENCE IN STRIDE TIME COMPARED TO VICON

	Kinect #1	Kinect #2
Mean % Diff	0.1	-0.1
Std. Deviation.	3.3	2.4

% DIFFERENCE IN STRIDE LENGTH COMPARED TO VICON

	Kinect #1	Kinect #2
Mean % Diff	-5.1	-0.7
Std. Deviation.	2.3	2.7

display this problem. During a follow-up investigation, it was determined that the garments which displayed this poor IR reflectance characteristic contained some percentage of spandex.

As the lack of depth data from the lower half of the body makes it impossible to extract the parameters of stride length and stride time using the methods presented in Section 4.3, only walking speed is reported for the 24 combined walking sequences of these participants, as measurements were returned from the upper body. Thus, stride time and length measurements are reported for 78 of the 102 walking sequences.

As the results in Figure 8 and Table II show, the Kinects yielded good performance, similar to that presented in Table I for the preliminary evaluation, on this dataset as compared to the Vicon. The maximum absolute percentage difference recorded for Kinect #1 compared to the Vicon was 9.6%, 4.1%, and 11.7% for walking speed, stride time, and stride length respectively. The maximum absolute percentage differences recorded for Kinect #2 were 4.9%, 8.4%, and 9.4%.

6.2. Discussion

On this dataset both Kinects displayed a consistent negative bias on the spatial measurements due to underestimating the distance traveled (and thus the walking speed and stride length) for the walking sequences compared to the Vicon. This effect was larger for Kinect #1 (-6.1%) than Kinect #2 (-2.1%), and similar to the results obtained in the preliminary evaluation. More investigation is needed to identify the exact cause for these biases. However, the standard deviation of the errors for each of the Kinects for the measure of walking speed is quite low, indicating good consistency.

For the temporal parameter of stride time no significant biases were observed and the standard deviations indicate 95% of measurements should have an error of 6.7% or less.

7. Conclusion

As the Kinect has only been available for a short period of time, this work focused on evaluating the accuracy and feasibility of using the depth data from the Kinect for passive, in-home gait assessment. Results showed good agreement between gait measurements obtained using the Kinect, as compared to those from an existing web-camera based system, and those from a Vicon motion capture system. Furthermore, the depth imagery from the Kinect not only addresses a major issue in foreground extraction from color imagery (changing lighting conditions), but significantly reduces the computational requirements necessary for robust foreground extraction; potentially further reducing the cost of an in-home vision based gait assessment system.

Although the precision of the distance measurements obtained from the Kinect decreases as the distance increases, this phenomenon seemed to have a minimal impact on the accuracy of the gait parameters measured in this work, even though participants were up to 8.1 meters from the Kinect at some point during the walking sequences. The minimal impact is likely due to the fact that a large number (equal to the number of pixels classified as foreground) of measurements are averaged together to obtain the centroid position for each frame. Furthermore, the high sampling rate (30 frames per second) allows significant temporal smoothing to be applied to the resulting centroid time series. However, a more detailed

investigation of the accuracy of the measurements from Kinect at large distances is certainly warranted.

Issues were encountered related to certain clothing not sufficiently reflecting the IR pattern emitted by the Kinect at distances over approximately four meters. For these cases, no depth values were returned from the Kinect, and, thus, no gait parameters could be computed other than walking speed. Similarly, the issue of foreground objects and people blending into the background, specifically the floor, could be problematic for fall detection using the depth imagery at significant distances from the Kinect.

Future work will look at further refining the algorithms for gait extraction, obtaining and evaluating additional physical parameters from the depth data of the Kinect in home environments, and will also explore a fusion of the depth and color imagery to achieve a fast, computationally inexpensive, and more robust foreground extraction than is possible with just the depth data or color imagery alone.

Acknowledgment

This work has been supported by the U.S. National Science Foundation under Grants IIS-0703692 and CNS-0931607. The authors would also like to thank the Eldertech research team members for their help in data collection.

References

[1] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic," *Computer Vision and Image Understanding*, vol. 113, pp. 80-89, 2009.

[2] D. Austin, T.L. Hayes, J. Kaye, N. Mattek, and M. Pavel, "On the disambiguation of passively measured in-home gait velocities from multi-person smart homes," *Journal of Ambient Intelligence and Smart Environments,* vol. 3, no. 2, pp. 165-174, April, 2011.

[3] T. Banerjee, J.M. Keller, M. Skubic, and C.C. Abbott, "Sit-To-Stand Detection Using Fuzzy Clustering Techniques," IEEE World Congress on Computational Intelligence, Barcelona, Spain, July 18-23, 2010, pp 1-8.

[4] F. Caillette and T. Howard, "Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction," In *Proc BMVC*. vol. 2, pp. 597.606, 2004.

[5] R. Camicioli, D. Howieson, B. Oken, G. Sexton and J. Kaye, "Motor slowing precedes cognitive impairment in the oldest old," Neurology 50 (1998), 1496–1498.

[6] CIR Systems, Inc., "GAITRite Electronic Walkway Technical Reference," Rev. C, Feb. 11, 2010.

[7] G. Demiris, M. Rantz, M. Aud, K. Marek, H. Tyrer, M. Skubic, and A. Hussam, "Older Adults' Attitudes Towards and Perceptions of 'Smarthome' Technologies: a Pilot Study," Medical Informatics and The Internet in Medicine, June, 2004, vol. 29, no. 2, pp. 87-94.

[8] G. Demiris, O.D. Parker, J. Giger, M. Skubic, and M. Rantz, "Older adults' privacy considerations for vision based recognition methods of eldercare applications," *Technology and Health Care*, vol. 17, pp. 41- 48, 2009.

[9] B.R. Greene, A. O'Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R.A. Kenny, "Quantitative Falls Risk Assessment Using the Timed Up and Go Test," IEEE Trans. on Biomedical Engineering, vol. 57, no. 12, Dec., 2010, pp. 2918-2926.

[10] J.M. Guralnik, E.M. Simonsick, L. Ferrucci, R.J. Glynn, L.F. Berkman, D.G. Blazer, et al. (1994). A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology, 49*(2), M85-94.

[11] S. Hagler, D. Austin, T. Hayes, J. Kaye, and M. Pavel, "Unobtrusive and Ubiquitous In-Home Monitoring: A Methodology for Continuous Assessment of Gait Velocity in Elders," *IEEE Trans Biomed Eng*, 2009.

[12] N. Hansen, "The CMA Evolution Strategy: A Comparing Review," In Towards a new evolutionary computation. Advances in estimation of distribution algorithms, pp. 75-102, Springer, 2006.

[13] I. Haritaoglu, D. Harwood, L.S. Davis, "W4: real-time surveillance of people and their activities," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 809–830.

[14] J.M. Hausdorff, D.A. Rios, H.K. Edelberg, "Gait variability and fall risk in community-living older adults: a 1-year prospective study," Arch Phys Med Rehabil 2001;82:1050–6.

[15] T.L. Hayes, M. Pavel, and J.A. Kaye, "An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline," *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; San Francisco, CA, 2004.

[16] D. Hodgins, "The Importance of Measuring Human Gait," Medical Device Technology. 2008 Sep;19(5):42, 44-7.

[17] B. Jansen, F. Temmermans and R. Deklerck, "3D human pose recognition for home monitoring of elderly," 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, Aug 23-26, 2007.

[18] K. Konolige and P. Mihelich, "Technical Description of Kinect Calibration," Ros.org, Dec. 2010, http://www.ros.org/wiki/kinect_calibration/technical

[19] J. Krumm et al., "Multi-Camera Multi-Person Tracking for Easy Living," *Proc. 3rd IEEE Int'l Workshop Visual Surveillance*, IEEE Press, Piscataway, N.J., 2000, pp. 3-10.

[20] T. Lee and A. Mihailidis, "An intelligent emergency response system: preliminary development and testing of automated fall detection," Journal of Telemedicine and Telecare, vol. 11, no. 4, 2005, pp. 194-198.

[21] R.H. Luke, "Moving Object Segmentation from Video Using Fused Color and Texture Features in Indoor Environments", Technical Report, CIRL, University of Missouri, 2008.

[22] B.E. Maki, "Gait changes in older adults: predictors of falls or indicators of fear," Journal of the American Geriatrics Society, vol. 45(3), pp. 313–20, 1997.

[23] A. Mittal and L.S. Davis, "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo," In: *Proc. 7th European Conf. Computer Vision, Kopenhagen, Danmark*, Vol. X, pages 18–33, 2002.

[24] T.B. Moeslund, A. Hilton, V. Kruger, “A survey of advances in vision-based human motion capture and analysis,” Computer Vision and Image Understanding (CVIU) 104 (2–3) (2006) 90–126.

[25] M. Montero-Odasso, M. Schapira, E.R. Soriano, M. Varela,R. Kaplan, L.A. Camera and L.M. Mayorga, “Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older,” J. of Gerontol. A Biol. Sci. Med. Sci. 60 (2005), 1304–1309.

[26] Motion Capture Systems from Vicon. http://www.vicon.com/

[27] H. Nait-Charif, S. McKenna, “Activity Summarization and Fall Detection in a Supportive Home Environment,” In: *Proc. of ICPR'04*, (2004) 323-326.

[28] OpenKinect. http://openkinect.org

[29] M. Runge and G. Hunter, “Determinants of musculoskeletal frailty and the risk of falls in old age,” *Journal of Musculoskeletal and Neuronal Interactions*, 6 (2006), 167–173.

[30] E. Stone, D. Anderson, M. Skubic, and J. Keller, "Extracting Footfalls from Voxel Data," 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, Aug 31-Sep 4, 2010.

[31] E. Stone, and M. Skubic, “Silhouette Classification Using Pixel and Voxel Features for Improved Elder Monitoring in Dynamic Environments,” *SmartE Workshop*, *2011 IEEE International Conference on Pervasive Computing and Communications*, Seattle, WA, Mar 21-25, 2011.

[32] E. Stone and M. Skubic, “Evaluation of an Inexpensive Depth Camera for Passive In-Home Fall Risk Assessment,” 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, Dublin, Ireland, May 21-23, 2011.

[33] F. Wang, M. Skubic, C. Abbott, and J. Keller, "Body Sway Measurement for Fall Risk Assessment Using Inexpensive Webcams," 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, Aug 31-Sep 4, 2010.

[34] S. Wang, M. Skubic, and Zhu Y, "Activity Density Map Dissimilarity Comparison for Eldercare Monitoring ," Proceedings, 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, September 2-6, 2009, pp 7232-7235.

[35] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, “Pfinder: Real-time tracking of the human body,” In: *Proc. SPIE,* Bellingham,WA, 1995.

[36] Z. Zhou, X. Chen, YC. Chung, Z. He, TX. Han, and JM. Keller, "Activity Analysis, Summarization, and Visualization for Indoor Human Activity Monitoring," IEEE Transactions on Circuits and Systems for Video Technology, 18:11, 2008.

[37] Z. Zhou, E. Stone, M. Skubic, J.M. Keller, and Z. He, "Nighttime In-Home Activity Monitoring For Elder-care," 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, Aug. 30 – Sept. 3, 2011.