33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Passive In-Home Measurement of Stride-to-Stride Gait Variability Comparing Vision and Kinect Sensing

Erik E. Stone, *Student Member, IEEE*, and Marjorie Skubic, *Member, IEEE*

Abstract— **We present an analysis of measuring stride-to-stride gait variability passively, in a home setting using two vision based monitoring techniques: anonymized video data from a system of two web-cameras, and depth imagery from a single Microsoft Kinect. Millions of older adults fall every year. The ability to assess the fall risk of elderly individuals is essential to allowing them to continue living safely in independent settings as they age. Studies have shown that measures of stride-to-stride gait variability are predictive of falls in older adults. For this analysis, a set of participants were asked to perform a number of short walks while being monitored by the two vision based systems, along with a marker based Vicon motion capture system for ground truth. Measures of stride-to-stride gait variability were computed using each of the systems and compared against those obtained from the Vicon.**

I. Introduction

It is estimated that between 25-35% of people 65 years or older fall each year [1]. Although only a small percentage of those falls cause severe injury, a percentage that increases with age, falls among the elderly are a major health concern. Clinical research has indicated the importance of monitoring gait information for a number of medical applications [2]. Additionally, research has identified specific measures of gait which may be predictive of future falls in older adults [3-5]. However, the majority of older adults do not have their gait assessed on a regular basis.

In-home monitoring systems capable of capturing gait parameters on a continuous, on-going basis would greatly facilitate the use of such information in clinical care. A number of methods exist to measure gait parameters, such as pressure sensitive mats, accelerometer based wearable devices, motion capture systems, and clinical observation [2,6,7]. Passive, in-home vision based monitoring systems can offer the resolution needed for detailed measurements of gait parameters on a continuous basis while still addressing the preference of older adults for passive sensing systems [8]. Furthermore, research has indicated that privacy concerns of older adults related to vision based monitoring may be addressed through the use of appropriate processing (e.g., anonymizing through the use of silhouettes), control, and handling of the vision data [9].

Studies have indicated that changes in an older adult's gait such as decreased stride length and speed may in fact be adaptations related to the fear of falling, and may not necessarily be predictive of future falls. However, these studies have also indicated that the amount of stride-to-stride variation in the measures of stride length, speed, and velocity are independent predictors of future falls, and, thus, may be useful for identifying high risk individuals [3-5].

Prior work has investigated the accuracy of two passive, in-home vision based monitoring systems for capturing information about daily activity, (including falls and gait information) for the purpose of fall risk assessment, early illness detection, and the detection of functional decline. The first system consists of two low cost web cameras, while the second system makes use of a single Microsoft Kinect sensor. The calibrated web camera based system utilizes multiple views of the same scene, along with a silhouette based foreground extraction algorithm, to construct a 3D representation of the subject being monitored. The second system utilizes the depth image produced by a single Microsoft Kinect system, along with a simple foreground extraction algorithm, to produce a 3D point cloud representation of the subject being monitored. These 3D representations allow for the accurate estimation of physical parameters that is largely independent of the viewing angle or direction. Comparisons of the two systems against a marker based Vicon motion capture system and a GAITRite electronic mat have shown good agreement in extracting gait parameters of average right/left stride length and stride time, along with walking speed, for short walking sequences [10, 11]. Prior analysis of the two vision based monitoring systems has not assessed the ability of those systems to capture these measures of stride-to-stride variation.

This paper looks to assess the ability of the two vision based monitoring systems to capture the stride-to-stride variations that have been shown to be predictive of future falls in older adults. Section II of this paper reviews the basic setup and operation of the different monitoring systems. Section III contains results and analysis of comparing stride parameters extracted from the two systems to those from a Vicon motion capture system for the purpose of assessing stride-to-stride variation. Finally, Section IV provides a brief summary and discussion of future work.

II. Systems

A. Web-Camera

Our web-camera based system, show in Figure 1, consists of two inexpensive web-cameras mounted high (~2.5m) in the environment. The cameras are positioned to be roughly orthogonal, and can be equipped with fisheye lenses with a

This work was supported in part by the U.S. National Science Foundation under Grants IIS-0703692 and CNS-0931607.

E. Stone, and M. Skubic are with the Electrical and Computer Engineering Department at the University of Missouri, Columbia, MO 65211. (e-mails: ees6c6@mizzou.edu, skubicm@missouri.edu)

horizontal field of view of 180 degrees for complete room coverage. For this work, images (640x480, RGB) were captured from the cameras at five frames per second (fps).

A background subtraction technique which fuses color and texture features is used to extract silhouettes from the raw images captured by the cameras, Figure 1 (c). Both intrinsic and extrinsic calibration parameters for each of the cameras are obtained *a priori*, allowing for the silhouettes to be projected into a discretized volume space, thus yielding a 3D intersection, Figure 1 (d). The discretization is typically done using 2.54cm cubic voxel elements, and the system operates in real-time. The background models used for silhouette extraction are continuously updated using a fusion of 2D and 3D features [12].

Given the 3D intersection associated with a person for each frame in a walking sequence, the locations of footfalls can be extracted, Figure 1 (d). The process is based on intersecting the ground plane projections of voxels (from the 3D representation) below four inches, thus filtering out all positions except those where the feet are stationary on the ground. Spatial and temporal gait parameters can then be estimated for the walking sequence using the locations and occurrence times of the footfalls [10].

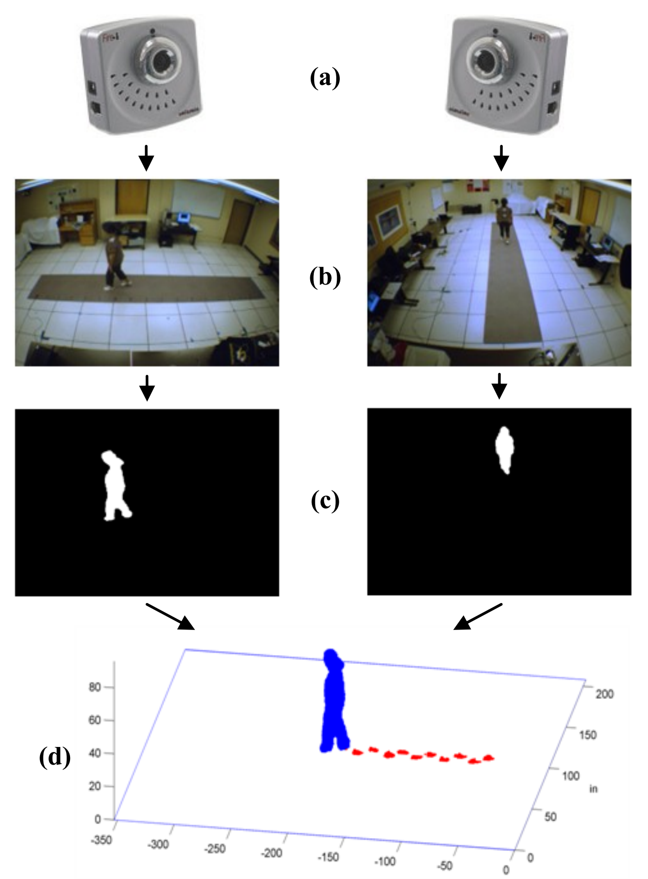

Fig. 1. Web camera based system. (a) Cameras (positioned orthogonally). (b) Views of the scene. (c) Extracted silhouettes (foreground). (d) Three-dimensional representation formed from silhouette projections, for current frame, along with history of extracted footfalls.

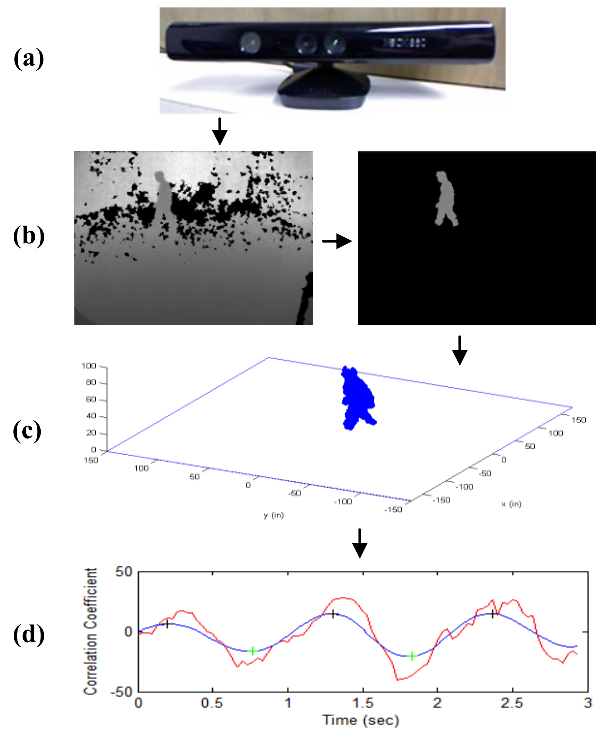

Fig. 2 Kinect based system. (a) Single Microsoft Kinect sensor. (b) Raw depth image from Kinect and extracted foreground (silhouette). (c) 3D point cloud representation of extracted foreground. (d) Plot of raw (red) and filtered (blue) correlation coefficient time series for walking sequence. Local maxima correspond to left footsteps, and local minima correspond to right footsteps.

B. Kinect

Our Kinect based system, shown in Figure 2, makes use of a single Microsoft Kinect sensor device. The Kinect, released by Microsoft, uses actively emitted structured infrared (IR) light to estimate depth at each pixel using a single IR sensitive camera. The depth image (640x480, 11 bit) is generated at 30 fps, and is invariant to changes in visible light. The Kinect also contains a standard RGB (color) camera. The device was designed to allow controller free game play on the Microsoft Xbox, which is able to perform skeletal tracking, gesture recognition, and more using the depth image [13].

Initially, the intrinsic, extrinsic, and stereo calibration parameters of the RGB and IR cameras, along with the parameters used for converting the raw depth values returned by the Kinect to distances, are estimated as described in [11]. A background subtraction technique is used to extract the foreground (silhouette) from the depth image, Figure 2 (b). A 3D point cloud representation of the extracted foreground can then be formed, Figure 2 (c). Our approach does not attempt to fit a skeletal model to the 3D point cloud.

Given the 3D point cloud representation of a person for each frame in a walking sequence, gait information can be estimated. Due to the nature of the depth image and current implementation of the system the actual footfall locations cannot be obtained in the same manner as with the two-camera system. Instead, temporal gait parameters are extracted from the time series of a correlation coefficient computed from normalized ground plane projections [11]. Specifically, at each frame those points from the 3D point cloud with a height below 20 inches are projected onto the ground plane. The projection is then normalized by subtracting the mean, and rotating based on the current estimated walking direction. Given the normalized projection, containing N points, the following correlation coefficient is computed:

$$\rho = \frac{\sum_{n=1}^{N} x_n y_n}{N}$$

where x_n and y_n correspond to the X and Y coordinates of the n^{th} point in the projection. The number of left and right steps in a walking sequence is obtained from the time series of the correlation coefficient, Figure 2 (d), as the number of local maxima and minima respectively.

Finally, spatial gait parameters are estimated using the temporal gait information, specifically the occurrence time of each footfall, combined with the movement of the centroid of the 3D point cloud during the walking sequence.

III. Experimental Analysis

For this analysis, two Kinects (each acting independently) and two web cameras (forming our web camera based system) were placed in a laboratory setting containing a Vicon motion capture system. Figure 3 shows the layout of the sensors along with the approximate location of the walking path used. The two Kinects were positioned at different angles with respect to the walking path in order to evaluate the impact of positioning on the gait parameters obtained.

A total of 18 walking sequences were collected from three participants. Each participant was asked to walk slowly for two sequences, normal for two sequences, and fast for two sequences. Each sequence contained between five and nine steps, and, thus, three to seven measurable strides. In total, the data set contained 87 individual strides.

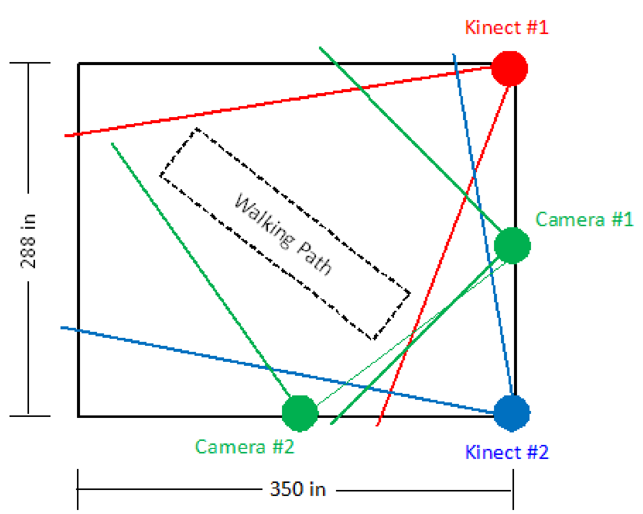

Fig. 3. Approximate position of Kinects, web cameras, and walking path in laboratory environment. Lines show field of view for each device.

A. Individual Stride Measurements

In order to evaluate the ability of the two systems to capture stride-to-stride variations, an initial analysis was performed to assess the measurement accuracy of individual strides. Specifically, the accuracy in measuring the parameters of length, time, and velocity associated with an individual stride was computed for each of the systems. The average difference and standard deviation of each of the stride parameters were computed to assess the variance inherent in the measurements.

Table I shows a comparison of the individual stride length, time, and velocity results obtained from each of the systems compared to the Vicon. Stride length was measured as the distance from one footfall to the next footfall of the same foot. Stride time was measured as the time elapsed from the occurrence of one footfall to the occurrence of the next footfall of the same foot. Finally, stride velocity was computed by dividing the stride length by the stride time.

The web camera system seems to outperform both of the Kinects when measuring the purely spatial parameter of stride length, whereas the results are mixed for the temporal parameter of stride time, and the composite parameter of stride velocity. Additionally, Kinect #2 appears to outperform Kinect #1 on all three measures.

TABLE I

INDIVIDUAL STRIDE LENGTH COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean Diff. (cm)	-0.70	-0.32	-0.09
Std. Dev. (cm)	**2.44**	**1.00**	**0.41**

INDIVIDUAL STRIDE TIME COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean Diff. (ms)	9.39	7.47	-29.31
Std. Dev. (ms)	**190.54**	**62.2**	**88.67**

INDIVIDUAL STRIDE VELOCITY COMPARED TO VICON

	Kinect #1	Kinect #2	Web-Camera
Mean Diff. (cm/s)	-0.95	-0.42	0.48
Std. Dev. (cm/s)	**1.54**	**0.64**	**1.63**

B. Stride-to-Stride Standard Deviation

The average difference and standard deviation was computed for the stride-to-stride standard deviation of stride length, time, and velocity for each walking sequences. The results are shown in Table II.

The statistics reported in [3] indicate that differences of approximately 1cm, 15.4ms, and 1cm/s in the stride-to-stride standard deviation of stride length, time, and velocity, respectively, separated the average faller and non-faller. Statistics reported in [4] indicate that a difference of approximately 50ms in the stride-to-stride standard deviation of stride time separated the average faller and non-faller. Based on those numbers, the results suggest the current implementation of the web camera system is capable of measuring stride length with sufficient accuracy to be useful, but lacks sufficient accuracy in measuring stride time and velocity. Interestingly, the results also suggest that the current implementation of the Kinect system measures stride velocity with sufficient accuracy to be useful, but lacks sufficient accuracy in measuring the separate parameters of stride length and time.

TABLE II

	Kinect #1	Kinect #2	Web-Camera
STRIDE-TO-STRIDE STANDARD DEVIATION OF STRIDE LENGTH COMPARED TO VICON			
Mean Diff. (cm)	2.60	1.09	0.53
Std. Dev. (cm)	**2.45**	**1.14**	**0.65**
STRIDE-TO-STRIDE STANDARD DEVIATION OF STRIDE TIME COMPARED TO VICON			
Mean Diff. (ms)	64.8	26.41	64.22
Std. Dev. (ms)	**104.36**	**42.84**	**33.01**
STRIDE-TO-STRIDE STANDARD DEVIATION OF STRIDE VELOCITY COMPARED TO VICON			
Mean Diff. (cm/s)	0.45	-0.04	2.77
Std. Dev. (cm/s)	**0.70**	**0.72**	**2.13**

IV. SUMMARY AND FUTURE WORK

We presented an analysis of the ability of two passive, in-home vision based monitoring systems to capture stride-to-stride gait variability which has been shown to be predictive of future falls in older adults. First, measurements of length, time, and velocity for individual strides obtained from the systems were compared against those from a Vicon motion capture system on a set of 87 individual strides. Next, the ability of the systems to measure the standard deviation in stride-to-stride gait parameters was compared against the Vicon for the set of 18 short walking sequences. Results suggest the current implementation of the web-camera system may have sufficient accuracy in measuring stride-to-stride variation in stride length, while the current implementation of the Kinect system may have sufficient accuracy in measuring stride-to-stride variation in stride velocity. Future work will look to refine algorithms and further assess the ability of these systems to measure stride-to-stride gait variability using a larger number and variety of subjects and walking sequences.

Additionally, many at-risk elderly individuals exhibit gait patterns for which these parameters are difficult to extract, such as shuffling, or the use of walking aides. We are currently investigating the use of the systems for quantifying the amount of shuffle (quick, short steps) in an individual's gait, independent of actually extracting footfalls or other gait parameters.

Finally, we are currently preparing to deploy the web camera based system in the apartments of a number of older adults living in an assisted care setting. Additionally, the frame rate of the system has been increased from five to ten fps which should improve the measurement of temporal gait parameters. Future work will look to further validate our passive, in-home vision based approaches for assessing fall risk in older adults in both laboratory and non-laboratory settings.

ACKNOWLEDGMENT

The authors would like to acknowledge the members of the Eldertech team at the University of Missouri for their help in collecting the data used in this work.

REFERENCES

[1] Sattin RW. Falls among older persons: a public health prospective. Annu Rev Public Health. 1992;13:489–508.
[2] D. Hodgins, "The Importance of Measuring Human Gait," *Medical Device Technology*, vol. 19, pp. 42-47, Sep. 2008.
[3] BE. 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.
[4] JM Hausdorff, DA Rios, HK Edelberg, Gait variability and fall risk in community-living older adults: a 1-year prospective study, Arch Phys Med Rehabil, 2001;82:1050–6.
[5] Y. Barak et al., "Gait Characteristics of Elderly People with a History of Falls: A Dynamic Approach," Phj-sical Therapy, 86, 11. 1501-1510 (2006).
[6] B. Bilney, M. Morris, and K. Webster, "Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait," *Gait and Posture*, vol. 17, pp. 68-74, 2003.
[7] 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.
[8] Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M & Hussam AA, "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.
[9] 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.
[10] E. Stone, D. Anderson, M. Skubic, and J. Keller, "Extracting Footfalls from Voxel Data," 32nd Annual Int. Conf. of the IEEE EMBS, Buenos Aires, Argentina, Aug 31-Sep 4, 2010.
[11] E. Stone and M. Skubic, "Evaluation of an Inexpensive Depth Camera for Passive In-Home Fall Risk Assessment," To appear 5th *Int. Conf. on Pervasive Computing Technologies for Healthcare*, Dublin, Ireland, May 2011.
[12] E. Stone and M. Skubic, "Silhouette Classification Using Pixel and Voxel Features for Improved Elder Monitoring in Dynamic Environments," *Workshop on Smart Environments to Enhance Health Care*, Seattle, WA, Mar., 2011.
[13] OpenKinect. http://openkinect.org