32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Extracting Footfalls from Voxel Data

Erik E. Stone, Derek Anderson, Marjorie Skubic, *Member, IEEE*, and James M. Keller, *Fellow, IEEE*

Abstract— **In this paper, we present a method for extracting footfall locations from three dimensional voxel data created from a pair of silhouettes. With the growth of the elderly population, there is a need for passive monitoring of physical activity to allow older adults to continue living in independent settings. Prior research using anonymized video data has shown good results in passively acquiring information useful for assessing physical function; and, additionally, research has shown that video data anonymized through the use of silhouettes alleviates privacy concerns of older adults towards the technology. Previous work in acquiring gait information from voxel data has not included a technique for identifying individual footfall locations, from which additional information useful for assessing asymmetric gait patterns and other physical parameters may be obtained. Furthermore, visualization of the footfall locations during a walking sequence may provide additional insight to care providers for assessing physical function. To evaluate our approach, participants were asked to walk across a GAITRite electronic mat, used to validate our results, while also being monitored by our camera system. Results show good agreement between the footfalls extracted by our system and those from the GAITRite.**

I. Introduction

As older adults are living longer, there is a need for passive monitoring systems to allow them to continue living independently while not risking their health and safety. This work is primarily focused on passively assessing physical function for the purpose of evaluating an elderly individual's fall risk on a continuous basis. Studies have shown a significant correlation between walking speed and physical function [1]. Additionally, research has indicated the clinical importance of measuring a person's gait, e.g., in determining fall risk [2].

Most existing tools for quantifying gait are very expensive, use wearable sensors, or require administration by an individual [2,3]. To address this need for continuous in-home assessment, a system utilizing two inexpensive web cameras has been developed. Recent research has shown that video data anonymized through the use of silhouettes alleviates the privacy concerns of older adults towards such technology [4]. By using silhouettes obtained from multiple cameras, a coarse, three dimensional representation of a human can be constructed. This three dimensional representation, termed voxel person, provides an accurate model from which spatial and temporal parameters can be extracted. Such voxel models have typically been used for markerless motion capture in laboratory settings [5,6]. Prior work has also been done in using such models for activity summarization and fall detection [7].

Manuscript received April 1, 2010. This work was supported in part by the U.S. National Science Foundation under Grants IIS-0703692 and CNS-0931607.

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

Prior work extracting gait information from voxel data has shown promising results. In [8], Wang et. al. were able to validate the extraction of walking speed, average step time, and average step length from voxel data against both a Vicon motion capture system, and a GAITRite electronic mat [3,9]. Furthermore, in [10], Wang et. al. were able demonstrate that different results are obtained in specific evaluation sequences as compared to those obtained during in-home scenarios, indicating the benefits of continuous monitoring.

Although accurate, the previously mentioned work does not yield individual footfall locations. Instead, the gait measurements are calculated based on the centroid of voxel person, and fluctuations in the width of the ground plane projection of the feet. In addition to helping analyze asymmetric gait patterns, individual footfall locations would help provide a visualization of an individual's walking pattern from which a care provider may gain additional insight.

This work describes a method for extracting and visualizing footfalls from voxel data. The resulting footfall locations, along with parameters measured from those footfall locations, are then validated against a GAITRite electronic mat. Section II of this paper provides a brief overview of the voxel person construction process. Section III gives a description of our method for extracting footfall locations from voxel data. Section IV contains the results obtained along with a comparison to the GAITRite results. Finally, Section V summarizes the major points and discusses future work.

II. Voxel Person

For the purpose of activity and gait analysis, a three dimensional human model, termed voxel person, is created using two inexpensive web cameras, operating at five frames per second, for which both the intrinsic and extrinsic camera parameters are estimated a priori. In order to address privacy concerns related to video monitoring, silhouettes are extracted from the raw video data using color and texture features; the background modeling and foreground extraction technique is described in [11].

Silhouettes from each of the cameras are back-projected into a discretized three dimensional space, referred to as voxel (volume element) space. For this work, the volume space, a 30x17x8 ft. room, is discretized into 1x1x1 in.

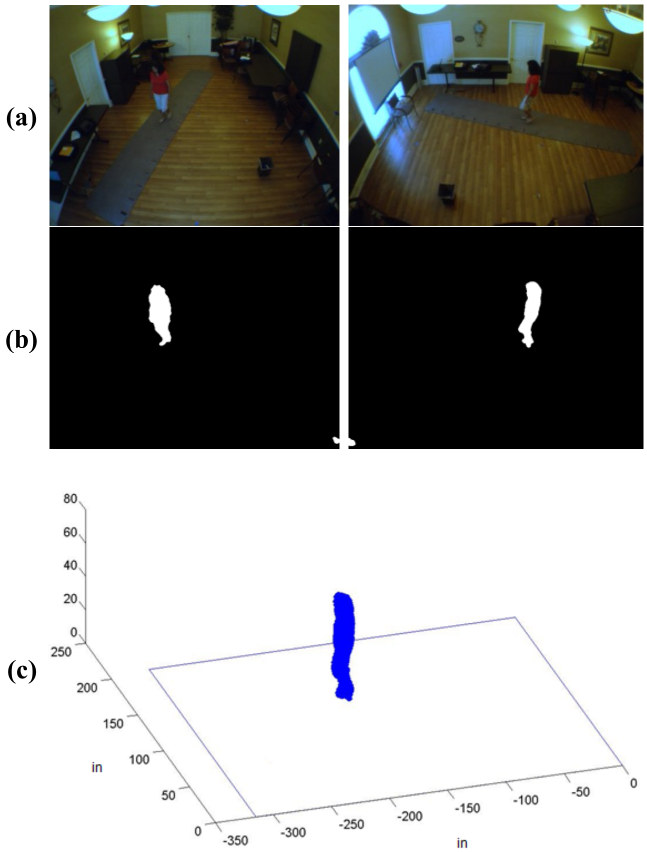


Fig. 1. (a) Two orthogonal camera views of the same scene. (b) Human silhouettes. (c) Three dimensional human model, voxel person, formed by the intersection of silhouette back-projections in voxel space.

rectangular voxels. The intersection of the two silhouette back-projections forms a three-dimensional human model. An illustration of the silhouette extraction and voxel person construction process is shown in Fig. 1. Note that the three-dimensional model allows us to capture gait parameters from any walking direction and is not dependent on the camera view. This is particularly important for continuous monitoring in an unstructured home setting.

III. Footfall Extraction

A. Projections

Our method for footfall extraction from voxel data is based on the assumption that during walking, one foot will remain stationary while the other foot is in motion. Similar to [10], voxels below two inches are used to capture information about the feet. Thus, the footfall extraction process is based on identifying 2D locations which are contained in the projection, P, of voxels below two inches onto the ground plane for a minimum number of consecutive frames, F. Such 2D locations should have a high likelihood of corresponding to a footfall.

Fig. 2 (a-b) shows the projection of voxels below two inches onto the ground plane for five consecutive frames. As the figure illustrates, only the section of the ground plane projection corresponding to a single footfall of voxel person is contained in the projection for all five frames. By taking the intersection of the five ground plane projections, P_1 thru P_5, a set, R, of those 2D locations which correspond to the footfall, shown in Fig. 2 (c), remain. This process of intersecting ground plane projections is repeated for every frame.

Specifically, a set of 2D locations, R_t, believed to correspond to footfalls, are extracted for each frame, t, by intersecting the ground plane projections of voxels below two inches over a window from $\left[t-\frac{F}{2}, t+\frac{F}{2}\right]$. Thus,

$$R_t = P_{t-\frac{F}{2}} \cap \ldots \cap P_{t+\frac{F}{2}}$$

The parameter F, which determines the number of frame projections to intersect, is estimated at each frame based on the velocity of voxel person, the height of voxel person, and the frame rate of the cameras using the following formula:

$$F_t = \left\lceil \frac{0.4\,h\,f}{2.0\,v_t} + C \right\rceil \qquad F_t \geq 2$$

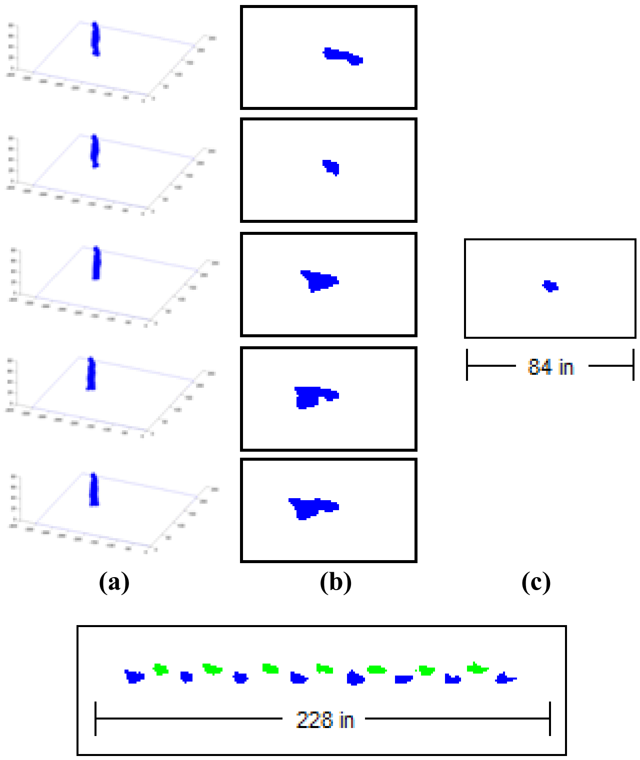
where h is the height, f is the frame rate, v_t is the velocity at time t, and C is a constant (for this work C = 0.75). F is required to be greater than or equal to two in order to limit the impact of noise from the silhouette extraction and voxel person construction process on the extracted footfalls [12].

It should be noted that given the five frame per second sampling rate of the camera system, and the requirement $F \geq 2$, the minimum time a foot needs to be in contact with the ground in order to guarantee detection is 0.4 seconds, thus limiting the maximum detectable step frequency. Furthermore, footfall extraction is only attempted during walking, which is identified using a predetermined minimum velocity threshold. Namely, if the velocity of voxel person is below 7.5 inches per second, footfalls are not extracted.

In addition to identifying the location of 2D points believed to belong to footfalls, each point is also given a right/left classification. The right/left classification is obtained by first fitting a line using least squares to the projection of the centroids of voxel person on to the ground plane within a local window, $[t-F_t, t+F_t]$ of frame t, excluding centroids within $\frac{F}{2}$ frames of t. Points extracted for frame t are then given a classification as either left, right, or unknown, based on which side of line they fall, or if they lie directly on the line. (Which side of the line is right vs. left is determined from the direction of voxel person.) Fig. 2 (d) shows a plot of the extracted footfalls for an example walking sequence, along with the right/left classification.

B. Clustering

Given the 2D point sets, R_t, for each frame, footfalls are identified by grouping the points into clusters, each of which correspond to a footfall. The points are clustered in an online fashion using both their 2D spatial location, their right/left



(a) (b) (c)

(d)

Fig. 2. (a) Three dimensional voxel person model for five consecutive frames, 68-72. (b) Ground plane projection of voxels below two inches, P_{68}-P_{72}. (c) Intersection of projections P_{68}-P_{72}, corresponding to R_{70} for $F = 5$. (d) Points belonging to footfalls for entire 86 frame sequence, with right/left classification of each footfall indicated by colors green/blue respectively.

classification, and the frame number associated with their extraction. Effectively, the algorithm attempts to add new points to existing footfall clusters; however, if a new point does not fall within a given threshold, D, from an existing cluster, using a dissimilarity measure defined below, then a new footfall cluster is formed. After each frame, new clusters are evaluated for merging into pre-existing clusters. The online clustering approach is shown in Algorithm I.

The weights in the dissimilarity measure, w_s, w_t, w_{lr}, are adjusted for each frame based on F_t; $d(p,c)$ is the Euclidean distance using spatial location; t_x and lr_x correspond to time (frame number) and left/right/undetermined (1, -1, 0) classification of the cluster, c, or point, p, respectively. C is the current number of clusters.

IV. Experimental Results

A set of 44 walking sequences, during which participants were asked to walk across a GAITRite electronic mat while also being monitored by our camera system, were used to evaluate the footfall extraction results. Of the 44 sequences, 40 yielded correct footfall extractions for the entire sequence recorded by the GAITRite. For the sequences in which footfall extraction failed, quick shuffling and limping gait patterns caused the point identification and clustering to incorrectly identify some footfall locations. Although the limited resolution of the three dimensional model does play a role, methods for addressing this limitation are discussed in Section V.

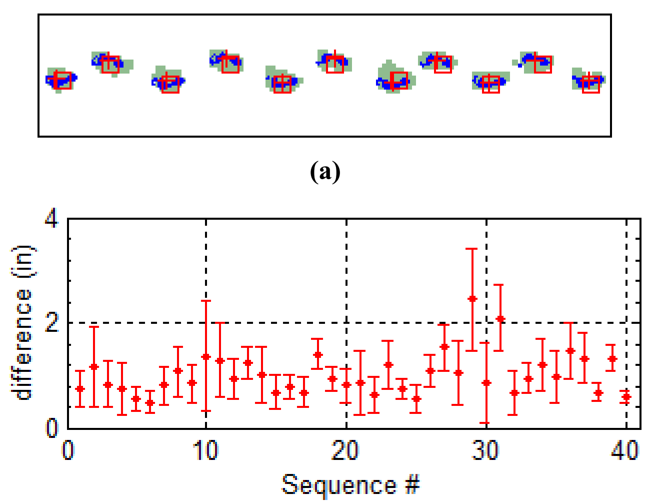
First, a direct comparison of the footfall locations identified by our camera system, compared to those obtained from the GAITRite, was conducted using the centroid of the footfall locations from each of the systems. In order to compare the two, a rigid transformation, combining rotation and translation, is estimated using least squares to best fit the centroids obtained from the camera system to those from the GAITRite. After applying the transformation, the distance between the points representing each footfall location (one from the camera system, one from the GAITRite) is

Algorithm I – Clustering

for each frame, t
  for every point, p, in R_t
    Compute dissimilarity, J, of p to current clusters
    $J_{min} = \min_{1 \le c \le C}(J_{p,c})$
    if $J_{min} < D$
      add p to cluster corresponding to J_{min}
      update parameters of cluster
    else
      form new cluster using p
    end
  end
  merge new clusters based on dissimilarity measure
end

Dissimilarity measure:

$$J_{p,c} = w_s d(p,c) + w_t|t_p - t_c| + w_{lr}|lr_p - lr_c|$$



(a)

(b)

Fig. 3. (a) Comparison of footfalls from voxel data to those from GAITRite for part of an example sequence. GAITRite footfalls shown in blue (on top) with red cross marking centroid. Voxel footfalls shown in green (beneath) with red square marking centroid. (b) Graph showing average difference in footfall location (along with standard deviation bars) for 40 test sequences compared to GAITRite results. Sequences contained between 5 and 16 steps.

computed. Fig. 3 (a) shows an example matching of the footfalls obtained from the camera system to those from the GAITRite. Additionally, Fig. 3 (b) shows a plot of the average error, with standard deviation, measured as the distance between centroids from the camera system to those of the GAITRite, for the footfall locations over all 40 sequences.

Next, the gait parameters of walking speed, right stride length, and left stride length, computed using the footfalls extracted from our camera system, were compared against the GAITRite for all 40 sequences. Fig. 4 shows a comparison of the above mentioned parameters against the GAITRite for each of the sequences, while Table I details additional statistical information about the results.

As Table I shows, although the spatial location of the footsteps (and thus the total distance and right and left stride length) is quite accurate, the low frame rate of the system introduces increased error in the time estimate as the sequence length decreases. As a result, due to the fixed length of the walking distance, increased error is observed in the walking speed as the subject moves faster.

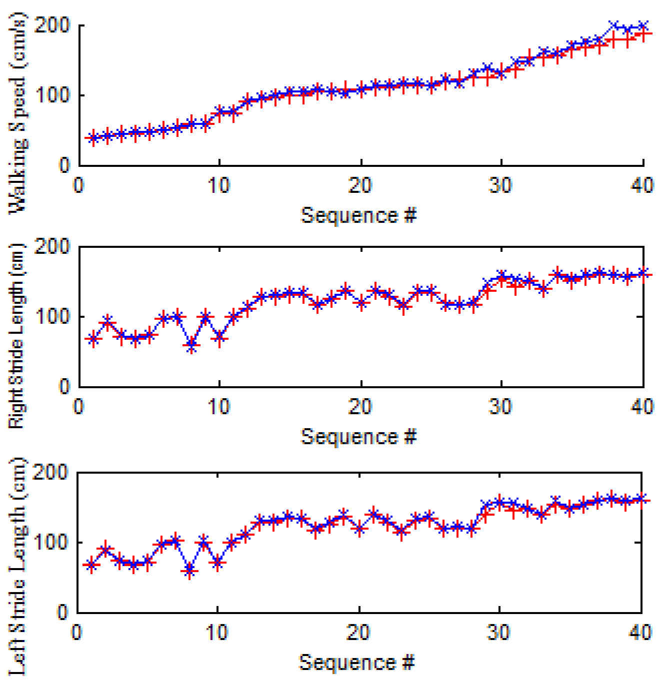


Fig. 4. Walking speed, right stride length, and left stride length results from the GAITRite (red +) and voxel data, using extracted footfalls, (blue x).

TABLE I
VOXEL DATA VS GAITRITE (40 SEQUENCES)

	avg. abs. diff (%)	std. dev.
total distance	1.4	1.5
total time	3.1	2.0
walking speed	3.7	2.6
right stride length	1.6	1.7
left stride length	1.9	3.2
right stride time	3.7	2.6
left stride time	4.1	2.7

V. SUMMARY AND FUTURE WORK

In this paper, we presented a method for extracting footfall locations from coarse voxel data obtained using a system comprised of two inexpensive web cameras, with the ultimate goal of continuous in home assessment of gait. A comparison of the extracted footfalls and corresponding gait parameters to those obtained from a GAITRite electronic mat indicated excellent agreement.

Of our 44 test sequences, four containing quick shuffling and limping gait patterns caused our footfall extraction approach to yield incorrect results. To address this, a new visualization technique based on a density plot of the ground plane projections is being developed to allow not only a visualization of all gait patterns, but also a method for quantifying the amount of shuffle present in a person's gait. Additionally, measures could be included to automatically assess the confidence of footfall extraction results.

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] F. Prince, H. Corriveau, R. Hebert, D. Winter ,"Review Article - Gait in the elderly," *Gait and Posture*, vol. 5, pp. 128-135, April 1997.

[2] D. Hodgins, "The Importance of Measuring Human Gait," *Medical Device Technology*, vol. 19, pp. 42-47, Sep. 2008.

[3] 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.

[4] 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.

[5] K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," In *Proc. of CVPR '00*, vol. 2, pp. 714-720, 2000.

[6] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *International Journal of Computer Vision*, vol. 53, pp. 199-223, 2003.

[7] 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.

[8] F. Wang, E. Stone, W. Dai, M. Skubic, and J. Keller, "Gait Analysis and Validation Using Voxel Data," In *Proc. of the IEEE Engineering in Medicine and Biology Society Conference*, pp. 6127-6130, Minneapolis, MN., Sep. 2009.

[9] K. E. Webster, J. E. Wittwer, and J. F. Feller, "Validity of GAITRite Walkway System for the Measurement of Averaged and Individual Step Parameters of Gait," *Gait and Posture*, vol. 22, pp. 317-321, 2005.

[10] F. Wang, E. Stone, W. Dai, T. Banerjee, J. Giger, J. Krampe, M. Rantz, and M. Skubic, "Testing an In-Home Gait Assessment Tool for Older Adults," In *Proc. of the IEEE Engineering in Medicine and Biology Society Conference*, pp. 6147-6150, Minneapolis, MN., Sep. 2009.

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

[12] D. Anderson, R. H. Luke, E. Stone, and J. M. Keller, "Fuzzy Voxel Object," *International Fuzzy Systems Association (IFSA)*, pp. 282-287, Lisbon, Portugal, July 2009.