Proceedings of Machine Learning Research 85:1–17, 2019 Machine Learning for Healthcare

Towards Intelligent Senior Homes: A Computer Vision-Based System for Gait Analysis with Depth Sensors

Haoye Cai[1] HCAIAA@STANFORD.EDU
Lu Yang[1] LUY@STANFORD.EDU
Zelun Luo[1] ALANZLUO@STANFORD.EDU
Jun-Ting Hsieh[1] JUNTING@STANFORD.EDU
Jay Luxenberg[3] JLUXENBERG@ONLOK.ORG
Grace Li[3] GRACELI@ONLOK.ORG
Li-Jia Li[2] LIJIALI@CS.STANFORD.EDU
N. Lance Downing[2] LDOWNING@STANFORD.EDU
Arnold Milstein[2] AMILSTEIN@STANFORD.EDU
Li Fei-Fei[1] FEIFEILI@CS.STANFORD.EDU

[1] *Department of Computer Science, Stanford University, United States*
[2] *Department of Medicine, Stanford University, United States*
[3] *On Lok Inc., United States*

Abstract

Due to the rapid increase of elderly population worldwide, cost-effective long-term care for seniors remains a daunting healthcare challenge yet to be resolved. As the shortage of traditional care resources arises, smart senior homes provide alternatives for monitoring elderly in home environments continuously and improving the well-being of seniors living independently. Non-intrusive and cost-effective sensors facilitate the discovery of the patterns in seniors' activities of daily living. However, an automated, privacy-preserving system for monitoring and analyzing gait has been long-awaited. Possible approaches to implement such a system involve using smartphones or wearing sensors, but these methods can be intrusive in that they requires adherence by seniors. In this paper, via a dataset collected from an assisted-living facility for seniors, we present a novel computer vision-based approach that uses non-intrusive, privacy-compliant depth sensors to continuously assess seniors' mobility and provide the corresponding long-term quantitative gait analytics. This analytics includes both qualitative and quantitative descriptions of walking patterns, which can provide caregivers valuable clues regarding seniors' health condition. We also assess the seniors' real-time locomotion by gait time to mimic Timed Up and Go (TUG) or walking tests which can provide clinically relevant evaluation and the early stage disease diagnosis. Our work is advancement towards a smart senior home that leverages ambient intelligence to support caregivers to help meet the healthcare challenges of an aging worldwide population.

1. Introduction

Many nations are experiencing substantial changes in the age structures of populations. In the United States, demographic projections portend an extraordinary rise in numbers of persons 65 years and older, placing increasing strain on healthcare services (Kinsella and

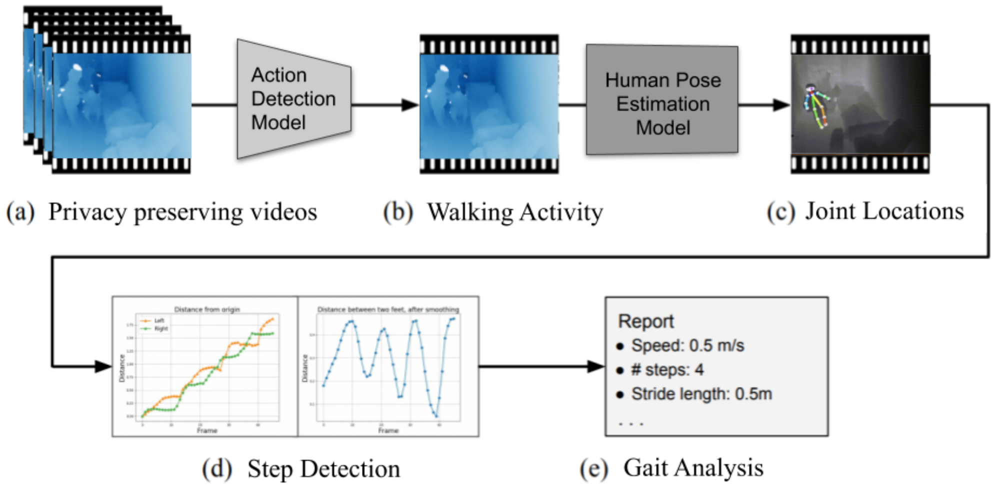


Figure 1: Pipeline of our system. Given a continuous video, we first extract the walking activity using an action detection model. Second, we train a depth pose estimator using pre-trained human pose estimation model OpenPose (Cao et al. [2017]) as initialization to re-train on public depth image dataset, then apply the trained estimator to extract the 3D joint locations of the seniors from recorded depth images. Finally, we detect steps based on poses and calculate clinically important gait parameters.

Velkoff [2001]). The global population of older persons - those aged 60 years or over - is expected to increase from 962 million to 1.4 billion in 2030, 2.1 billion in 2050, and potentially to 3.1 billion in 2100 (DESA [2017]). Despite an aging population that continues to grow, however, support for geriatric medicine and healthcare is becoming more tenuous (Butler [1997], for Disease Control et al. [2003]). It is estimated that the average number of potential caregivers per senior will drop from 7 in 2010 to 4 in 2030, and further fall below 3 in 2050, implying resultant challenges to quality elder care. (Redfoot et al. [2013]). The aging of the population also causes increased outlays for Social Security and Medicare, underlying significant societal impact. The Congressional Budget office projects the spending of major health care programs such as Medicare and Medicaid to grow from 5.2% of the U.S. GDP in 2018 to 9.2% in 2048 (Office [2018]). Concerns regarding capacity constraints for senior care is not limited to the the United States; even nations with vastly different healthcare systems, such as Japan and Great Britain, are experiencing a rapid rise in the cost for their elder healthcare systems (Akiyama et al. [2018]; Schneider and Guralnik [1990]). Therefore, it is crucial to establish cost-effective and quality long-term care for seniors worldwide.

In an effort to assist caregivers, there have been an increasing number of inventions around smart senior homes, where the elderly can be monitored with intelligent devices. These technologies not only enable frail seniors to remain safely independent in a more sustainable fashion, but can also continuously monitor their health status to preemptively assess disease risk (Courtney [2008]; Demiris et al. [2008]; Luo et al. [2018]). While most of the current senior care models have depended heavily on in-person care, many intelligent home projects have been initiated to assist the elderly without limiting or disturbing their daily routine, providing greater comfort, independence, and quality of life (Chan et al. [2008]). Data collected by the intelligent systems that monitor the seniors' key activities and track their physical well-being provide valuable clues of acute health declines. Specifically, gait assessment in home environments has recently become of increasing interest and can help identify a high risk of fall, as well as gauge the stage of some diseases. One approach to implement such a system is to use wearable devices to capture the various signals that characterize the human gait. However, this approach requires seniors to wear dedicated sensors such as accelerometers, gyroscopic sensors, etc. on several parts of the body throughout the day, which can be cumbersome or intrusive, and there has been limited success in calculating the paths and distances travelled accurately (Muro-De-La-Herran et al. [2014]; Liu et al. [2009]; Tong and Granat [1999]).

Video acquisition provides a non-intrusive alternative to wearable sensors for automated monitoring of seniors' mobility and gait patterns. Recent advances in computer vision have gained progress in analysis of human activity from video data (Wang et al. [2012]; Simonyan and Zisserman [2014]; Karpathy et al. [2014]; Ji et al. [2013]). These techniques allow us to to track individuals, recognize human behaviors, and detect unusual activities in an effective and passive manner. Though these are promising, computer vision-based descriptive analytics of seniors' daily activities using privacy-protected sensor data for clinically relevant evaluation remains a major challenge. Gait characterization is a widely-investigated field which provides useful information for multiple health-related applications such as rehabilitation, and diagnosis of medical condition Perry et al. [1992]. Gait deficits have served as an indicator of diseases such as dementia, progressive supranuclear palsy and Parkinsons disease (McIntosh et al. [1997]; Williams et al. [2007]; Verghese et al. [2002]). In cases where a senior has a neurodegenerative disease such as dementia, constant evaluation of gait-related metrics to identify markers of preclinical dementia may lead to new insights into early disease stages. Therefore, it is essential to construct an ambient vision-based gait characterization framework to enhance diagnostic assessments and discover new preventive strategies. Prior works for vision-based gait characterization have various limitations, such as extracting poses with privacy-violating RGB camera (Mehta et al. [2017]) or analyzing gait patterns with young and healthy individuals' video (Stone and Skubic [2011]), which make them not suitable for the senior home environment. To our knowledge, a non-intrusive and privacy-compliant vision-based system for continuous elderly mobility monitoring and automatic gait analysis in daily senior home environment has not been developed prior to our work.

In this work, we introduce a system that uses computer vision to automatically analyze seniors' long-term gait patterns using sensor data collected from a single depth camera placed within the home in a privacy-compliant manner. A schematic of our system implementation for gait analysis in an assisted-living facility for seniors is illustrated in Figure 1. Firstly, we acquire Activities of Daily Living (ADL) data of seniors via easy-to-install and privacy-preserving sensors in seniors' rooms (Figure 1a). Secondly, we extract the walking activity sequences using an action detection model. Thirdly, we deploy a pose estimation model to extract the skeletal joint locations from depth video of the seniors. Finally, we perform gait characterization for quantitative long-term gait analytics of the seniors. We show that we are able to achieve accurate clinically relevant gait parameter estimation that may offer valuable information regarding the seniors' health condition.

Here we showcase a real-world implementation and its efficacy of an automated, non-intrusive, and privacy-compliant vision-based system for continuous monitoring of seniors' walking activity and gait characterization. Notably, our system is privacy preserving and easy to implement as it only requires a single depth camera installed at the seniors home. Furthermore, this system can be incorporated in any smart system to help medical professionals assess the long-term gait of seniors and obtain clinically-relevant indicators of the seniors' health.

Technical Significance. We propose an automated, privacy-preserving, and non-intrusive sensing system that fits well in home care settings as well as a vision-based algorithm to conduct gait analysis over long-term monitoring. For gait analysis, we detect walking sequences using an action detection model, and then we train a state-of-the-art Convolutional Neural Networks-based model to effectively and accurately detect human joints (human pose estimation) from depth images alone, instead of RGB-D images as in previous works. Finally, we detect walking steps given detected pose sequences and further estimate various gait parameters to generate useful gait analysis.

Clinical Relevance. We propose an automated low-cost daily monitoring system to be employed in ordinary senior homes for long-term recording and analysis of seniors' gait parameters, which otherwise must be tested specifically in complicated hospital or laboratory settings. The long-term gait analysis generated by our system can enable better understanding of normal and abnormal gait statistics, and hence possibly help identify potential health problems, e.g. evaluate fall risks.

2. Cohort

We conduct this study in an assisted living facility for seniors, where sensors are deployed upon each participants consent. We deploy a depth sensing setup to identify and assess their gait in privacy-preserving manner. In addition, we simulate a test walking dataset in an effort to evaluate our pose estimation and gait analysis algorithm with annotated ground truth gait parameters since the senior home dataset did not record these.

2.1. Daily Senior Home Monitoring

Visual sensing with RGB-D cameras that are commonly utilized for computer vision tasks can raise privacy concerns in home care settings due to the presence of highly identifiable data. In contrast, depth-sensing technology provides precise measurement of the environment without revealing the individual's identity. Depth sensors capture range information where pixel values in the image correspond to a calibrated distance between objects and the sensor. Depth images offer several benefits over RGB: they are robust to the change in illumination conditions, color, and background complexity, and they carry rich 3D information of a scene in a simple way. Person localization using depth sensors like Kinect enable effective monitoring without compromising privacy to a large extent, and this has motivated the use of such sensors in various healthcare settings (Ghose et al. [2013], Haque et al. [2017]).

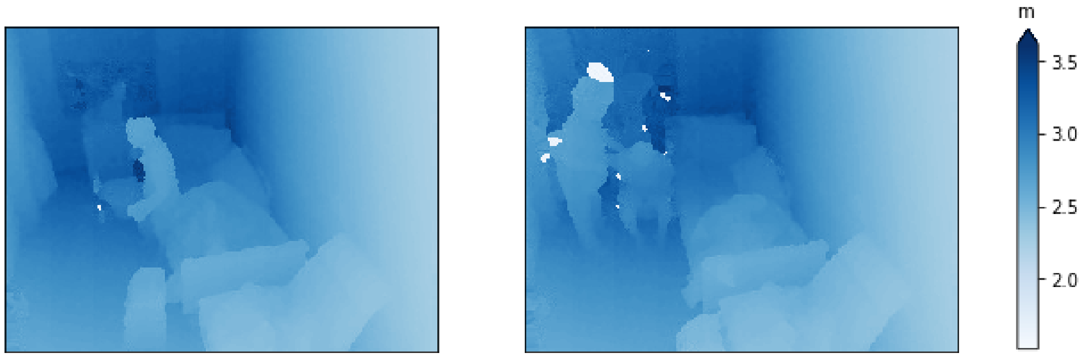


Figure 2: Examples of depth images color-coded for better visualization. Our framework utilizes privacy-compliant depth signals to detect and monitor senior mobility and walking patterns.

In particular, we deploy a depth sensing setup where an ASUS Xtion PRO depth sensor is mounted near the ceiling of a senior home room, with a 45° downward viewing angle. The depth sensor records images at 240 × 320 resolution and 30 frames per second (FPS), with distance range of 0.4 to 4.5 meters. Figure 2 illustrates the sample images provided by depth sensors, demonstrating how a depth sensor can represent accurate localization of people in complex indoor environments. We collected a dataset of depth video in several senior home rooms over the course of 1 month, using our visual sensing setup. Following depth video collection, we annotated the dataset with 6 types of daily living activities including sleeping, sitting, standing, walking, using bedside commode, and getting assistance from a caregiver, for training action detection model. For the purpose of this paper, we utilize the depth data in the walking category.

2.2. Simulation Test Data

Since the senior home dataset does not provide any ground truth gait parameters, we propose to record another test dataset to evaluate our algorithm's accuracy. We construct the test dataset of both depth and RGB videos in a lab environment using a depth sensing setup similar to that in the senior home room, where ASUS Xtion PRO depth sensors are mounted near the ceiling, with a 45° downward viewing angle at 8 locations. The vertical distances from the camera to the floor during capture are about 2 meters. The resolutions of both RGB and depth data are 240 × 320 in pixel. Both sequences are synchronized and the frame rates are 30 FPS. The depth sensor records images with distance range of 0.4 to 4.5 meters. We perform the camera calibration procedure at the beginning of video capture session and the cameras are fixed at all 8 locations.

We invited six students - three male and three female volunteers to perform the walking activity and recorded by the sensors at 8 different locations. Figure 3 displays the layout of the depth sensors along with the approximate location of the walking paths used to obtain gait parameters. The subjects are asked to perform walking along the path under each sensor location. Finally, we capture approximately 5,000 frames for a total of 48 video samples. Each video sample spans about 3 to 10 seconds. Two example frames are illustrated in

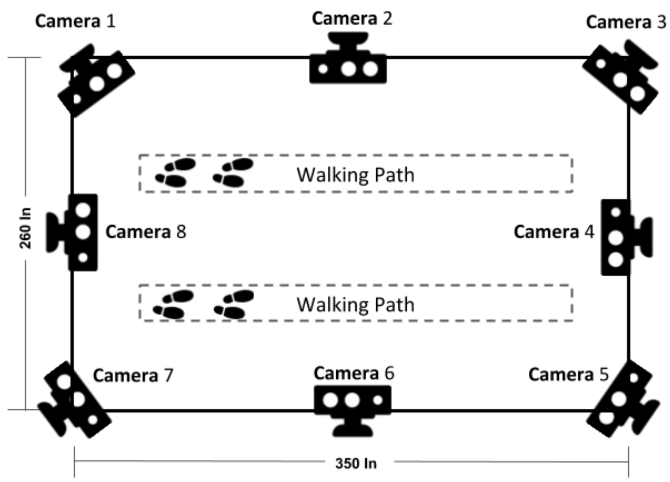


Figure 3: Approximate positions of depth cameras, and walking path in laboratory environment.

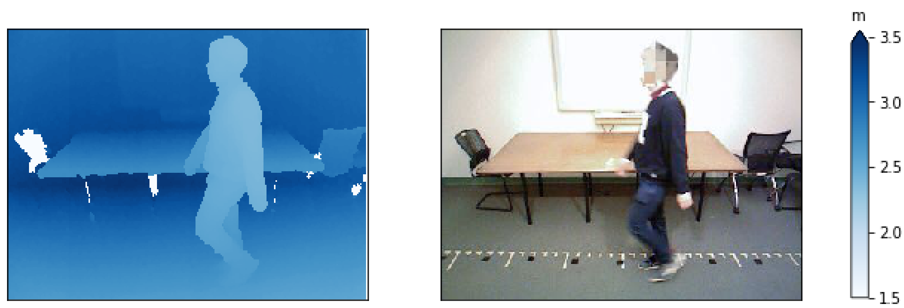


Figure 4: Examples of aligned depth modality (left) and RGB modality (right). The depth image is color-coded for better visualization. The distance travelled by the volunteer is annotated by markers on the floor.

Figure 4, including the depth modality (left) and RGB counterpart (right). The background environment of the scene is varied at different camera locations to simulate the indoor home setting. In addition, to obtain ground truth gait parameters, we visually identify each step in each video sample, and the step length and total distance travelled in depth video can be measured and visually validated by the floor markers in RGB counterparts. Then other gait parameters can be calculated accordingly.

3. Methods

As shown in Figure 1, the pipeline of our system to achieve gait analysis from depth images consists of four major steps. Given a continuous recording of a senior's daily living in depth modality, the first step is to extract clips of the senior walking with an action detection model. The second step is to train a human pose estimation model and apply the estimator to the extracted clips to obtain pose sequences of the seniors. The third and fourth steps of our pipeline involve running our gait analysis algorithm on the pose sequences to detect walking steps and calculate gait parameters accordingly. In subsequent sections, we will discuss the details of walking detection, pose estimation from depth, step detection, and gait analysis respectively.

3.1. Walking Detection

To extract walking sequences from continuous videos, we apply temporal action detection models that have been well explored in computer vision literature. Formally, given a video sequence $V = \{V_1, V_2, ..., V_T\}$ where $V_i, i \in \{1, ..., T\}$ denotes a frame at time i, the action detection model extract all video segments $\{V_{x_i}, ..., V_{y_i}\}$ corresponding to certain classes and predict the class labels. Here (x_i, y_i) denotes the start/end index of the i-th segment. In this paper we use the model to extract video segments corresponding to the walking class.

For the trainig of the detection model, we follow the same procedure as in Luo et al. [2018]. We train the model using our temporally annotated senior home dataset that has six action classes. During test time we apply our model to detect and extract only the walking class for later processing.

3.2. Pose Estimation from Depth Images

Given the extracted video frames of the walking activity, we train a pose estimation model to get the skeletal joint locations of the senior in each frame. Formally, our system $\rho : \mathbb{R}^{w\times h} \rightarrow \mathbb{R}^{w\times h\times J}$ takes as input a depth image I of size $w\times h$, and outputs 2D locations of all keypoints, which are represented as a set of joint heat maps S. The set $S = \{S_1, S_2, ..., S_J\}$ contains J heat maps, one for each joint, and $S_j \in \mathbb{R}^{w\times h}$ for all $j \in \{1, ..., J\}$ is a 2D representation of the probability that a particular joint occurs at each pixel location.

To train on our depth image datset, we first convert our ground truth joint locations into the same form. Let $\mathbf{l_j} \in \mathbb{R}^2, (\mathbf{l_j} = (x_{2j}, x_{2j+1}))$ be the 2D ground truth position for joint j. The value at location $\mathbf{p} \in \mathbb{R}^2$ in the ground truth heat map S_j^* is then defined as,

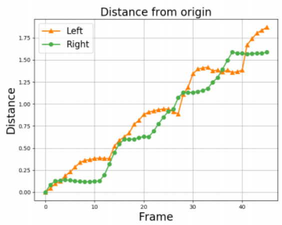
$$S_j^*(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{l_j}\|_2^2}{\sigma^2}\right) \tag{1}$$

where σ controls the variance. Then we train the network in a supervised fashion, where the loss function is defined as the Mean-Square-Error (MSE) between the predicted and ground truth joint heatmaps:

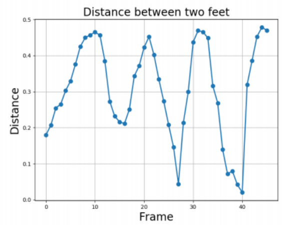
$$\mathcal{L} = \sum_{j=1}^{J} \sum_{\mathbf{p}} ||S_j^*(\mathbf{p}) - S_j(\mathbf{p})||_2 \tag{2}$$

where $S_j, j \in \{1, ..., J\}$ is the output of our network ρ, i.e $S = \rho(I)$.

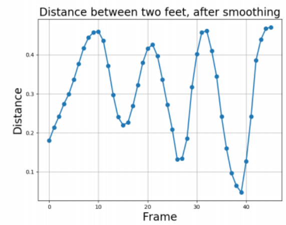
To train this model, we utilize pre-trained pose estimation model OpenPose (Cao et al. [2017]) trained on large-scale RGB datasets as our initialization and continue training using depth images with human joint annotations from CAD-120 dataset (Koppula et al. [2012]).



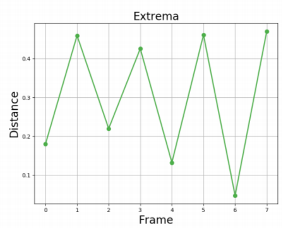
(a) Displacement of each foot from origin.



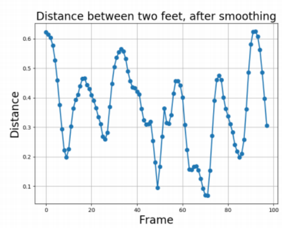
(b) Distance between two feet, d_t.



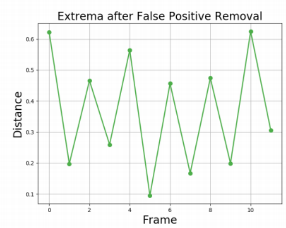
(c) Distance between feet after smoothing, $\bar{d}_t$.



(d) Extrema of the smoothed curve, m_i and n_i.



(e) Example of a $\bar{d}_t$ curve with false local extrema.



(f) Correct extrema after false positive removal.

Figure 5: Our gait analysis algorithm: (a) Extract the 3D feet locations in each frame, (b) calculate the L2 distance between the two feet, (c) apply smoothing to remove false positives, and finally (d) locate the extrema in the curve. (e) In the rare case where the $\bar{d}_t$ curve has false local extrema even after smoothing, (f) we identify the true extrema after false positive removal.

Moreover, as a post-processing step, we remove false positives with temporal smoothing among all candidates:

$$y_t = \arg\min_{\hat{y}_t} \sum_{i=t-k}^{t+k} ||y_i - \hat{y}_t||_2 \tag{3}$$

where y_t is the final joint location collection of a person for time t, $\hat{y}_t$ is a collection selected from all person candidates, and k is the window size.

Finally, with a proper transformation using the extrinsic and intrinsic camera matrix, we can calculate the 3D joint locations in the video.

Implementation Details We use the network architecture of OpenPose (Cao et al. [2017]), with VGG-Net Simonyan and Zisserman [2015] backbone. We normalize the inputs by the mean and standard deviation required by VGG-Net (Simonyan and Zisserman [2015]), which are $(0.485, 0.456, 0.406)$ and $(0.229, 0.224, 0.225)$ for BGR channels respectively. . For data augmentation, we only use random horizontal flipping. We initialize our network with pre-trained OpenPose weights, then train the model using SGD for 40 epochs with batch size 48, momentum 0.9 and weight decay 5×10^{-4}. Our initial learning rate is 0.001, and we decay the learning rate by a factor of 10 at epoch 20 and 30.

3.3. Step Detection

Once we have the sequence of joint locations of T frames, we use the feet locations across time to perform gait analysis. Figure 5 shows the pipeline of our step detection approach. Figure 5a shows the horizontal displacement from the origin of the left and right foot, $(l_t, r_t)|_{t=1}^{T}$. Next, we calculate the Euclidean distance between the two feet, $d_t = ||l_t - r_t||_2$. As shown in Figure 5b, this forms a periodic pattern across time. When a person swings a foot forward to take a step, the distance between the feet first decreases and then increases. Thus, the local maxima in the curve indicate the points when the two feet are wide apart, which correspond to the steps. However, due to the gait pattern and the errors in pose estimation, the feet distance d_t may not be smooth, resulting in false local extrema. Thus, we first apply a convolution smoothing with a uniform kernel to get a smoothed curve $\bar{d}_t$. Figure 5c shows the curve after smoothing, in which some false extrema are removed while the correct ones are preserved. Finally, as shown in Figure 5d, we determine the correct extrema $m_1, n_1, m_2, n_2, ...$, where m_i, n_i refer to the maxima and minima, respectively. Since a step occurs when the feet are farthest apart, the maxima m_i indicate when each step happens.

We observe that the above algorithm works well in practice for most video sequences. However, in some extreme cases, the distance between feet curve still contains false extrema even after smoothing. Therefore, we design an algorithm to remove them. First, we calculate the range of the extrema, $r = \max_i m_i - \min_j n_j$. Then, we determine a threshold $\theta = \alpha r$, where $\alpha < 1$. For any pair of consecutive extrema, if the difference is smaller than the threshold θ, we remove that pair. Figure 5e shows a rare example where the smoothed $\bar{d}_t$ curve still contains false local extrema, and Figure 5f shows that the algorithm can remove the false positives and obtain the correct ones.

3.4. Gait Analysis

Following our step detection methodology above, we can calculate the following medically relevant gait parameters: step length, stride length, swing time, frequency and gait speed, as previously defined in literature Muro-De-La-Herran et al. [2014]. Each local maximum m_i corresponds to each step, and each local minimum n_i occurs between steps when the feet are next to each other. We determine step length d_i^1 by the projected distance when at local maximum, stride length d_i^2 by the distance between two successive placements of the same foot, and swing time t_i by the time difference between two maximum. Finally, gait speed v is calculated by the sum of the stride length divided over the total time, to account for the case when the senior does not walk in a straight line.

Formally,

$$
\begin{aligned}
d_i^1 &= \sqrt{m_i^2 - n_i^2} \\
d_i^2 &= \sqrt{m_i^2 - n_i^2} + \sqrt{m_{i+1}^2 - n_i^2} \\
t_i &= (f_{i+1} - f_i)/fps \\
v &= \frac{\sum_i d_i^1}{(f_N - f_1)/fps}
\end{aligned}
\tag{4}
$$

Table 1: Quantitative Evaluation of Our Method. We calculate L1 errors (both mean and standard deviation) for various gait parameters. The measure units used for each parameter are shown in parentheses.

Step L1	Step length (cm)	Stride length (cm)	Swing time (ms)	Frequency	Speed (m/s)
0.21 ± 0.73	3.15 ± 2.17	4.74 ± 3.67	38.1 ± 30.0	0.24 ± 0.22	0.067 ± 0.060

Table 2: Quantitative Comparison with Other Methods. We compare L1 error of various gait parameters for each method, in terms of both absolute error and error percentage. “-” indicates that the number is unavailable in their reported results. The comparison shows our accuracy is comparable to other state-of-the-art methods.

		“Dubois 2014”	“Rocha 2018”	Ours
Step L1	ms	-	20.6 ± 20.3	7.0 ± 24.3
Step length	cm	-	2.7 ± 2.4	3.2 ± 2.2
	%	12.5 ± 12.6	-	6.3 ± 4.3
Stride length	cm	-	5.5 ± 4.0	4.7 ± 3.6
	%	-	-	4.6 ± 3.5
Swing time	ms	-	18.0 ± 15.6	38.1 ± 30.0
	%	11.5 ± 8.3	-	9.9 ± 8.7
Speed	m/s	-	0.029 ± 0.024	0.067 ± 0.060
	%	10.0 ± 7.3	-	7.2 ± 7.8

where f_i is the frame number index for m_i, fps is frames-per-second, and N is the total number of maxima found.

4. Results

The following sections first describe the quantitative results and comparisons evaluated on our simulation test dataset to illustrate the accuracy of our method. Then we present a comprehensive descriptive analytics obtained by applying our method on our real-world senior home dataset to demonstrate our system's capability to provide long-term gait monitoring and analysis.

4.1. Evaluation on Simulation Test Data

Evaluation Metrics. As the ground truth gait parameters are annotated during our simulation test data collection, our evaluation then compares our estimated parameters against the ground truth via L1 error. Specifically, for step detection accuracy (denoted “Step L1”), we calculate the L1 distances between each annotated time step and the closest detected step, and then average over all steps from all videos. “Step length” and “stride length" calculate the average of L1 distances between the ground truth step/stride length of each step and its corresponding estimated step/stride length in *cm*, and "swing time" the L1 distances to ground truth swing time in *ms*. "Frequency" is defined as the total number of steps in a video divided by its duration, and "speed" is defined as the total distance traveled (sum of step lengths) in a video divided by its duration.

Quantitative Results. Table 1 shows the quantitative error measurements for each gait parameter. We can see from the table that the accuracy is quite good since the distance errors are on the magnitude of several centimeters.

We also show quantitative comparisons with other vision-based gait analysis systems in Table 2. Here we compare with two recent methods "Dubois 2014" (Dubois and Charpillet [2014]) and "Rocha 2018" (Rocha et al. [2018]). For methods that do not provide error percentages, we assume the ground truth values are similar and directly compare the absolute errors. In this way, we can still compare and understand the magnitude of our error.

By evaluating the errors, we can see that our method achieves errors comparable to other state-of-the-art methods (sometimes better), while only requiring single depth images alone instead of RGB-D images.

4.2. Descriptive Analytics on Real-World Data.

We apply our algorithm on our unlabeled real-world senior home dataset, and provide quantitative descriptive analytics of the seniors' gait statistics. These quantitative analytics can reliably provide useful clinical information to be interpreted by a caregiver.

Gait Statistics. First we calculate the gait statistics (namely step length, swing time, frequency and gait speed) of three randomly chosen subjects over a week as an illustration. Note that we do not calculate stride length here because stride lengths are the same as step lengths in the way our seniors move, i.e. only one foot moves forward and the other foot follows rather than two feet move forward alternately.

The results are shown in Table 3, where we can see a summary of gait statistics (both mean and standard deviation) for each subject over one week's time. The calculated statistics is consistent with typical senior mobility, e.g. step length of around 20 cm. We can then further analyze each subject's gait characteristics given the recorded profiles. For example, we can infer from the table that subject 2 has a lower swing time and a higher frequency than subject 1, which implies subject 2 moves legs faster. We can also see that subject 3 (compared with 1 and 2) has a much lower swing time and higher frequency, resulting in a much faster moving speed.

Monitoring Changes. In this section, we show how our system can help monitor the gait parameter changes over a long period of time. As shown in Figure 6, we plot the gait parameter changes for step length, swing time frequency and gait speed, over one week's time monitored by our system. We show the results for subject 1. We can see in the figures that the statistics vary from day to day (for example, on the fourth day the swing time is shorter while the frequency and speed is higher, i.e. the subject moves faster, which may indicate the subject was in good mental and physical condition), but are maintained in a small fluctuation range.

Table 3: Gait Statistics Summary over One Week. We show the average statistics (both mean and standard deviation) over one week for all three subjects. These gait profiles may help keep track of the subjects' health conditions.

Subject	Step length (cm)	Swing time (s)	Frequency	Speed (m/s)
1	23.16 ± 7.27	1.77 ± 0.70	0.71 ± 0.27	0.26 ± 0.09
2	22.66 ± 4.11	1.68 ± 0.65	0.72 ± 0.29	0.27 ± 0.10
3	21.53 ± 1.14	1.25 ± 0.50	0.95 ± 0.38	0.33 ± 0.07

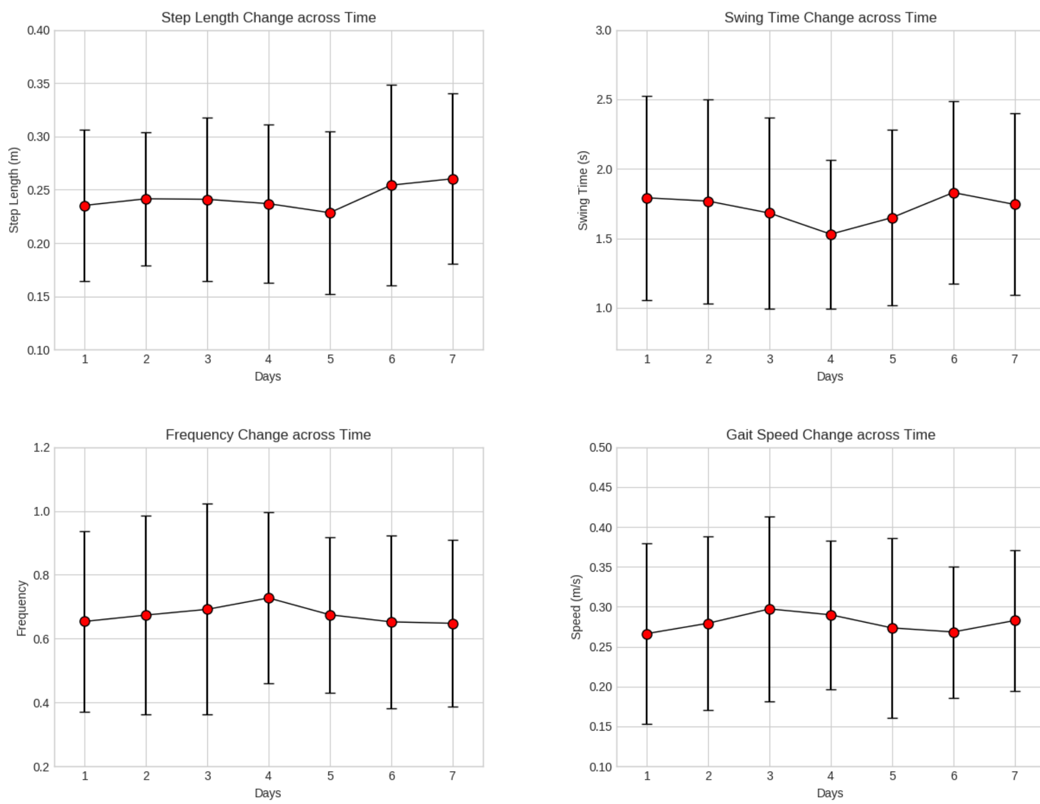


Figure 6: Gait parameter changes over time. The red dots denote the means of the parameters every day, while the black bars denote the standard deviations of the parameters every day. The curves indicate the trend of the change.

These results allows caregivers to easily observe changes over time and potentially identify anomalous mobility changes, e.g. sudden drop in step length showing decrease in senior's daily movement, which may indicate health decline or symptoms of certain disease.

Table 4: Daily Gait Time Summary over One Week. We show the average gait time over one week in each scenario. The statistics may serve as an indicator of the subject's mobility status.

Subject	From bed to door (s)	From bed to desk (s)
1	17.0 ± 3.6	23.3 ± 4.0
2	18.1 ± 5.1	22.8 ± 3.3
3	20.6 ± 3.4	25.6 ± 2.2

Adapted Standard Test. In addition to the gait statistics calculation above, we also propose and provide some standard gait analysis metrics (e.g. Timed-Up-and-Go), that are normally measured in hospital/laboratory settings, adapted to our daily monitoring setting.

We propose *Daily Gait Time*, analogous to "gait time" in laboratory test context (Freter and Fruchter [2000]). For this metric, we just measure the time that the subject spend in daily movements to walk from one fixed location to another, and these locations are chosen such that the corresponding walking sequences occur with high frequency. For our experiment here, we use walking "from bed to door" and "from bed to desk".

We show our estimated *Daily Gait Time* in Table 4, where both means and standard deviations over one week are calculated. These statistics can help monitor long-term mobility status for walking along a fixed distance, which may provide medical insights on the senior's health condition.

5. Discussion and Related Work

In this era where aging societies are becoming a reality, smart senior homes have started to gain increasing interest as more elderly people desire to live independently (Wiles et al. [2012]). A number of intelligent homes have been equipped with various sensors to accommodate people who need special care, especially elderly individuals (Demiris et al. [2008]; Elger and Furugren [1998]). However, most of these approaches require installation of multiple sensors, each of which associate with specific applications and raise concerns about privacy. Recently, a privacy-preserving and non-intrusive system has shown promise in detecting activities of daily living from a single camera, providing an efficient way to monitoring health conditions of seniors (Luo et al. [2018]).

Some of the existing approaches using wearable device-based systems for gait measurement and analysis can provide detailed statistics of a person's stride length, velocity, and swing time, but they require subjects to wear different types of motion senors and systems throughout the clinic session often when they are being attended in person (Muro-De-La-Herran et al. [2014]). A few related works propose to use low cost devices, such as attaching small sensors on mobile devices to extract pose information (Bao and Intille [2004]; Unuma et al. [2004]). However, a high degree of accuracy can only be achieved with certain style of carrying the cell phone, which raises feasibility concerns for activity monitoring of elderly individuals. Above all, most previous work on gait analysis can only be performed over a short period of time, giving biased evaluations and not capturing the long-term gait pattern (Muro-De-La-Herran et al. [2014]). A number of vision-based approaches have also been developed for automated gait monitoring, many of which lack adaption to gait variety and uses privacy-violating RGB videos or simulated data (Cheng et al. [2014]). While these methods are an important development for gait characterization, they have not focused on detecting gait deviations from normal patterns in the senior home environment.

The Timed-Up-and-Go (TUG) test and its variants have been commonly used to measure the physical and functional mobility of frail older adults, and to help identify those who are prone to falls (Shumway-Cook et al. [2000]). Despite its ubiquity in clinical assessment of balance and mobility, TUG has its own limitations such that it only examines a senior's fall risk at a particular point in time. In addition, clinical tests of mobility are usually performed in laboratory, and are not realistic to be adapted to the senior home setting. Therefore, effective long-term monitoring paves a novel way for reliable characterization of gait, and allows medical professionals to track seniors' mobility over time (Muro-De-La-Herran et al. [2014]).

Our paper therefore implements an automated, vision-based, and privacy-preserving system that accurately monitors clinically-relevant walking patterns of seniors. We experiment on depth sensor data from a realistic senior home environment and a simulation walking test dataset in both depth and RGB modalities. In addition, our model achieves excellent results on gait characterization compared to other state-of-the-art algorithms of mobility assessment while using fewer input signals. Furthermore, we have provided interpretable quantitative descriptive analytics of the senior's walking pattern over a long period. These analytics can enhances safety of senior residents, and allow caregivers and medical professionals to more efficiently identify patterns of abnormal activity that might be representative of health concerns.

References

Naomi Akiyama, Takeru Shiroiwa, Takashi Fukuda, Sachiyo Murashima, and Kenshi Hayashida. Healthcare costs for the elderly in japan: Analysis of medical care and long-term care claim records. *PloS one*, 13(5):e0190392, 2018.

Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*, pages 1–17. Springer, 2004.

Robert N Butler. Population aging and health. *Bmj*, 315(7115):1082–1084, 1997.

Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

Marie Chan, Daniel Estève, Christophe Escriba, and Eric Campo. A review of smart homes-present state and future challenges. *Computer methods and programs in biomedicine*, 91 (1):55–81, 2008.

Hong Cheng, Zicheng Liu, Yang Zhao, Guo Ye, and Xinghai Sun. Real world activity summary for senior home monitoring. *Multimedia Tools and Applications*, 70(1):177–197, 2014.

Karen L Courtney. Privacy and senior willingness to adopt smart home information technology in residential care facilities. *Methods of information in medicine*, 47(01):76–81, 2008.

George Demiris, Brian K Hensel, Marjorie Skubic, and Marilyn Rantz. Senior residents perceived need of and preferences for smart home sensor technologies. *International journal of technology assessment in health care*, 24(1):120–124, 2008.

UN DESA. World population prospects, the 2017 revision, volume i: comprehensive tables. *New York United Nations Department of Economic & Social Affairs*, 2017.

A. Dubois and F. Charpillet. A gait analysis method based on a depth camera for fall prevention. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4515–4518, Aug 2014. doi: 10.1109/EMBC.2014.6944627.

Gerhard Elger and Barbro Furugren. Smartbo-an ict and computer-based demonstration home for disabled people. In *Proceedings of the 3rd TIDE Congress: Technology for Inclusive Design and Equality Improving the Quality of Life for the European Citizen. Helsinki, Finland June*, volume 1, 1998.

Centers for Disease Control, Prevention (CDC, et al. Trends in aging–united states and worldwide. *MMWR. Morbidity and mortality weekly report*, 52(6):101, 2003.

Susan H Freter and Nadine Fruchter. Relationship between timed up and go and gait time in an elderly orthopaedic rehabilitation population. *Clinical Rehabilitation*, 14(1): 96–101, 2000. doi: 10.1191/026921500675545616. URL https://doi.org/10.1191/026921500675545616. PMID: 10688350.

Avik Ghose, Priyanka Sinha, Chirabrata Bhaumik, Aniruddha Sinha, Amit Agrawal, and Anirban Dutta Choudhury. Ubiheld: ubiquitous healthcare monitoring system for elderly and chronic patients. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 1255–1264. ACM, 2013.

Albert Haque, Michelle Guo, Alexandre Alahi, Serena Yeung, Zelun Luo, Alisha Rege, Jeffrey Jopling, Lance Downing, William Beninati, Amit Singh, et al. Towards vision-based smart hospitals: A system for tracking and monitoring hand hygiene compliance. *arXiv preprint arXiv:1708.00163*, 2017.

Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35 (1):221–231, 2013.

Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

Kevin G Kinsella and Victoria Averil Velkoff. *An aging world: 2001.* Number 1. Bureau of Census, 2001.

Hema Swetha Koppula, Rudhir Gupta, and Ashutosh Saxena. Learning human activities and object affordances from RGB-D videos. *CoRR*, abs/1210.1207, 2012. URL http://arxiv.org/abs/1210.1207.

Tao Liu, Yoshio Inoue, and Kyoko Shibata. Development of a wearable sensor system for quantitative gait analysis. *Measurement*, 42(7):978–988, 2009.

Zelun Luo, Jun-Ting Hsieh, Niranjan Balachandar, Serena Yeung, Guido Pusiol, Jay Luxenberg, Grace Li, Li-Jia Li, N Lance Downing, Arnold Milstein, et al. Computer vision-based descriptive analytics of seniors daily activities for long-term health monitoring. *Machine Learning for Healthcare (MLHC)*, 2018.

Gerald C McIntosh, Susan H Brown, Ruth R Rice, and Michael H Thaut. Rhythmic auditory-motor facilitation of gait patterns in patients with parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 62(1):22–26, 1997.

Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. Vnect: Real-time 3d human pose estimation with a single rgb camera. *ACM Transactions on Graphics (TOG)*, 36(4):44, 2017.

Alvaro Muro-De-La-Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2):3362–3394, 2014.

Congressional Budget Office. The 2018 long-term budget outlook. *(US Congress) Washington DC*, 2018.

Jacquelin Perry, Jon R Davids, et al. Gait analysis: normal and pathological function. *Journal of Pediatric Orthopaedics*, 12(6):815, 1992.

Donald Redfoot, Lynn Feinberg, and Ari N Houser. *The aging of the baby boom and the growing care gap: A look at future declines in the availability of family caregivers*. AARP Public Policy Institute Washington, DC, 2013.

Ana Patrcia Rocha, Hugo Miguel Pereira Choupina, Maria do Carmo Vilas-Boas, Jos Maria Fernandes, and Joo Paulo Silva Cunha. System for automatic gait analysis based on a single rgb-d camera. *PLOS ONE*, 13(8):1–24, 08 2018. doi: 10.1371/journal.pone.0201728.

Edward L Schneider and Jack M Guralnik. The aging of america: impact on health care costs. *Jama*, 263(17):2335–2340, 1990.

Anne Shumway-Cook, Sandy Brauer, and Marjorie Woollacott. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Physical therapy*, 80(9):896–903, 2000.

Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. URL http://arxiv.org/abs/1409.1556.

Erik E Stone and Marjorie Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *2011 Annual international conference of the IEEE engineering in medicine and biology society*, pages 6491–6494. IEEE, 2011.

Kaiyu Tong and Malcolm H Granat. A practical gait analysis system using gyroscopes. *Medical engineering & physics*, 21(2):87–94, 1999.

M Unuma, K Kurata, A Toyama, and T Horie. Autonomy position detection by using recognition of human walking motion. *Trans. IEICE*, 87(1):78–86, 2004.

Joe Verghese, Richard B Lipton, Charles B Hall, Gail Kuslansky, Mindy J Katz, and Herman Buschke. Abnormality of gait as a predictor of non-alzheimer's dementia. *New England Journal of Medicine*, 347(22):1761–1768, 2002.

Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1290–1297. IEEE, 2012.

Janine L Wiles, Annette Leibing, Nancy Guberman, Jeanne Reeve, and Ruth ES Allen. The meaning of aging in place to older people. *The gerontologist*, 52(3):357–366, 2012.

David R Williams, Janice L Holton, Kate Strand, Tamas Revesz, and Andrew J Lees. Pure akinesia with gait freezing: a third clinical phenotype of progressive supranuclear palsy. *Movement disorders: official journal of the Movement Disorder Society*, 22(15): 2235–2241, 2007.