Noname manuscript No.
(will be inserted by the editor)

Classification of Gait Anomalies from Kinect

Qiannan Li · Yafang Wang* · Andrei Sharf · Ya Cao · Changhe Tu · Baoquan Chen · Shengyuan Yu

Received: date / Accepted: date

Abstract A persons manner of walking, or their gait are important features in human recognition and classification tasks. Gait serves as an unobtrusive biometric modality which yields high quality results. In comparison with other biometric modalities, its main strength is its performance even in data that is captured at distance or at low resolution. In this paper we present an algorithm for classification of gait disorders arising from neuro-degenerative diseases such as Parkinson and Hemiplegia. We focus on motion anomalies such as tremor, partial paralysis, gestural rigidity and postural instability. The analysis and classification of such motions is challenging since they consist a multiplicity of subtle formations while lacking a regular pattern or major cycle. We introduce a gait representation which is invariant to the walking cycle and yields an efficient similarity metric. Our method performs on the joints' motion trajectories of a 3D human skeleton captured by a Kinect sensor. The algorithm is robust in that it does not require calibration, synchronization or a careful capturing setup. We demonstrate its efficiency by classifying different degenerative cases with high accuracy even in presence of noise and low resolution acquisition.

Qiannan Li, Yafang Wang *Corresponding author, Changhe Tu, Baoquan Chen
School of Computer Science and Technology, Shandong University, Jinan, P.R.China
E-mail: yafang.wang@sdu.edu.cn

Andrei Sharf
Computer Science Department, Ben-Gurion University, Beer-Sheva 84105, Israel

Ya Cao, Shengyuan Yu
Department of Neurology, Chinese PLA General Hospital, Beijing 100853, P.R.China

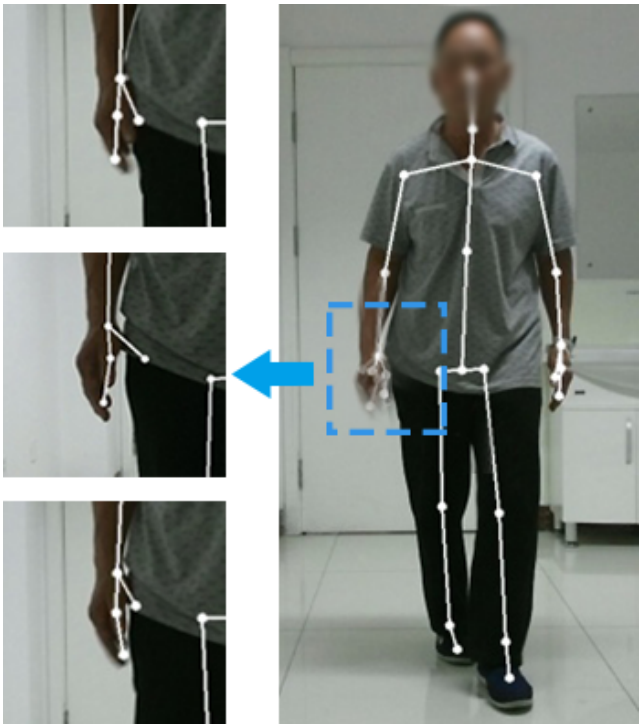


Fig. 1 Hand tremors and gait anomalies in a Parkinson patient. Extracted skeleton from Kinect show the different hand pose geometries (left column).

1 Introduction

Human gait serves as an important biometric feature for recognizing people based on their individual walking styles. Unlike other biometric features such as iris or fingerprints, gait does not require high resolution capturing or special equipment. Its main strength is its performance even in data that is captured at low resolution or that is noisy and partial. A seminal early work in this area was carried out by psy-

chologists in 1971, when Johansson attached light points to the joints of people bodies in a dark room. The results suggested that people can recognize each other merely by walking styles [16]. Suggested computational tools for gait analysis, try to mimic these capabilities of humans to recognize people based on their walking style even at a long distance.

Our work follows this path, taking a step further and exploring gait deficiencies and disturbances. Such disturbances, generally denoted *gait anomalies*, are motion patterns which are typically very complex. They may include phenomena such as tremor, partial paralysis, gestural rigidity, postural instability and etc. These motions are challenging as they are typically non-periodic, lacking a regular synchronized repetitive pattern and without a clearly defined start and end points. Furthermore, their patterns are erratic and may consist complex configurations as several phenomena interfere with each other. In contrast, normal walk cycles have a well defined repetitive pattern controlled by a simple state machine which synchronizes and balances the motion.

Gait disorders are very common in neuro-degenerative diseases as Parkinson and Hemiplegia. These diseases are characterized by the loss of neurons in brain, resulting in disfunction of circuites that mediate motor functions. As a result, there can be a multitude of motor symptoms such as rigidity, akinesia, bradykinesia, rest tremor, and postural abnormalities, most of them affecting walking ability [6]. Current standard for assessment of symptoms is highly subjective using clinical scales such as the 'Unified Parkinson Disease Rating Scale (UPDRS)' [33]. For example, in an observation-based gait analysis by a clinician, the approximations are made based on the deviation of gait parameters from normal gait symmetry.

In this work, we record human gait using a single Kinect camera. Such low-cost non-intrusive depth scanners allow capturing 3D human walking cycles at nearly video rates within a very simple setup. In addition to a raw space-time volume, Kinect extracts a 3D virtual skeleton of the body which stays coherent in time [34]. These capabilities, packed in an affordable and compact device, already led several researchers to propose Kinect as appropriate means for gait acquisition and recognition [36, 11].

We introduce a novel gait representation algorithm that is invariant to walking state and pose in front of the camera. Our intuition is that the relative movements of body parts with each other form a principal gait feature. Thus, to efficiently represent human gait, we use the covariance between skeletal joints combining their trajectory positions and speed to model their relative movement. Gait classification performs by learning a similarity distance from the trained gait models.

Thus our paper makes the following novel contributions:

- An enhanced gait descriptor, encoding joints' positional and speed data in a covariance-based descriptor
- A segmentation of the motion sequence into local time-windows which allows walk-cycle invariance thus relieving the need to synchronize walking sequences.
- Learning a similarity distance from training set which allows improving accuracy rates

We demonstrate our method using a set of recorded walking cycles originating from persons with Parkinson and Hemiplegia diseases as well as healthy persons. Data is typically noisy as it is captured using a simple setup consisting of a single $Kinect^1$ camera. Nevertheless, results show a high classification accuracy which supports human gait being a powerful and efficient biometric feature.

Due to its simplicity, our system naturally lends to health providers and medical centers, allowing non-expert personnel and staff to easily capture patients with neuro-degenerative diseases. Furthermore, no special setup is required and humans are recorded in a natural environment such as a standard-sized room where they may walk in front of the camera.

2 Related Work

Human gait has been shown to be an important biometric feature which has been applied in a wide range of tasks [14, 15, 18]. It is beyond our scope to survey the full literature on gait processing and instead we focus our discussion on recent approaches in 2D and 3D as well as in the realm of medical apps.

2.1 Video-based

Videos have been a traditional means for human gait acquisition and exploration. Researchers have showed that video-based gait is an indicative feature for the efficient recognition of a person's gender [21, 30], age [26] and identity. [23] provides a thorough survey of video-based gait recognition methods.

Silhouette plays a fundamental role in video-based approaches [29, 5, 25]. [21] propose a gate representation based on features extracted from silhouettes of human walking motion for the purpose of person identification and classification. Similarly, [20] introduce an algorithm that utilizes spatio-temporal silhouette templates for gait recognition. [22] partitions the human silhouette into meaningful components and performs classification by studying the gait of each component and their combinations. [3] represent the motion trajectory space as a set of deforming basis shapes. Similar to us, they use the discrete cosine transform (DCT) as means to process motion in dual space in a compact and stable manner.

The Gait Energy Image (GEI) was introduced as an effective representation for gait classification [13, 26]. [28] use

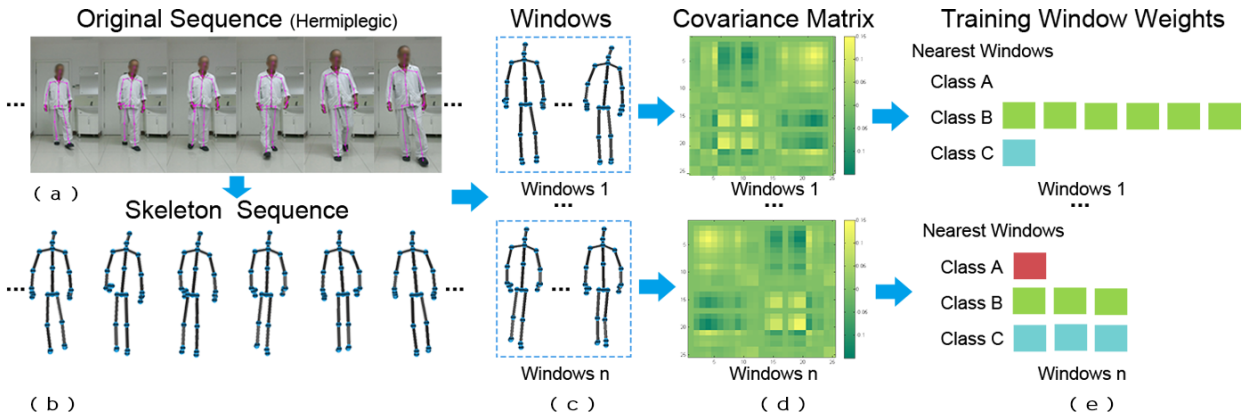


Fig. 2 Overview of our pipeline. Starting from a depth captured motion sequence (a), we extract the skeletal structure (b) and segment the motion sequence into fixed size motion windows (c). We use a covariance based descriptor to represent joints motions and speeds (d) and compute a classifier accordingly (e).

GEI as gait representation and cast the gait recognition problem as a bipartite ranking problem thus leveraging training samples from different classes/people and even from different datasets.

[2] proposes a super-resolution method for low framerate videos by performing gait recognition in high framerate examples. [24] perform gait-based gender classification for humans walking freely in any direction. Similar to us, they learn a distance metric for silhouettes from a training set which yields the optimal gender classification. Similar to us, [27] consider the subject's speed in their gait recognition algorithm. Nevertheless, they model speed using a cylindrical manifold whose azimuth and height correspond to phase and stride and use it to generate a constant speed gait sequence.

Nevertheless, video-based approaches are essentially 2D, relying on high quality video and coherent silhouette extraction. In contrast, we introduce a simple acquisition system using a single depth camera which captures high level 3D gait information. Furthermore, our algorithm does not assume a controlled environment and is invariant to walking pose and step, yielding highly accurate results.

2.2 Kinect-based

Recently, significant efforts have been made to process depth data obtained from Kinect cameras [35,11,7]. Kinect offers an attractive processing platform due to its low-cost build, non-intrusive acquisition, available software and coherent skeleton extraction [4,9].

Kinect allows the extraction of a skeleton structure from the depth frames, coherently connecting a set of joints by rigid segments in time. Since it offers an efficient and compact representation of 3D human motion, it has been successfully applied for human gait recognition [38,39] and action classification [4,40] purposes.

[19] introduce a gait recognition algorithm, which computes a covariance matrix from the skeleton joints' trajectories. This is an efficient representation which accounts for the relative body parts movement and thus is invariant to pose and walk cycle. In our work we use a similar covariance based gait representation. Nevertheless, our method focuses on gait anomalies which are more challenging due to their irregular and complex motion patterns which are accounted. Therefore, we enhance our gait representation with both positional and pace information. Furthermore, our classification method is more elaborate to suffice data challenges.

[1] use horizontal and vertical distances between skeletal joints to define an efficient feature vector which they use for gait based human recognition. Authors demonstrate the simplicity of extracting such features from depth camera as opposed to 2D video based techniques and their significant improvement to recognition accuracy.

In [40], the 3D rotational and translational relationships between various body parts are represented as curved manifolds in the Lie group. Action classification performs by mapping action curves to vector space in Lie algebra. While this representation performs well for some actions, our anomaly gait data consists of many subtle fine motions which require representing and weighting them in when computing our classifier.

[8] proposes an algorithm to reliably compute joints trajectories in scanned human motion from a side view. They show the application of their joint extraction to gait analysis for the "Get Up and Go Test" in the field of rehabilitation. While joints and trajectories may be noisy and even disappear due to occlusions, our gait representation and classification are robust and can handle such problems properly.

Finally, an evaluation of different Kinect-based feature sets for gait recognition is performed in [9]. Among their

findings, they show that mutual distances between joints are more robust than angles and lower body parts are more reliable than upper body parts. In our work, we follow this direction and encode relative joint positions as well as considering both lower and upper parts motions which have proved to be significant in neural-degenerative diseases.

2.3 Clinical apps

Kinect captured motion data has been used recently also in the context of clinical applications. Stone and Skubic [36, 37] analyze Kinect motion data to detect feet touching the ground and automatically generate alerts to clinicians in response to specific gaits of in-home residents.

A similar Kinect-based system is also used in [31] to help preventing falls of elderly people at home. They analyze gait, comparing normal versus abnormal walking as well as transitions between sitting and standing. Thus, their system monitors and detects events when elderly people are likely to fall.

Gait analysis has been widely researched in the context of Parkinson disease due to its prominent gait patterns such as tremors and paralysis. In [17] a review is presented on patents in the area of computerized gait disorder analysis using computer vision methods.

Recently, it has been shown that Kinect sensors may be efficiently utilized in the context of Parkinson disease (PD) [36]. In [12], authors test accuracy of Kinect in comparison to an expert-level scanner, in measuring clinically relevant movements in people with PD. Their study shows that Kinect can accurately measure timing and gross spatial characteristics of clinically relevant movements, although very small movements, such as hand clasping are captured with lower spatial accuracy. Nevertheless, we show that by learning the full body joints representation from Kinect in a robust manner, our classifier recognizes gait anomalies of degenerative diseases with good accuracy.

Similar to us, [32] have recently shown the utilization of Kinect for gait analysis in PD patients. In their work, they detect specific gaits parameters which yield the best recognition results (e.g. the variance of the shoulder center velocity). While both works demonstrate the potential of using Kinect as a PD assessment tool, ours takes a general and computes a classifier for gait anomalies arising from various degenerative diseases such as PD and Hemiplegia. Thus, we do not search specific gaits but instead represent the full body motion as a generic vector storing joints positions and velocities and compute the all-to-all covariance matrix.

3 Overview

Our aim here is to compute a robust gait anomaly classifier which allows a simple and straightforward capture setup which may be utilized by non-experts easily. Therefore, we use an off-the-shelf $Kinect^1$ camera and capture humans walking freely in its field of view. To extract the moving skeleton from the raw depth sequence, we use $Kinect^1$'s published SDK, yielding a 25 joints skeleton for un-occluded body motion (see Figure 2 for an overview).

In a preprocessing step, we first normalize the extracted skeletons to obtain a position and scale invariant skeleton data. The full skeleton motion sequence is then segmented into fixed sized possibly overlapping time windows which we then process independently. Thus, our method is invariant to walking state and does not assume a synchronized walking cycle in time .

For each time window, we compute its trajectory descriptor by first removing its high-frequency noise and then computing all joints' relative positional and speed information in a covariance matrix.

Finally, we use a covariance matrix distance to learn an optimal similarity metric from a training set. Our metric learning is based on the term frequency inverse document frequency (tf-idf) approach which yields an importance weight for different time windows.

In the testing step, given an unknown motion sequence, we first normalize and compute its time windows descriptors. Then, we classify the sequence using a K-nearest neighbors (KNN) voting scheme which returns a matching score for each class thus picking the highest scoring class as the classification result.

4 Technical Details

4.1 Preprocessing and notations

Our human gait acquisition setup consists of a single $Kinect^1$ depth camera mounted on a tripod positioned in a normally lighted room. Humans are walking naturally along the rooms length as the camera captures their walking motion in a frontal view with their walking direction. This allows capturing longer sequences as the depth field is 4.5 meters while the field of view is limited to 0.5 meters.

$Kinect^1$ allows capturing a human at 15 frames per seconds and extracting its skeleton consisting of 25 joints in time. Note that although better acquisition setups may be assembled, a simple and generic low-cost acquisition is adequate as our focus here is a robust and efficient gait processing method.

Given a motion sequence in time $t \in [1 \dots T]$, it consist of a set of skeletons, each of which is defined by 25 joints.

We denote the i^{th} joint at time t:

$$p_i^t = \begin{bmatrix} x_i^t \\ y_i^t \\ z_i^t \end{bmatrix}$$

Due to the frontal view direction, the skeleton size grows as the subject moves towards the camera. To obtain a scale invariant skeleton, we follow the observation that the euclidean distance between the shoulder centre and the hip centre should be fixed when the subject walks [19]. Therefore, we normalize the skeleton joints w.r.t. their distance from the hip center $\|p_i^t - p_{hipCenter}^t\|$.

We further normalize the skeletons to a common size by computing an average skeleton in terms of its skeletal parts and normalizing all skeletons to have the same parts length and thus yielding a scale invariant representation. Finally, we also rotate skeletons to a common global orientation by aligning the vector connecting the two hips with the global X-axis.

$Kinect$ typically generates noisy trajectories due to imperfect acquisition and distance from camera. Therefore, we apply a noise removal step in which we remove the highest frequencies in the trajectory data using a low pass filer. Following [19], we transform trajectories to frequency domain using a discrete cosine transform (DCT) and filter the k highest frequencies ($k = 50\%$) and transform back the smoothed signal using inverse DCT.

A motion sequence encodes joint transformations in time and may depend on the motion cycle and speed. To overcome walking-cycle alignment and speed normalization problems, we segment each motion sequence into subsequences by automatically sliding a fixed sized windows and cutting the motion frames along time. Thus we represent a motion sequence by multiple windows of small size, yielding an walking-cycle and speed-invariant representation.

4.2 Gait covariance descriptor

To compute our gait descriptor, we encode body joints trajectories using a covariance matrix which computes the correlation between different joints in time in the spirit of [19].

To obtain invariance to absolute body positions, we use a local relative coordinate system to represent joints. Thus, we use the hip center as the origin of the local coordinate system and represent the remaining 24 skeletal joints there, yielding a normalized set of joints in relative coordinates per frame denoted:

$$\bar{p}^t = \begin{bmatrix} \bar{x}_1^t, ..., \bar{x}_{24}^t \\ \bar{y}_1^t, ..., \bar{y}_{24}^t \\ \bar{z}_1^t, ..., \bar{z}_{24}^t \end{bmatrix}$$

Thus, the covariance matrix of skeletal joints positions $\bar{p}$, captures the relative positions of joints along the motion trajectories and is defined as:

$$Cov(\bar{p}) = \frac{1}{N-1} \sum_{t=1}^{T} (\bar{p}^t - \mu)(\bar{p}^t - \mu)^{\mathrm{T}}$$

where, $t \in [1, T]$, $\bar{p}^t$ is a 3×24 dimensional feature point at time t, $N = 24$, and μ is the mean of $\bar{p}^t$.

The positional covariance matrix considers only mutual joint positions along time. This is mainly efficient in representing gaits with a strong spatial variance. For example it may easily recognize very distinct joint trajectory patterns such as running vs. walking, raising a hand and etc. In our case, Parkinson and other related neuro-degenerative diseases consist motion deficiencies with very complex patterns. Motion patterns such as Parkinsonian tremors have very little spatial variance. Therefore, taking only the spatial pattern of motion is insufficient for classification of such subtle patterns.

Thus, we enhance our motion representation with joints motion rate, i.e. speed. Similar to the positional covariance matrix, we encode relative speeds of joints in a covariance matrix. Let g be the gradient set of all joints, then:

$$g_i^t = \begin{bmatrix} x_i^t - x_i^{t-1} \\ y_i^t - y_i^{t-1} \\ z_i^t - z_i^{t-1} \end{bmatrix}$$

where i = [1,2,...,24].

The covariance matrix of g denoted $Cov(g)$ is:

$$Cov(g) = \frac{1}{N-1} \sum_{t=1}^{T} (g^t - \mu)(g^t - \mu)^{\mathrm{T}}$$

where, $t \in [1, T]$, g^t is a 24 dimensional point at time t, $N = 24$, and μ is the mean of g^t.

4.3 Learning gait similarity

We apply a supervised learning approach where we initially learn a classifier from a set of known motion exemplars. Thus, we are given a set of pre-classified motion windows which we cut from longer sequences belonging to Parkinsonian, Hemiplegia and normal walking gaits. Windows segment motion sequences into arbitrary sets and therefore some windows may contain distinctive gait features while others may contain insignificant information. Thus, to measure a window's importance, we take a term frequency inverse document frequency (tf-idf) approach with regards to the window dissimilarity value.

To compute the dissimilarity between two covariance matrices we use the matrix distance formulation introduced in [10]:

$$\delta(Cov(a), Cov(b)) = \sqrt{\sum_{i=1}^{n} ln^2 \lambda_i(Cov(a), Cov(b))}$$

where $\lambda_i(Cov(a), Cov(b))$ are the generalized eigenvalues of $Cov(a)$ and $Cov(b)$ that satisfy $Cov(a)x = \lambda Cov(b)x$, and x is the corresponding generalized right eigenvector. The dissimilarity measure between two symmetric positive definite matrices $Cov(a)$ and $Cov(b)$ satisfies the following:

$\delta(Cov(a), Cov(b)) \geq 0$ and $\delta(Cov(a), Cov(b)) = 0$ only if $Cov(a) = Cov(b)$
$\delta(Cov(a), Cov(b)) = \delta(Cov(b), Cov(a))$
$\delta(Cov(a), Cov(b)) + \delta(Cov(b), Cov(c)) \geq \delta(Cov(a), Cov(c))$

Having defined a covariance matrix distance metric, we follow to learning a distinctive classifier for time windows using tf-idf. Given a training window w_q^c where c denotes its class and q indicates its index (among class windows), $tf(w_q^c)$ is the term frequency that measures the occurrence frequency (tf) of w_q^c in class c:

$$tf(w_q^c) = \sum_{i=1}^{|w^c|} \frac{sim(w_q^c, w_i^c)}{|w^c|}$$

where $sim()$ is the binary distance between two windows covariance matrices. For a window w_q^c its positional p and speed g matrices are simply concatenated and termed $Covw_q^c$ for simplicity:

$$sim(w_q^c, w_i^c) = 1 \quad \text{if} \quad \delta(Covw_q^c, Covw_i^c) < \varepsilon$$
$$= 0 \quad \text{otherwise}$$

$|w^c|$ is the number of windows in class c.

Next we measure the distinctiveness of the window by measuring the inverse frequency of its appearance in other classes (idf), by counting the number of occurrences of w_q^c in all classes:

$$idf(w_q^c) = \log \frac{|c|}{\sum_j^{|c|} occ(w_q^c, j)}$$

where $occ(w_q^c, j)$ is defined as:

$$occ(w_q^c, j) = 1 \quad \text{if} \quad \sum_k^{|w^j|} sim(w_q^c, w_k^j) > 0$$
$$= 0 \quad \text{otherwise}$$

Finally, for each window in our training set w_q^c , we learn its classification power, i.e. compute its distinctiveness w.r.t. class c in the form of a weight:

$$weight(w_q^c) = tf(w_q^c) \times idf(w_q^c)$$

4.4 Gait classification

Given a test motion sequence s, we first segment it into fixed sized time-windows of size h in the number of frames. In our experiments, we evaluate for both different h values as well as for overlapping and non-overlapping windows. For each window we then compute its joints positional and speed covariance matrices.

Next, for each time window in the test sequence $w_q^* \in s$ ($*$ denotes an unknown class), we use k-nearest-neighbors using our matrix similarity metric $\delta()$ to compute the k-closest nearest neighbors $k\_closest(w_q^*)$ in the training data (we evaluate for different k in our experiments).

Classification performs per class c, using an averaged weighted distance of w_q^* to its k-closest neighbors in c as:

$$w\_distance(w_q^*, c) = \sum_{w_p^c \in k\_closest(w_q^*)} \frac{weight(w_p^c) \cdot \delta(Cov(w_q^*), Cov(w_p^c))}{|w_p^c|}$$

where $|w_p^c|$ is the number of windows in the knn set of w_q^* of class c.

Finally, we compute the class similarity of s and c, denoted $class\_sim(s, c)$ as the average weighted distance of all windows in s to their knn windows in c:

$$class\_sim(s, c) = \frac{1}{|w_q^* \in s|} \sum_{w_q^* \in s} w\_distance(w_q^*, c)$$

and classify s as the class obtaining the maximal $class\_sim(s, c)$ (see algorithm pseudo-code in 1).

Algorithm 1 Gait Classification.

Input: classes $C = \{Normal(no), Hemiplegia(he), Parkinson(pa)\}$;
training windows set: w^{no}, w^{he}, w^{pa}
test sequence s;
Output: class of s;

1: **for** each window $w_q^* \in s$ **do**
2: set $k\_closest(w_q^*)$ as the k nearest neighbor windows of w_q^*;
3: **for** each class $c \in C$ **and** $w_p^c \in k\_closest(w_q^*)$ **do**
4: $w\_distance(w_q^*, c)$+=$weight(w_p^c)\delta(Cov(w_q^*), Cov(w_p^c))$
5: $class\_sim(s, c) += w\_distance(w_q^*, c)$
6: set class of s as $\max_c class\_sim(s, c)$

5 Results

To test our method, we have captured with a $Kinect^1$ a dataset of walking persons in different disease classes. Our setup consists of a normal sized room of $5X6meters$, with normal daylight, in which we position the $Kinect^1$ camera on a tripod at the height of $1.0meters$. The captured data

was processed on a computer with Intel Core i7 312GHZ with 64GB RAM. Computation times in terms of training the classifier are in the range of few seconds.

To exploit the camera's depth of field of $4.5meters$ we captured humans in a frontal view as they walk towards and away the camera, thus allowing to capture long sequences. Thus, the number of the frames per walking sequence was between 40 and 120. Differences between sequences lengths were due to different step sizes and speed among difference people.

Dataset captures three different classes of walking people: normal (46 to 75 years old), Hemiplegia (50 to 83 years old) and Parkinson (60 to 85 years old). For each class, 14 different people were captured walking along a line and facing the Kinect camera, at normal speed, wearing usual shoes and clothes. For proper evaluation we have captured for each person 5 independent walking sequences resulting in a 5×14 walking sets. Data also consists additional information on the age, gender, pathogenic condition, clinical characteristic of the subjects for an enhanced representation of the data.

Figure 3 consists excerpts from this dataset, showing three walking humans of the three classes as they walk towards the camera. The extracted skeleton which we use in our processing is shown on the right.

5.1 Classification accuracy

To measure classification results we use the accuracy definition of $acc = \frac{tp+tn}{tp+fp+tn+fn}$, where tp is the number of true positives, fp is the number of false positives, tn is the number of true negatives, and fn is the number of false negatives.

The parameters used in our experiments are:

- h determines the time-window size, i.e. the segmentation step size of the sequence.
- k determines the number of nearest neighbors
- ε determines the threshold in the binary distance ($sim()$) between two windows covariance matrices. Specifically, it defines a percent in the relative ordered distances.

In all our experiments, the parameters values we used are $[\varepsilon, k, h] = [0.3, 10, 10]$. Out of 14 humans per disease, we use 8 as training set and 6 for testing. There are possible $\binom{14}{8}$ different selections of training and testing sets. In practice we run 100 times, and each time randomly select a different training and testing set.

Table 1 summarizes our classification accuracy for each of the classes. Hemiplegia, depending on its severity and type, has a significant effect on body functionalities. For Parkinson, it's cardinal clinical symptoms include bradykinesia, rigidity, rest tremor and disturbances in balance. Nevertheless, some subtle gaits such as distinctive hand temors cannot be observed by our system. This is due to its low capture frame rate and resolution as well as due to the filtering out of high frequencies in the denoising step. Our method obtains good recognition rates for Hemiplegia and has a slightly lower classification accuracy for Parkinson.

	Avg	Best	Worst
Normal	83.3%	94.4%	71.1%
Hemiplegic	77.1%	92.2%	67.8%
Parkinson	76.7%	90.0%	61.1%

Table 1 Classification accuracy

5.2 Comparisons

In Tables 2and 3, we compare our method with the state-of-art methods introduced in [40] and [19]. "Covariance" denotes the covariance matrix method of [19]. "Lie relative"and "Lie absolute" denote the Lie algebra methods introduced in [40]. "Joint relative", "joints quat. (quaternions)" and "joints absolute" are straightforward representations of the skeletal joints gaits. Relative are the concatenation of all bones vectors, starting from the head root node and encoding each vector in its ancestor coordinate system. Quaternions are the concatenation of all joints rotations along the time window in quaternion representation. Absolute are the simple concatenation of joints 3D coordinates.

Our method is better than state-of-the-art techniques for classification of Hemiplegia and Parkinson. In Table 2 we show that our method performs significantly better than others in the case of Hemiplegia while for Parkinson the improvement is of a lower magnitude.

In Table 3, we summarize the average accuracy of the methods for all classes. Note that our method is significantly better, crossing the 90% for the best case and nearly 80% on average. In fact, it outperforms state-of-art methods by 4.7% for average accuracy and by 3.0% for best accuracy.

Note also that our method's average accuracy performs 18.2% better than the average accuracy of the covariance based approach in [19]. This is due to the enhanced representation used by our method utilizing both joints speeds and positions as well as our customized time window segmentation

Figure 4 shows confusion matrices for the 3 classes comparing our method, with Lie relative and Relative joints. Our method obtains a confusion matrix thus predicts better in both positive and negative cases. Specifically, normal data is classified correctly by all methods. Hemiplegia is slightly confused as Parkinson and misclassfied by Lie and relative joints methods. Parkinson is the most challenging and gets lower classification accuracy than the other 2 classes in general. Nevertheless, our method significantly outperforms the others even in this challenging case.

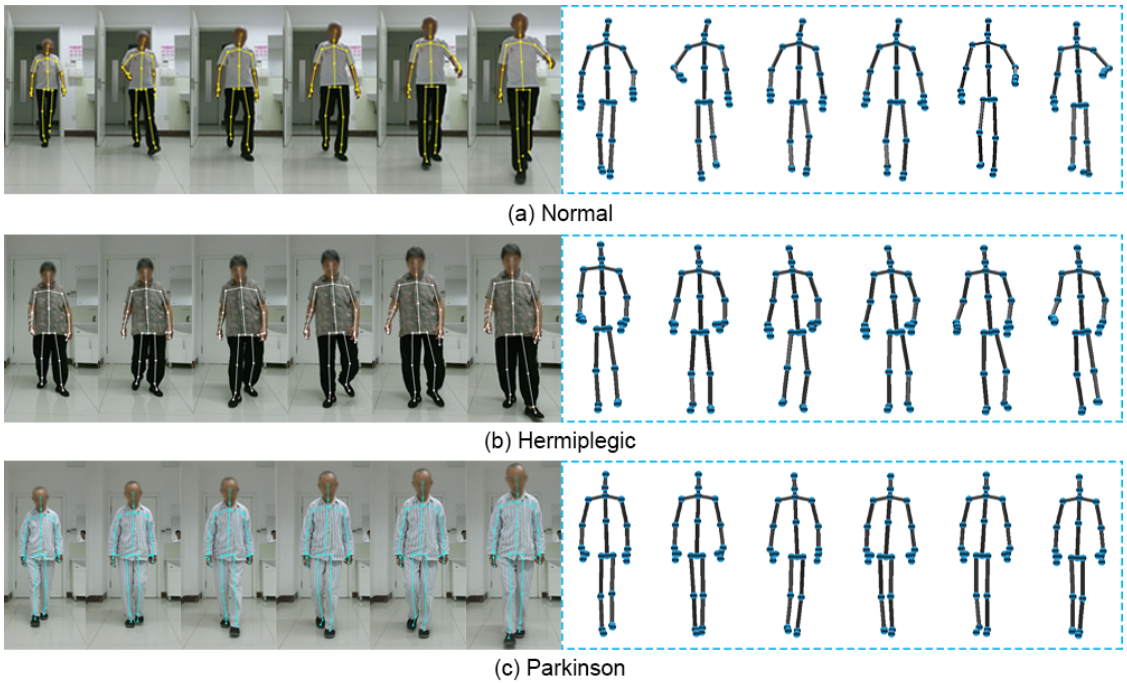


Fig. 3 Three excerpts from our dataset, showing normal, Hemiplegic and Parkinson walking sequences (top-to-bottom rows). Extracted skeletons are on the rightside.

	Normal	Hemiplegic	Parkinson
Ours	83.3%	77.1%	76.7%
Covariance	63.8%	61.1%	57.5%
Lie relative	79.2%	69.3%	74.3%
Lie absolute	74.5%	67.0%	65.2%
Joints relative	78.2%	70.2%	72.6%
Joints quat.	76.1%	67.0%	70.8%
Joints absolute	77.6%	67.4%	70.2%

Table 2 Classification comparison per class

Finally in Figure 5 we compare learning rates and convergence of the 3 methods above, plotting accuracy vs. training time. Note that note that our method's learning rate is the steepest converging faster than the others.

Method	**Avg. Acc.**	**Best Acc.**
Ours	**79.0%**	**91.1%**
Covariance	60.8%	77.8%
Lie relative	**74.3%**	**88.1%**
Lie absolute	68.9%	78.5%
Joints relative	73.7%	85.2%
Joints quat.	71.3%	83.0%
Joints absolute	71.7%	84.5%

Table 3 Classification comparison summary

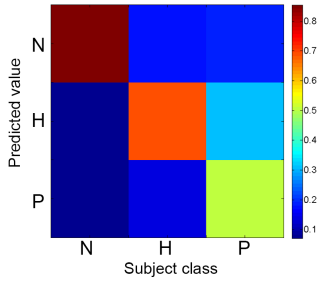
5.3 Robustness evaluation

To thoroughly evaluate our method, we have tested different parameters and aspects of our method.

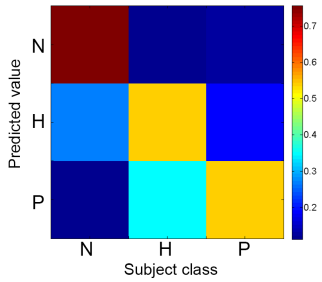
We evaluate the effectiveness of combination of joints positional and speed information in our gait descriptor. Using only joints positions in our gait description (joint positions covariance), the average and best accuracy were 58.3% and 70.0% respectively. Similarly, using only joints speeds in our gait descriptor, the average and best accuracy were 55.0% and 73.3% respectively. As can be seen in Table 3 the combination of both positional and speed descriptors (Ours) improved significantly the performance (in comparison to Covariance).

To compute a robust gait classifier, we choose various combinations of walk sequences for training and testing. In two different experiments, A and B, we have selected different training sets: in A 7 out of 14 and in B 9 out of 14. We have also accounted for a minimum overlap of 2 subjects between the two sets. Classification results were: in A 78.1 and 88.6 for average and best accuracy respectively; in B 80.1 and 89.3 for average and best accuracy respectively. We observe a natural overall increase in the accuracy from 7 to 9 subjects in the training set. Nevertheless the variance is small and the increase rate is moderate, reflecting on our classifier's robustness to the training set.

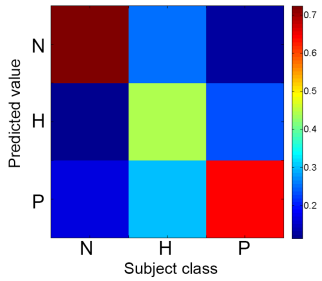
Time window size. We evaluate the influence of time window size on our results in Table 4. Walk cycle refers to cut-



(a) Our method

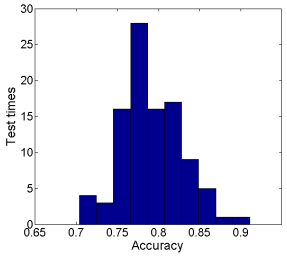


(b) Lie relative

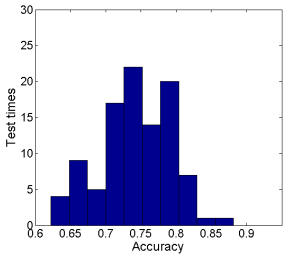


(c) Relative

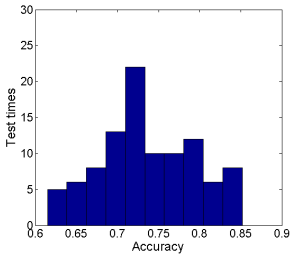
Fig. 4 Confusion matrix of disease classification. N, H and P refer to Normal, Hermiplegy and Parkinson respectively.



(a) Our method



(b) Lie relative



(c) Relative

Fig. 5 Histogram of disease classification results.

Window Size	Avg. Acc.	Best Acc.
3	72.0%	76.1%
10	**79.0%**	**91.1%**
15	75.8%	85.1%
20	75.2%	84.4%
Walk Cycle	74.6%	82.8%
Full sequence	70.7%	80.0%

Table 4 Classification with different time windows .

ting walk sequences into synchronized walk cycles starting from a specific walk configuration and returning to it.

Results show that time-windows of approximately size 10 yield an optimal result. By further increasing the time-window size, we observe a constant decrease in accuracy. This may be explained as large time-windows fail to represent local patterns in the motion due to its large size and the accumulation of the sequence into one large matrix. Similarly, shorter windows than 10 have lower accuracy since small windows may cut significant patterns into simple segments without any distinctive features.

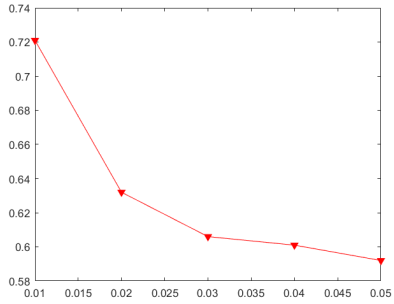
Robustness to noise. In order to evaluate the robustness of our method to noise, we have performed several experiments in which we insert different noise types.

In the first experiment, we insert noise to the positions of the joints in the sequence, testing the robustness of our classifier to such noise. Specifically, we move each joint position a random value in each of the X, Y, Z axes. The values range is defined as a percentage of the body's bounding box diagonal. In Figure 6(a), the graph describes our method's accuracy (y-axis) against the noise level inserted to the joints (x-axis). Note that the accuracy decreases moderately together with the increase in the noise magnitude.

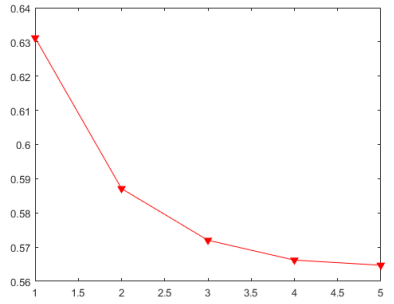
Similarly, in a second experiment, we simulate missing data. This experiment demonstrates the case where motion data cannot be acquired due to occlusions and rapid movement. This may cause parts of the skeleton and even full frames to completely disappear.

To simulate this, we remove a percentage of data from the captured sequences in terms of number of joints 6(b). We demonstrate the relation between our method's accuracy (y-axis) and the percentage of data we remove (x-axis).

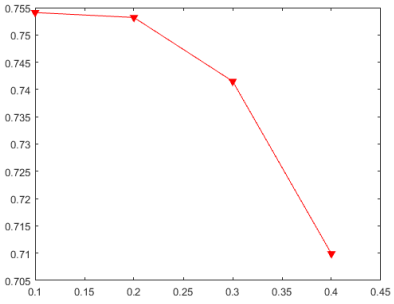
Specifically, in each sequence, we remove a random percent of joints in the range of the current level. In Figure 6(c) we randomly remove entire frames from the sequence up to a percent (x-axis) Note that in both cases the accuracy drops at a moderate rate.



(a) Joints noise



(b) Missing joints



(c) Missing frames

Fig. 6 Robustness to noise.

6 Discussion and conclusions

In this paper, we study the problem of identifying 2 different degenerative diseases by using 3D human skeleton and trajectories of joints captured by a Kinect sensor. We propose an enhanced gait representation which considers both positional and speed data as well as segmenting the motion sequence into optimal windows. Experiments on real data validate the effectiveness of the proposed in comparison with state-of-the-art.

Limitations. Obviously the most prominent limitation of our method is the low resolution of the data acquisition which hamper representation and recognition of subtle motions. These are of significant importance in gait anomalies and their better acquisition may result in accuracy improvement. Another limitation of our system is the field of view which is limited to the camera's depth of field. This may be short in terms of natural walking and longer paths may be required. Nevertheless, this limitation may be attended relatively easy by combining and registering several Kinects together to obtain a larger area coverage.

Future work. A possible extension of our work is to exploit 2D image features to improve the recognition rate. Kinect sensor cannot capture accurately subtle movements of joints. For example the hand tremor, which is the distinguished feature of Parkinson, may not fully reflect in the Kinect skeletal data. This weakness might be complemented by analyzing a corresponding 2D HD and high FPS video. Another possible future work is to extend our method to recognize more diseases which have obvious clinical manifestations in human gait such as Ataxia as well as recognizing the severity and level of each degenerative disease. It is meaningful to develop a real system for the recognition and classification of degenerative diseases to help medical personnel in their examination and diagnosis both qualitatively and quantitatively.

References

1. M. Ahmed, N. Al-Jawad, and A. Sabir. Gait recognition based on kinect sensor. In *SPIE Photonics Europe*, pages 91390B–91390B. International Society for Optics and Photonics, 2014.
2. N. Akae, A. Mansur, Y. Makihara, and Y. Yagi. Video from nearly still: An application to low frame-rate gait recognition. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 1537–1543, 2012.
3. I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1442–1456, 2011.
4. A. Ball, D. Rye, F. Ramos, and M. Velonaki. Unsupervised clustering of people from'skeleton'data. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 225–226. ACM, 2012.
5. N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos. Gait recognition using dynamic time warping. In *Multimedia Signal Processing, 2004 IEEE 6th Workshop on*, pages 263–266. IEEE, 2004.
6. P. Brodal. *The Central Nervous System: Structure and Function*. Oxford University Press, 2010.
7. P. Chattopadhyay, S. Sural, and J. Mukherjee. Gait recognition from front and back view sequences captured using kinect. In *Pattern Recognition and Machine Intelligence*, pages 196–203. Springer, 2013.

8. E. Cippitelli, S. Gasparrini, S. Spinsante, and E. Gambi. Kinect as a tool for gait analysis: Validation of a real-time joint extraction algorithm working in side view. *Sensors*, 15(1):1417–434, 2015.
9. B. Dikovski, G. Madjarov, and D. Gjorgjevikj. Evaluation of different feature sets for gait recognition using skeletal data from kinect. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*, pages 1304–1308. IEEE, 2014.
10. W. Förstner and B. Moonen. *A Metric for Covariance Matrices*, pages 299–309. 2003.
11. M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1964–1967. IEEE, 2012.
12. B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester. Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & posture*, 39(4):1062–1068, 2014.
13. J. Han and B. Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):316–322, 2006.
14. J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of applied physiology*, 88(6):2045–2053, 2000.
15. J. M. Hausdorff, D. A. Rios, and H. K. Edelberg. Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Archives of physical medicine and rehabilitation*, 82(8):1050–1056, 2001.
16. G. Johansson. *Visual Motion Perception*. Scientific American offprints. Freeman, 1975.
17. T. Kahn and D. Nyholm. Computer vision methods for parkinsonian gait analysis: A review on patents. *Recent Patents on Biomedical Engineering*, 2013(6):97–108, February 2013.
18. N. L. Keijsers, M. W. Horstink, and S. C. Gielen. Ambulatory motor assessment in parkinson's disease. *Movement Disorders*, 21(1):34–44, 2006.
19. M. Kumar and R. V. Babu. Human gait recognition using depth camera: a covariance based approach. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, page 20. ACM, 2012.
20. T. H. Lam, R. S. Lee, and D. Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9):2563–2573, 2007.
21. L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 148–155. IEEE, 2002.
22. X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 38(2):145–155, 2008.
23. L.-F. Liu, W. Jia, and Y.-H. Zhu. Survey of gait recognition. In *Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence*, pages 652–659. Springer, 2009.
24. J. Lu, G. Wang, and T. S. Huang. Gait-based gender classification in unconstrained environments. In *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, November 11-15, 2012*, pages 3284–3287, 2012.
25. J. Lu, G. Wang, and P. Moulin. Human identity and gender recognition from gait sequences with arbitrary walking directions. *Information Forensics and Security, IEEE Transactions on*, 9(1):51–61, 2014.
26. Y. Makihara, H. Mannami, and Y. Yagi. Gait analysis of gender and age using a large-scale multi-view gait database. In *Computer Vision–ACCV 2010*, pages 440–451. Springer, 2011.
27. A. Mansur, Y. Makihara, R. Aqmar, and Y. Yagi. Gait recognition under speed transition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 2521–2528, 2014.
28. R. Martín-Félez and T. Xiang. Gait recognition by ranking. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I*, pages 328–341, 2012.
29. H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern recognition letters*, 17(2):155–162, 1996.
30. C. B. Ng, Y. H. Tay, and B.-M. Goi. Recognizing human gender in computer vision: a survey. In *PRICAI 2012: Trends in Artificial Intelligence*, pages 335–346. Springer, 2012.
31. M. Parajuli, D. Tran, W. Ma, and D. Sharma. Senior health monitoring using kinect. In *Communications and Electronics (ICCE), 2012 Fourth International Conference on*, pages 309–312. IEEE, 2012.
32. A. P. Rocha, H. Choupina, J. M. Fernandes, M. J. Rosas, R. Vaz, S. Cunha, and J. Paulo. Parkinson's disease assessment based on gait analysis using an innovative rgb-d camera system. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 3126–3129. IEEE, 2014.
33. P. J. S. Fahn, C. D. Mardsen and P. Teychenne. *Recent developments in Parkinson's disease*. American Neurological Association, 1987.
34. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 1297–1304, 2011.
35. S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. Gait energy volumes and frontal gait recognition using depth images. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE, 2011.
36. E. E. Stone and M. Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 6491–6494. IEEE, 2011.
37. E. E. Stone, M. Skubic, and J. Back. Automated health alerts from kinect-based in-home gait measurements. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 2961–2964. IEEE, 2014.
38. J. Tang, J. Luo, T. Tjahjadi, and Y. Gao. 2.5d multi-view gait recognition based on point cloud registration. *Sensors*, 14(4):6124, 2014.
39. M. Z. Uddin, J. T. Kim, and T.-S. Kim. Depth video-based gait recognition for smart home using local directional pattern features and hidden markov model. *Indoor and Built Environment*, 23(1):133140, 2014.
40. R. Vemulapalli, F. Arrate, and R. Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 588–595. IEEE, 2014.