Spatiotemporal Gait Measurement With a Side-View Depth Sensor Using Human Joint Proposals

Andrew Hynes [ID], Stephen Czarnuch [ID], Megan C. Kirkland [ID], and Michelle Ploughman [ID]

Abstract—We propose a method for calculating standard spatiotemporal gait parameters from individual human joints with a side-view depth sensor. Clinical walking trials were measured concurrently by a side-view Kinect and a pressure-sensitive walkway, the Zeno Walkway. Multiple joint proposals were generated from depth images by a stochastic predictor based on the Kinect algorithm. The proposals are represented as vertices in a weighted graph, where the weights depend on the expected and measured lengths between body parts. A shortest path through the graph is a set of joints from head to foot. Accurate foot positions are selected by comparing pairs of shortest paths. Stance phases of the feet are detected by examining the motion of the feet over time. The stance phases are used to calculate four gait parameters: stride length, step length, stride width, and stance percentage. A constant frame rate was assumed for the calculation of stance percentage because time stamps were not captured during the experiment. Gait parameters from 52 trials were compared to the ground truth walkway using Bland-Altman analysis and intraclass correlation coefficients. The large spatial parameters had the strongest agreements with the walkway (ICC(2, 1) = 1.00 and 0.98 for stride and step length with normal pace, respectively). The presented system directly calculates gait parameters from individual foot positions while previous side-view systems relied on indirect measures. Using a side-view system allows for tracking walking in both directions with one camera, extending the range in which the subject is in the field of view.

Index Terms—Depth sensor, gait analysis, joint proposals, shortest paths, side-view.

Andrew Hynes is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: ajhynes@mun.ca).

Stephen Czarnuch is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada, and also with the Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6, Canada (e-mail: sczarnuch@mun.ca).

Megan C. Kirkland and Michelle Ploughman are with the Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6, Canada (e-mail: megan.kirkland@mun.ca; michelle.ploughman@med.mun.ca).

I. INTRODUCTION

THE analysis of human gait is an important component of treating walking disorders [1], which arise from neurological diseases including cerebral palsy [2] and multiple sclerosis (MS) [3]–[6]. Clinical gait analysis is commonly performed with timed walking tests [7], [8]. For a deeper analysis, quantitative gait measures can be obtained using pressure-sensitive walkways such as GAITRite [5] or the Zeno Walkway [9]. The walkways measure spatial and temporal gait parameters by recording the positions of the feet over time. They can also measure kinetic properties such as the centre of pressure of the foot. However, walkways are unable to directly measure the kinematics of body parts other than the feet. Full-body gait analysis has been performed using sensors attached to the body [10]–[12], or by tracking markers on the body with a motion capture system [13], but these approaches typically require significant setup time, expert knowledge, and specialized locations [14].

Markerless gait analysis has been performed using RGB cameras, depth sensors, and other devices such as laser scanners. Recently, Zago *et al.* [15] measured spatiotemporal parameters without markers using two RGB cameras for stereoscopic vision. Iwai *et al.* [16] used 2D laser range sensors placed at shin height to estimate where the foot contacts the ground, but required subjects to walk barefoot with shins exposed (e.g., wearing shorts). Castelli *et al.* [17] used a single RGB camera to measure spatiotemporal parameters without conventional markers, but this still required subjects to wear white undergarments to mark the pelvis and foot segments, and the subjects had to walk in front of a homogeneous blue background. Other approaches using single RGB cameras also exist (e.g., [18], [19]), but these studies only evaluated angular gait parameters (e.g., dorsiflexion) rather than spatiotemporal. Unlike an RGB camera, a single depth sensor is sufficient for measuring the scene in 3D.

Human pose estimation from depth sensors has recently seen large advances, notably with the release of the Microsoft Kinect [20]. Depth sensors typically provide traditional RGB data as well as depth data (i.e., a measure of distance from the sensor for each pixel in the field of view), providing a 3D understanding of the scene [21]. A large volume of research has now investigated the Kinect as a device for gait analysis [13], [22]–[31]. The advantages of gait analysis with a depth sensor include long-term monitoring in a home setting [23] and tracking

the full body in 3D at a low cost with natural clothing and without wearable sensors or markers.

Gait analysis with the Kinect is often conducted using the Kinect Software Development Kit (SDK) to process the depth images captured by the camera [22], [25]–[28], [30]–[32]. The SDK outputs a skeleton model of the human body with 20 joints at an approximate frame rate of 30 frames per second [22] (we adopt the terminology of [22] and [33] by referring to all these positions as joints, while some are technically segments or parts, such as the head and feet). Behrens *et al.* introduced a computerized measure for gait analysis in persons with MS, the Short Maximum Speed Walk (SMSW) test [32]. The parameters of SMSW were calculated using the positions of the hip-centre joint extracted from the SDK. This new test was found to be correlated with established clinical measures including the Timed 25-Foot Walk. Gabel *et al.* derived a feature vector from multiple consecutive frames of skeleton data from the SDK [22]. A regression model used this vector to predict stride durations and arm angular velocities, which were validated against data from wearable sensors. Gait parameters have been measured concurrently with the Kinect SDK and the GAITRite mat in both children [28] and healthy adults [24], [30].

The Kinect SDK is intended to track the human skeleton from a frontal perspective [29], so one camera is insufficient for tracking a person walking both ways on a walkway. In [30], a Kinect camera was placed at each end of a GAITRite mat so the subject would be tracked from a frontal perspective while walking in either direction on the mat. Participants in [28] walked in only one direction on the GAITRite (towards a Kinect at the end of the mat). A second side-view camera captured depth images for assisting with manual annotation. A frontal view is also inconvenient because the practical range of the depth sensor is less than 4 m. If the entire human shape needs to be seen, the minimum distance from the sensor must be at least 2 m, resulting in a measurable walking distance of at most 2 m [34]. From a side view, the body is visible for the entire horizontal field of view (about 4 m).

In response to the front-view limitations of the SDK, gait analysis with a non-frontal Kinect has been explored. Cippitelli *et al.* presented an algorithm for a side-view Kinect that functions without machine learning [29]. A calibration step was required in which the subject faces the sensor with outstretched arms. The lengths between adjacent body joints were calculated from this calibration image. The system tracked six joints visible on one side of the body (head, shoulder, elbow, hip, knee, and ankle), in order to produce an objective score for the Get Up and Go Test (GUGT), which involves standing from an armless chair and beginning to walk. While the six joints were sufficient for GUGT, spatial gait parameters such as stride length require separate foot positions to be measured directly. Baldewijns *et al.* used the SDK to extract the binary image of the person from a side view, but not to track the full skeleton model [24]. Step length and step time were calculated indirectly by analyzing the centre of mass of the binary image. Stone and Skubic [23] performed continuous and long-term monitoring of older adults with a Kinect mounted in their apartments. A probabilistic model was used to estimate gait parameters rather than tracking a skeleton, limiting the applicability of this approach to clinical gait assessment.

The tracking ability of the Kinect SDK is rooted in a machine learning algorithm developed by Shotton *et al.* [33]. Given a single depth image, the trained system produces multiple proposals for the 3D positions of human joints. Each joint proposal is associated with a confidence value indicating the likelihood that the position is correct. The different human joints are identified independently (i.e., without information from other image frames or the kinematic constraints of the body). Unfortunately, the process from the joint proposals to the final smooth tracking of the human skeleton is a proprietary and unpublished algorithm [35].

We use a predictor developed for overhead hand tracking [36] which is based on the algorithm of Shotton *et al.* [33]. Each pixel in a random subset from a depth image is described with a multiclass probability density function (PDF) and the information in the underlying PDFs is aggregated using a local mode-find approach, resulting in multiple proposals for each tracked part. We retrained our predictor to output multiple joint proposals from side-view depth images of the human body. However, the predictor can generate inaccurate proposals by mistaking one body part for another, mixing up left/right parts, or detecting background noise. Therefore, we present a method to select accurate joints from the proposals which we group by part type, removing the left/right distinction provided by the predictor. The problem of accurately selecting from multiple joint proposals has also been applied to pose estimation in RGB videos [37] and multi-person pose estimation in RGB images [38].

We present three main contributions:

1) A method to select accurate head and foot positions per-frame from multiple joint proposals after estimating the fixed lengths of links between parts. The feet are then assigned to left and right sides based on the direction of walking motion.
2) A method to calculate standard spatiotemporal gait parameters from the left and right foot positions using the same equations used by the Zeno Walkway.
3) A validation against the Zeno Walkway.

Our system is tested on a data set of 52 walking trials recorded at the Recovery and Performance Laboratory, Memorial University. The study was approved by the Health Research Ethics Board of Newfoundland and Labrador (#14.102). Participants with MS were measured concurrently by a Zeno Walkway and a Kinect v1 camera from a side view. Participants were recruited into our study out of convenience during their participation in a separate aerobic exercise study. The camera was positioned 1m above the ground, 2.5m from the midpoint of the 14ft walkway (walkway specifications available from Protokinetics [39]). Each trial consisted of multiple passes along the walkway in both directions. Gait parameters were calculated from the Zeno Walkway data by the Protokinetics Movement Analysis Software (PKMAS), which uses calculations as defined by Huxham *et al.* [40].

II. POSE ESTIMATION

Let $\mathcal{P}$ be the set of all joint proposals on one frame captured by the depth sensor. These proposals are positions in 3D space. $\mathcal{P}$ is partitioned into subsets representing body part types. We
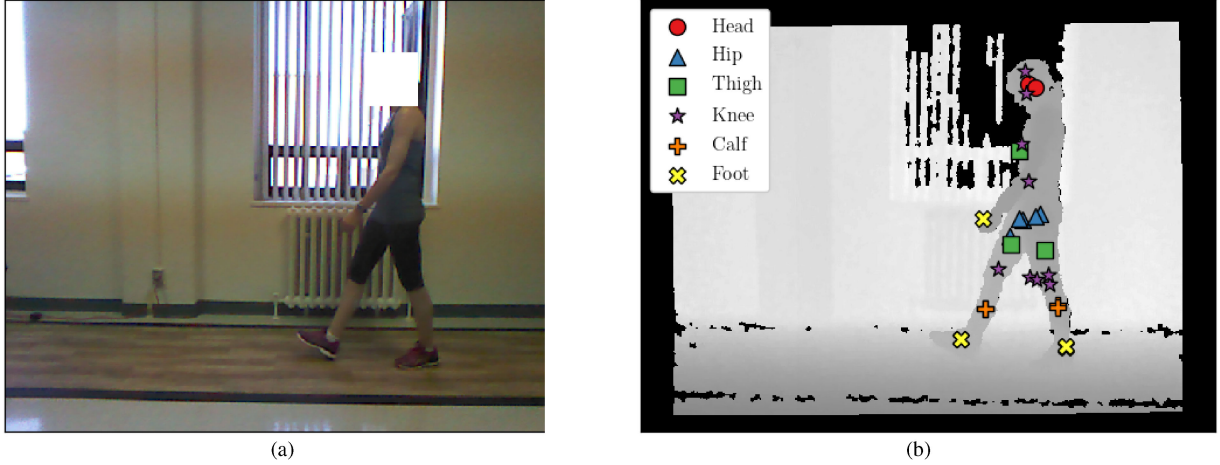
Fig. 1. Example frame in a walking trial. (a) RGB image captured by the Kinect camera. (b) Corresponding depth image showing all joint proposals.

utilize six part types: head, hip, thigh, knee, calf, and foot. Fig. 1 shows a two-dimensional view of the positions in $\mathcal{P}$ labelled by part type.

Our method is based on the assumption that the links between consecutive pairs of parts (head to hip, hip to thigh, etc.) have fixed lengths [41]. This allows us to represent the joint proposals on each frame as a weighted graph, with edge weights dependent on the difference between the expected lengths for the trial and the measured lengths on the frame. A shortest path from head to foot in this graph finds a combination of parts with lengths similar to the expected lengths, by minimizing the cumulative error between the measured and expected lengths. The process of estimating the expected lengths for the trial is explained later in Section II-C.

A. Graph Representation

The joint proposals of $\mathcal{P}$ are represented as the vertices of a weighted graph G. The vertices of part type t form a complete bipartite graph with the vertices of part type $t + 1$ (i.e., there is an edge between each vertex of type t and each vertex of type $t + 1$, and no edges between vertices of the same type [42]). The edges are directed from type t to type $t + 1$. An example of graph G is shown in Fig. 2.

Each proposal i has a 3D position $\mathbf{p}_i$ and a part type t_i. The measured length L_{ij} between two proposals i and j is

$$L_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\| \tag{1}$$

If $t_j = t_i + 1$, then proposals i and j are connected by a directed edge $i \to j$ in G, and there is an expected length $L_{t_i t_j}$ between parts of type t_i and parts of type t_j. As introduced in [41], the weight W_{ij} of the edge is

$$W_{ij} = \left(L_{ij} - L_{t_i t_j}\right)^2 \tag{2}$$

B. Shortest Paths

After the weighted graph G has been constructed, an algorithm is run to find the shortest path to each vertex representing
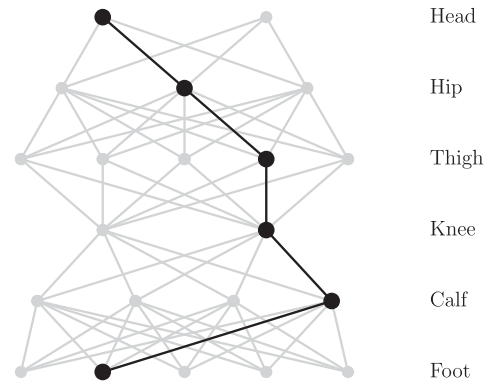

Fig. 2. Graph representation of joint proposals. Each proposal is represented by one vertex in the graph. The black vertices and edges show a possible shortest path from head to foot. The path consists of one proposal for each part type.

a foot proposal. A shortest path between two vertices u and v is a path along the edges from u to v with the lowest possible sum of edge weights [43]. Fig. 2 shows a possible shortest path in G from head to foot.

Since each edge in G is directed from a vertex of part type t to one of type $t + 1$, there are no paths in the graph that can begin and end on the same vertex. Therefore, G is classified as a directed acyclic graph (DAG). We use an algorithm that finds all shortest paths from a single source vertex on a DAG, described in [43]. This is similar to Dijkstra's algorithm, but is specifically optimized for a DAG.

The vertices of the graph must be in topological order before the algorithm is run. A topological ordering is a sequence of vertices such that for each edge $u \to v$, u appears before v in the ordering. A topological ordering for G is obtained by listing the vertices of each part type in order from head to foot.

G can be viewed as a single-source graph where the source vertex is connected to each head vertex with a zero-weight edge. Thus, the algorithm for single-source shortest paths on a DAG can be run on G.

The algorithm finds a shortest path to every vertex in G, but our use of the term 'shortest path' will refer only to a path ending on a foot vertex for the remainder of the paper. The structure of G guarantees that such a path consists of exactly one vertex for each part type, as demonstrated in Fig. 2.

C. Length Estimation

An iterative algorithm is run on the walking trial to estimate the expected lengths. The lengths are first assumed to all be zero. The graph G is weighted using these expected lengths. The shortest path algorithm is executed on G, and the path with the lowest total weight is selected. The measured lengths from this path are recorded for the frame. The algorithm calculates the medians of the measured lengths recorded so far (from the first frame to the current frame). If the medians have not changed from the previous frame, the expected lengths are updated to be these medians. The process restarts from the beginning of the trial with the new expected lengths. The algorithm terminates when each expected length has converged to a stable value. The final expected lengths are then used on the full walking trial for selecting the head and feet.

D. Foot Selection

There are n_{foot} proposals for foot positions in a frame. From these, two must be selected as the best estimates for the actual feet of the walking person.

The simplest solution would be to select the two paths with the lowest total weights. However, it is possible that these paths both emanate from the same actual leg in the image, while a different path is the correct one for the other leg. Another solution would be to select the two paths with foot positions furthest from each other. This would be effective if all the foot proposals emanated from the actual two feet. But a noisy (highly inaccurate) foot proposal can be generated on a frame, and this approach would tend to select that proposal.

Therefore, the challenge is to select two paths that are not too close together and yet do not end on noisy foot positions. Our solution is an algorithm which evaluates the paths in pairs rather than individually. The links between proposals are now given scores rather than weights, and a pair of paths receives a total score. The algorithm finds the pair with the highest total score over a series of rounds. Paths that are spaced apart will achieve a higher score than paths that are close together, but only if they do not contain noisy proposals.

A subset of proposals, $\mathcal{P}_{paths}$, is taken from the set of all joint proposals $\mathcal{P}$. A proposal is in $\mathcal{P}_{paths}$ if it is included in any of the shortest paths. Many noisy joint proposals are absent in $\mathcal{P}_{paths}$, as evident in Fig. 3.

A score S_{ij} is assigned to the link between two proposals i and j in $\mathcal{P}_{paths}$ if both of the following conditions are met:
1) There is a fixed length between the part types t_i and t_j.
2) Proposals i and j are on the same shortest path.

The score is calculated with a simple quadratic function.

$$S_{ij} = -(x-1)^2 + 1 \qquad (3)$$

where x is the ratio between the measured length L_{ij} and expected length $L_{t_i t_j}$. The ratio is calculated by dividing the
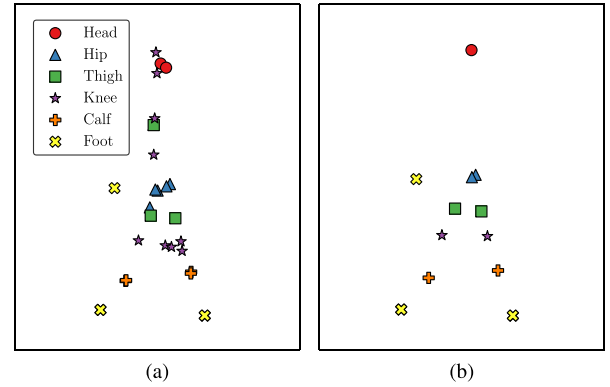

Fig. 3. Removing noisy joint proposals. (a) $\mathcal{P}$, the set of all joint proposals on a frame. (b) $\mathcal{P}_{paths}$, a subset of $\mathcal{P}$. A proposal is in $\mathcal{P}_{paths}$ if it is included in any of the shortest paths to the feet.
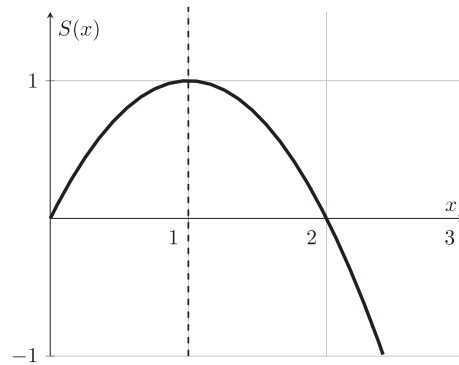

Fig. 4. Plot of the score function defined in Equation 3. The dashed vertical line marks the beginning of the domain $x \geq 1$.

greater length by the lesser length, so that $x \geq 1$.

$$x = \frac{\max(L_{ij}, \ L_{t_i t_j})}{\min(L_{ij}, \ L_{t_i t_j})} \qquad (4)$$

The links between proposals with consecutive part types are the only links represented by edges in G. However, there are two additional links of fixed length: hip to knee and knee to foot. The expected length from hip to knee is calculated as the sum of the expected lengths from hip to thigh and thigh to knee, since all three parts should lie in a straight line. The same applies to the knee, calf, and foot. Scores are assigned to these additional links as well as the links represented by edges in G.

Like the edge weight W_{ij} in G, the score S_{ij} is dependent on the expected and measured lengths between joint proposals i and j. While W_{ij} is restricted to non-negative values (a consequence of Equation 2), S_{ij} can be positive or negative. The highest possible score is one, occurring when the measured length equals the expected length. The score becomes negative when the ratio of the lengths is greater than two. The quadratic function defined in Equation 3 is shown in Fig. 4.

Once the scores are assigned, all possible pairs of shortest paths are compared. Algorithm 1 summarizes the process to select the best pair of shortest paths. $\mathcal{P}_{pair}$ is the set of positions included in a pair of paths. A sphere of radius r is centred on each position in $\mathcal{P}_{pair}$. If positions $\mathbf{p}_i$ and $\mathbf{p}_j$ from $\mathcal{P}_{paths}$ both
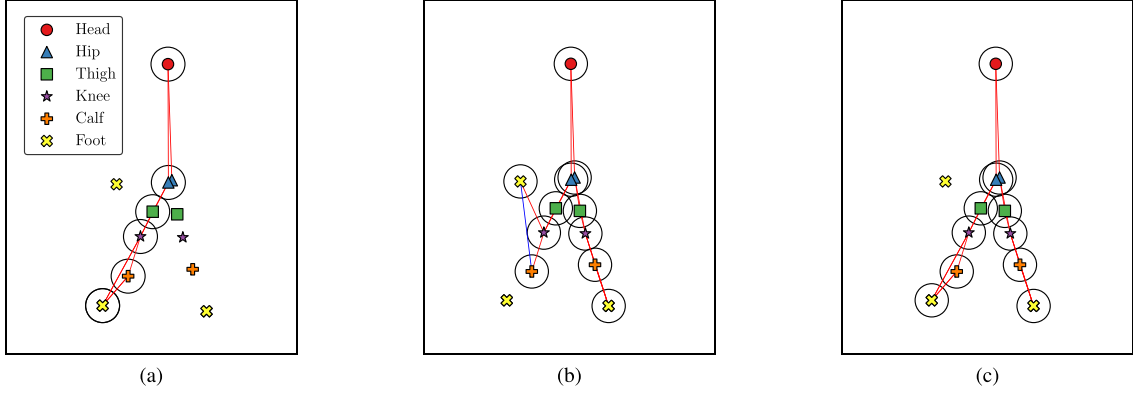
Fig. 5. Comparing pairs of shortest paths. The visible points constitute $\mathcal{P}_{paths}$, a subset of $\mathcal{P}$ (see Fig. 3). A sphere of radius r is centred on each joint proposal in a pair of paths. The lines between points indicate links with non-zero scores, ranging from negative (blue) to positive (red). (a) The two paths end on nearby foot positions. Several links with good scores (from the other leg) are missed by the spheres. (b) One path ends on a noisy foot position. The link from calf to foot has a negative score, reducing the total score of the pair. (c) The spheres include links with good scores from both sides of the body. This pair of paths has the highest total score for radius r.

Algorithm 1: Selecting the Best Pair of Shortest Paths.

Input:
 $pairs$ All pairs of shortest paths
 $radii$ Array of radii for the spheres
 $\mathcal{P}_{paths}$ Set of proposals along the paths
 S Matrix of scores between proposals
Output: $pair_{best}$ Best pair of shortest paths

1: **function** SelectBestPair($pairs$, $radii$, $\mathcal{P}_{paths}$, S)
2: $n_p \leftarrow$ number of pairs
3: $votes \leftarrow$ array of n_p zeros
4: **for** $r \in radii$ **do**
5:   $scores \leftarrow$ array of n_p zeros
6:   **for** $pair \in pairs$ **do**
7:    $\mathcal{P}_{pair} \leftarrow$ Set of positions in $pair$
8:    $V_{spheres} \leftarrow$ combined volume of spheres centred on positions in $\mathcal{P}_{pair}$
9:    $s_{total} \leftarrow 0$
10:    **for** $\mathbf{p}_i \in \mathcal{P}_{paths}$ **do**
11:     **for** $\mathbf{p}_j \in \mathcal{P}_{paths}$ **do**
12:      **if** $\mathbf{p}_i$ and $\mathbf{p}_j$ are both in $V_{spheres}$ **then**
13:       $s_{total} \leftarrow s_{total} + S_{ij}$
14:    $scores[pair] \leftarrow s_{total}$
15:   $winners_r \leftarrow$ all pairs with a score equal to $\max(scores)$ for radius r
16:   $votes(winners_r) \leftarrow votes(winners_r) + 1$
17: $pair_{best} \leftarrow pairs[\arg\max(votes)]$
18: **return** $pair_{best}$

---

lie inside the combined volume of spheres, the score S_{ij} is added to the total score for the pair of paths (note that $S_{ij} = 0$ unless $\mathbf{p}_i$ and $\mathbf{p}_j$ are on the same path). Fig. 5 shows spheres of one radius on different pairs of paths, and the links with non-zero scores included by these spheres.

Negative scores discourage the selection of a noisy foot proposal. Consider the scenario where the two correct foot proposals are close together while an incorrect proposal is far away. If scores were restricted to positive values, the link to a noisy foot proposal would have a small positive score, still contributing to the total score. When the score is negative, the noisy proposal causes a net decrease in the total score.

After a total score has been calculated for each pair of paths, a vote is given to the pair with the highest score. In the case of a tie, a vote is given to each pair tied for the top score. The process repeats with a new radius for the spheres, and the votes for the pairs are accumulated.

When the votes have been counted over a range of radii, the pair with the most number of votes is selected. The two foot positions from this pair are deemed to be the best estimates for the actual feet on the frame.

E. Head Selection

When a frame includes multiple head proposals, the two shortest paths selected in Section II-D could include two different head proposals. When this occurs, the path with the lower total weight defines the selected head position.

III. GAIT ANALYSIS

A. Walking Passes

The walking trials have a varying number of passes along the Zeno Walkway. For each pass, participants enter the field of view, walk along the walkway, and exit on the opposite side. This process ensured a number of empty frames between each pass. To identify the passes, the indices of the non-empty frames in a trial are clustered with DBSCAN (density-based spatial clustering for applications with noise) [44], which determines the number of clusters automatically. Each detected cluster of frame indices is treated as one walking pass. DBSCAN also labels data points as noise if they are too far from the core clusters. Any frames identified as noise are excluded from the following calculations.

B. Orientation

The general direction of the walking pass is needed for detecting stance phases and assigning them to left and right feet.
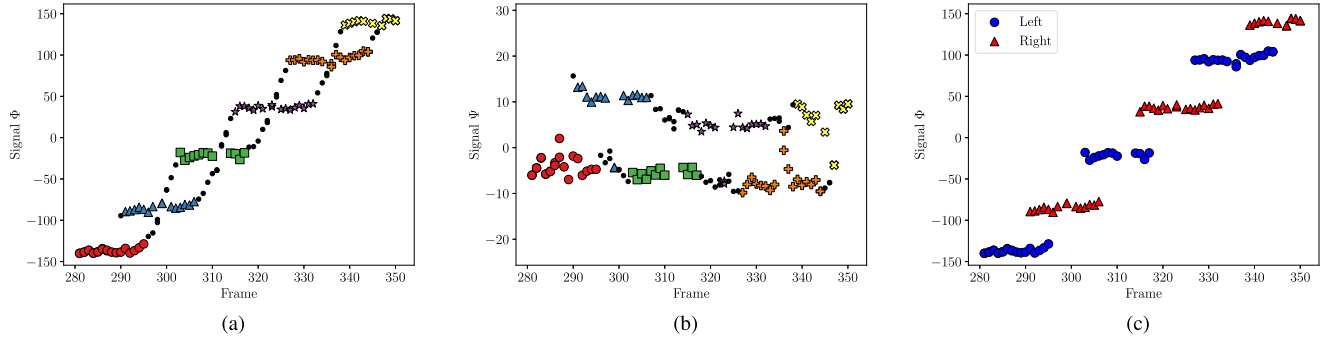
Fig. 6.   Detecting stance phases and assigning left/right sides. (a) The stance phases are detected by clustering the values of signal Φ. The unique clusters are shown with different colours, while values marked as noise are shown in black. (b) The signal Ψ for the same walking pass. This signal is used to assign sides to the clusters. High values of Ψ indicate that a cluster is on the right. (c) The detected clusters of Φ are shown again after assigning sides. These correspond to the stance phases of the left and right feet.

Because the foot positions are 3D, an up direction is also needed to assign left and right sides.

$\mathbf{P}_{foot,A}$ and $\mathbf{P}_{foot,B}$ are the arrays of selected foot positions for the walking pass. The vector from the mean foot position to the head position is computed on each frame. The overall up direction $\hat{\mathbf{v}}_{up}$ is the median of these vectors (as a unit vector).

$$\mathbf{P}_{foot,mean} = \left(\mathbf{P}_{foot,A} + \mathbf{P}_{foot,B}\right)/2$$

$$\mathbf{V}_{up} = \mathbf{P}_{head} - \mathbf{P}_{foot,mean}$$

$$\hat{\mathbf{v}}_{up} = \mathrm{unit}(\mathrm{median}(\mathbf{V}_{up})) \qquad (5)$$

A new array $\mathbf{P}_{foot}$ is obtained by interweaving the rows of $\mathbf{P}_{foot,A}$ and $\mathbf{P}_{foot,B}$ to ensure temporal order. The frames of the walking pass are grouped similarly.

A linear model is fitted to $\mathbf{P}_{foot}$ with the RANSAC (random sample consensus) algorithm [45]. RANSAC is an iterative algorithm which evaluates the goodness of fit for multiple random samples of the data. The output linear model is defined by a point $\mathbf{p}_{centroid}$ and a unit vector $\hat{\mathbf{v}}_{forward}$. The vector estimates the general forward direction of the walking pass.

Finally, the perpendicular vector $\hat{\mathbf{p}}_{perp}$ is the cross product of the up and forward vectors.

$$\hat{\mathbf{v}}_{perp} = \hat{\mathbf{v}}_{up} \times \hat{\mathbf{v}}_{forward} \qquad (6)$$

The RANSAC algorithm also classifies each point as an inlier or outlier. The array of grouped foot points $\mathbf{P}_{foot}$ is revised to only contain inlier points for further calculations.

C. Stance Phase Detection

The stance phases of the feet are identified by analyzing the motion of the feet over time. The displacement of a foot is expected to be close to zero for a stance phase since the foot is planted on the floor. The forward direction $\hat{\mathbf{v}}_{forward}$ is used to create a one-dimensional signal for detecting the phases. $\mathbf{V}_{foot}$ is the set of vectors from $\mathbf{p}_{centroid}$ to the grouped foot positions in the pass, $\mathbf{P}_{foot}$.

$$\mathbf{V}_{foot} = \mathbf{P}_{foot} - \mathbf{p}_{centroid} \qquad (7)$$

The signal Φ is found by taking the dot product of the direction vector $\hat{\mathbf{v}}_{forward}$ with each vector in $\mathbf{V}_{foot}$. This transforms the array of vectors into a 1D array of values, which are coordinates along the line of walking motion. If the line of motion is perpendicular to the camera, the values are analogous to x coordinates in the camera view.

$$\Phi = [\hat{\mathbf{v}}_{forward} \cdot \mathbf{v} \mid \mathbf{v} \in \mathbf{V}_{foot}] \qquad (8)$$

Fig. 6 shows a scatter plot of the signal with frames on the x-axis. Because the foot points A and B have been grouped together in $\mathbf{P}_{foot}$, there are two values in Φ for each frame (unless one of the foot points was marked as an outlier by RANSAC).

The stance phases are detected by clustering the one-dimensional values of Φ with DBSCAN. Each cluster returned by the algorithm is a unique stance phase, while values marked as noise correspond to swing phases.

Since the values of the signal are also temporal in nature, we use one of modifications proposed in ST-DBSCAN [46], which adapts DBSCAN for spatiotemporal data. The original algorithm requires two parameters: min_pts, the minimum number of neighbours a point needs to be considered a core point, and ϵ, the radius defining the spatial neighbourhood of the point. The modification introduces a third parameter ϵ_t, which defines the temporal neighbourhood. The final neighbourhood is the intersection of the spatial and temporal neighbourhoods. Thus, if a foot point generates a value in Φ that is similar to a cluster in space but not in time, it will be correctly marked as noise.

D. Side Assignment

The detected stance clusters are now assigned to left and right sides. A second one-dimensional signal, Ψ, is calculated using $\hat{\mathbf{v}}_{perp}$ instead of $\hat{\mathbf{v}}_{forward}$. The values of Ψ are analogous to z (depth) coordinates in the camera view if the subject is walking perpendicular to the camera.

$$\Psi = [\hat{\mathbf{v}}_{perp} \cdot \mathbf{v} \mid \mathbf{v} \in \mathbf{V}_{foot}] \qquad (9)$$

Each cluster is independently assigned to the left or right side. The value $v_{side,stance}$ is the median of Ψ values in the cluster. As
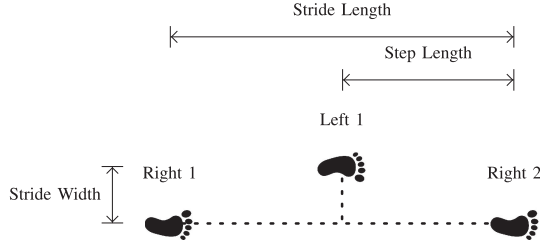
Fig. 7.   Spatial gait parameters for right foot 1.

TABLE I
EXAMPLE OF STANCE PHASES ORDERED BY INITIAL FRAME

Stance no.	f_{first}	f_{last}	Position	Side
0	281	295	[157, -90, 267]	R
0	291	306	[103, -90, 280]	L
1	303	317	[40, -89, 259]	R
1	315	332	[-17, -86, 266]	L
2	327	342	[-73, -82, 249]	R
2	339	350	[-121, -78, 262]	L

described in the previous section, there are usually two values of the signal on one frame, emanating from foot points A and B. Therefore, $v_{side,swing}$ is calculated as the median of Ψ values that correspond to frames in the cluster but are not in the cluster themselves. If none of these values exist, $v_{side,swing}$ is assumed to be zero. All clusters are initially assumed to belong to the left foot. If $v_{side,stance} > v_{side,swing}$, the cluster is assigned to the right foot instead.

E. Gait Parameters

PKMAS defines a stride as the first contact of one foot on the floor to the proceeding first contact of the same foot. The other foot is in a stance phase during this stride. Thus, stride i for foot a is defined by three positions: $\mathbf{p}_{a,i}$, $\mathbf{p}_{b,i}$, and $\mathbf{p}_{a,i+1}$, illustrated in Fig. 7. In this diagram, $\mathbf{p}_{a,i}$ is right foot 1, $\mathbf{p}_{b,i}$ is left foot 1, and $\mathbf{p}_{a,i+1}$ is right foot 2.

The positions and times defining a stride are estimated from the detected stance phases. The first and last frames (f_{first} and f_{last}) of each stance phase are recorded, and the stance position is calculated as the median of all positions in the phase. The left and right stance phases are grouped together and ordered by initial frame. An example is shown in Table I. Each group of three consecutive stance phases represents a stride, where the three median positions are $\mathbf{p}_{a,i}$, $\mathbf{p}_{b,i}$, and $\mathbf{p}_{a,i+1}$, respectively. However, gait parameters are only computed if the three phases have alternating sides (either L-R-L or R-L-R). This is intended to prevent the recording of incorrect gait parameters if the clustering algorithm misses a stance phase or adds an extra one.

The stride length is the distance from $\mathbf{p}_{a,i}$ to $\mathbf{p}_{a,i+1}$.

$$l_{stride,a,i} = \|\mathbf{p}_{a,i+1} - \mathbf{p}_{a,i}\| \qquad (10)$$

The step length and stride width depend on $\mathbf{p}_{b,i,proj}$. This is the projection of $\mathbf{p}_{b,i}$ onto the line defined by $\mathbf{p}_{a,i}$ and $\mathbf{p}_{a,i+1}$.

$$l_{step,a,i} = \|\mathbf{p}_{a,i+1} - \mathbf{p}_{b,i,proj}\| \qquad (11)$$

$$w_{stride,a,i} = \|\mathbf{p}_{b,i} - \mathbf{p}_{b,i,proj}\| \qquad (12)$$

The stride time is the time from the first contact of one foot to the next first contact of the same foot. Assuming a constant frame rate (frames per second; fps), this can be calculated by dividing the difference of frames by the frame rate.

$$t_{stride,a,i} = \frac{f_{first,a,i+1} - f_{first,a,i}}{fps} \qquad (13)$$

The stance time is the time from the first contact to the last contact of the same foot.

$$t_{stance,a,i} = \frac{f_{last,a,i} - f_{first,a,i}}{fps} \qquad (14)$$

The stance percentage is the stance time divided by the stride time.

$$p_{stance,a,i} = \frac{t_{stance,a,i}}{t_{stride,a,i}} \cdot 100\% \qquad (15)$$

Unfortunately, time stamps were not recorded during the data collection phase of this study, so we could not reliably assess the frame rate value or its consistency (and if time stamps were available, the stride/stance time could be calculated by simply subtracting the corresponding time stamps, ignoring the frame rate altogether). For this reason, we only report on stance percentage instead of stride and stance time.

If the frame rate is assumed to be constant, the stance percentage can be calculated without it because it is cancelled from the equation.

$$p_{stance,a,i} = \frac{f_{first,a,i+1} - f_{first,a,i}}{f_{last,a,i} - f_{first,a,i}} \cdot 100\% \qquad (16)$$

The stride length, step length, stride width, and stance percentage are calculated for each stride in the walking pass and for each pass in the trial.

The PKMAS system also outputs negative gait measurements in the case of atypical step length (when $\mathbf{p}_{a,i+1}$ falls behind $\mathbf{p}_{b,i}$) or stride width (when $\mathbf{p}_{b,i}$ crosses over the line from $\mathbf{p}_{a,i}$ to $\mathbf{p}_{a,i+1}$). The equations for these atypical cases are described in [40]. However, accounting for atypical gait was not in the scope of this project. Instead, we compared the absolute values of the Zeno gait parameters to our system.

IV. EXPERIMENTS AND RESULTS

A. Data Sets

Our data set consisted of 52 walking trials measured concurrently by a Zeno Walkway and Kinect v1 depth sensor. Four female participants with MS were instructed to wear their normal clothing and walk at a natural pace. For each trial, participants were either instructed to walk while completing a cognitive task (i.e., dual-tasking) or to just walk at a normal pace without a cognitive dual-task (see Table II).

Two additional walking trials were captured only by the Kinect in the same environment. A label image was created from each depth image in these trials by segmenting the human form using the same technique used to train the part predictor, as seen in Fig. 8. The approximate true positions of body parts were obtained for these two trials by computing the median position of

TABLE II
PARTICIPANT TRIAL BREAKDOWN, DEMOGRAPHIC AND CLINICAL DETAILS, INCLUDING EXPANDED DISABILITY STATUS SCALE (EDSS) SCORES

| Normal Walk	Dual Task	Age	Height (cm)	Weight (kg)	EDSS	Handed	Comorbidities	Medications
2	4	60	160	70	3.5	Left	Hypertension Dyslipidemia	Teriflunomide, Perindopril, Atenolol Nifedipine, Simvastatin, Hydralazine Carbamazepine
4	10	40	171	60	2.5	Right	Raynaud's Dyslipidemia	Fingolimod, Pantoprazole, Valacyclovir Carbamazepine, Venlafaxine Vitamin D, Co-cyprindiol
4	10	47	175	75	3.0	Right	-	-
5	13	48	168.5	90	6.0	-	Migraines	Glatiramer Acetate, Oxybutynin Amitriptyline, Citalopram Vitamin D, Ketorolac
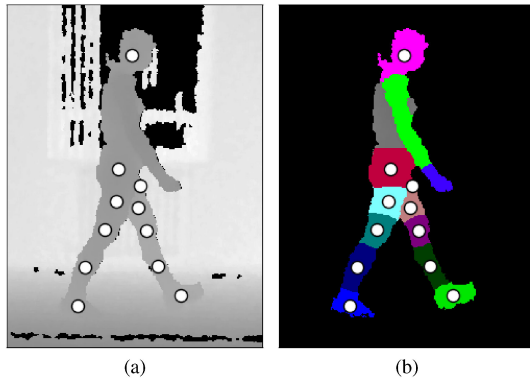
Fig. 8. Example frame from a labelled walking trial. The white dots indicate the median positions of body part segments in the label image. (a) Depth image. (b) Corresponding label image.

Fig. 9. Modified accuracy of selected foot positions compared to the radii used in Section II-D. A radius r on the x axis indicates that the radii $\{0, 1, \ldots, r\}$ were used. This accuracy was calculated as the ratio of frames where both selected foot positions were within the distance D of the corresponding modified truth positions.

each segment of pixels, then converting from image coordinates to real world coordinates.

Our method for selecting joint proposals (Section II) was applied to frames containing at least one proposal for each part type of interest (head, hip, thigh, knee, calf, and foot). A total of 19399 frames were processed from the 52 Zeno trials and two labelled trials.

The method and results were implemented in Python using scientific computing libraries [47]–[54].

B. Pose Estimation

The head and two foot positions were first selected on each frame without assigning left/right sides to the feet (Section II-D and Section II-E). The approximate true positions from the two labelled trials were used to evaluate the accuracy of this selection.

We defined accuracy as the ratio of frames where the selected position is within a distance D of the true position. Following the convention of [33], we set $D = 10$ cm. The selected head positions achieved an accuracy of 0.98.

A limitation of approximating true foot positions from the labelled trials is that the feet occlude each other as the person walks, pushing the centre of the pixels away from the true centre of the foot. A gold-standard motion tracking system such as Vicon [13] would be needed to establish a ground truth for the foot positions. Furthermore, because our method selects foot positions from the available proposals on a frame, the
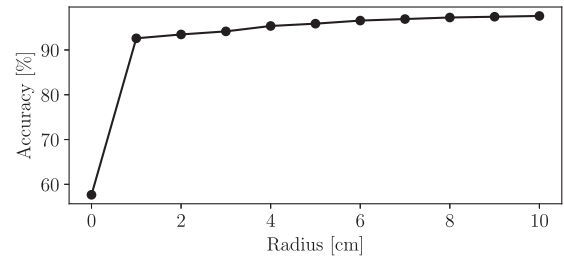
selected positions can only be as accurate as the most accurate proposals. There can be frames where none of the proposals are within the distance D of either approximate true position. For these reasons, we computed modified accuracies by comparing the selected foot positions to a modified truth. The modified left/right truth position was set as the proposal closest to the approximate true position from the label image.

In order to compare the selected foot positions to the modified truth, they were matched with the left and right modified truth positions by taking the pairing with the smaller total distance from matched to truth. Then, the left/right foot modified accuracy was computed as the ratio of frames where the left/right matched position is within the distance D of the corresponding modified truth position. The resulting modified accuracies were 0.98 and 0.98 compared to the left and right modified truth positions. We also found the ratio of frames where both of the matched foot positions were within the distance D of their corresponding modified truth positions. This resulted in a modified accuracy of 0.97.

The feet were selected using spheres of various radii (Section II-D). Fig. 9 shows the modified accuracy of the selected feet versus radii. Each radius r on the horizontal axis indicates the range of radii $\{0, 1, \ldots, r\}$cm. The modified accuracy improved significantly from a radius of zero to one. Only small improvements were observed afterwards.

C. Gait Analysis

There were 214 walking passes over the 52 walking trials measured by the Kinect and Zeno Walkway. Fifteen (28.8%) of

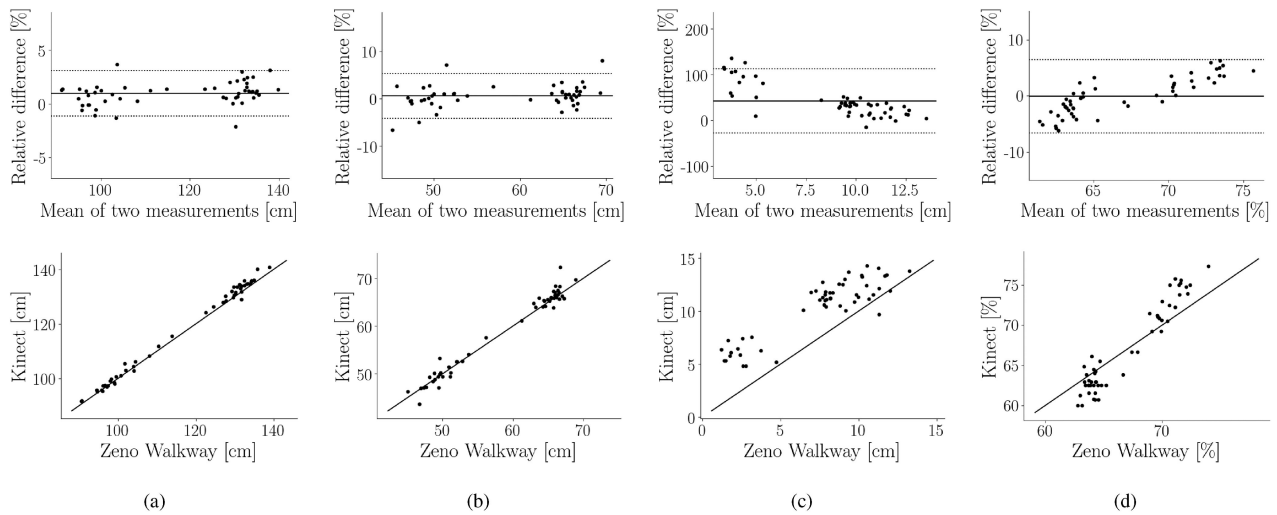|	Normal walk				Dual-task walk			
	Bland-Altman [%]		ICC		Bland-Altman [%]		ICC	
	Bias	1.96σ	$ICC_{2,1}$	$ICC_{3,1}$	Bias	1.96σ	$ICC_{2,1}$	$ICC_{3,1}$
Stride length [cm]	1.09	1.71	1.00	1.00	0.93	2.28	0.99	1.00
Step length [cm]	1.58	4.61	0.98	0.98	0.15	4.56	0.99	0.99
Stride width [cm]	39.14	71.56	0.66	0.87	43.67	70.19	0.61	0.89
Stance percentage [%]	−0.91	7.45	0.83	0.83	0.24	6.26	0.89	0.89

Fig. 10. Comparison of gait parameters calculated from the Kinect and Zeno Walkway. The upper plots are Bland-Altman plots. The horizontal lines show the bias (solid) and limits of agreement (dotted). The lower plots are direct comparisons of the values. The diagonal line shows the ideal agreement. (a) Stride length. (b) Step length. (c) Stride width. (d) Stance percentage.

the trials were normal walking, and the remainder were dual-task. Gait parameters were calculated for a total of 709 strides. The median number of strides was 11.5 per trial and 3.0 per pass. As stated in Section III-A, frames were grouped into walking passes with DBSCAN, which can mark data points as noise. Only one frame was marked as noise out of all the trials.

1) Stance Positions: As described in Section III-D, the detected stance positions of the feet were assigned to left and right sides. Therefore, these positions can be directly compared to the left/right modified truth positions from the labelled trials, rather than matching pairs of foot positions as done in the previous section. The left/right accuracies were both 1.0 when compared to the modified truth positions.

2) Bland-Altman: Bland-Altman analysis [55] is a common technique to quantify the agreement between two measurement devices. The difference between two 1D arrays of measurements X_A and X_B are computed, and the bias of device A compared to device B is the mean of these differences. In this case, X_A is the array of measurements from our system for one gait parameter and X_B is the array of corresponding walkway measurements.

Since the gait parameters have different magnitudes (e.g., stride length is longer than stride width), a direct comparison of Bland-Altman differences would be biased towards the smaller parameters. Furthermore, the parameters have a variety of dimensions (length, speed, and percentage of time),

making the comparison invalid. Because of this, we computed relative differences as suggested in [56]. The relative difference between two measurements x_A and x_B was calculated as $(x_A - x_B)/\text{mean}(x_A, x_B)$.

The limits of agreement are defined as the bias $\pm 1.96\sigma$, where σ is the standard deviation of the differences. Assuming that the differences are normally distributed, then 95% of the differences are expected to lie between the limits of agreement [56]. Thus, a low tolerance (1.96σ) defining the limits indicates a strong agreement.

Table III displays the bias and limits of agreement of gait parameters calculated by our method when compared to the ground truth walkway, separated by walking type (normal or dual-task). For normal walking, stance percentage had the lowest relative bias magnitude (0.91%) and stride width had the highest (39.14%). Stride length had the lowest tolerance ($1.96\sigma = 1.71\%$) and stride width the highest (71.56%). For dual-task walking, step length had the lowest bias magnitude (0.15%) and stride width the highest (43.67%). Stride length had the lowest tolerance (2.28%) and stride width the highest (70.19%). The results are visualized in Fig. 10, showing the bias and limits of agreement of all walking trials grouped together.

3) Intraclass Correlation: Interclass correlation coefficients, such as Pearson's coefficient, quantify the correlation between variables of different classes. By contrast, intraclass correlation
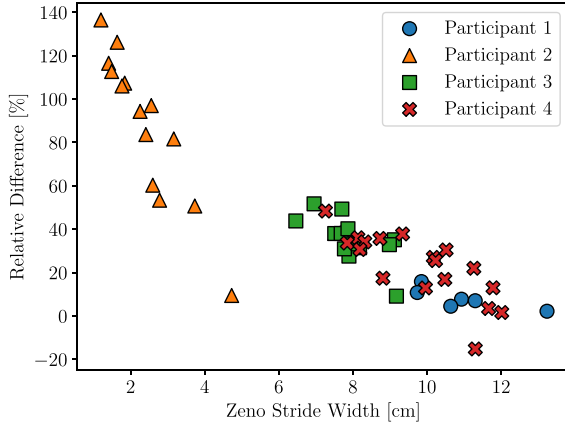
Fig. 11. Relative difference of stride width between the two systems plotted against the actual (Zeno) stride widths. The relative difference is highest when actual stride width is lowest.

coefficients (ICCs) quantify both the correlation and agreement between variables of the same class. An ICC value ranges from zero to one, where one is the highest reliability [57].

We calculated ICCs of the form $ICC_{2,1}$ and $ICC_{3,1}$. The former quantifies the absolute agreement between raters (the Kinect and Zeno Walkway), and the latter quantifies consistency across the walking trials. Both are calculated from the same matrix, where each row represents a walking trial and each column represents a rater. A detailed description of the required calculations can be found in [58].

The two forms of ICC are reported in Table III for the gait parameters. For normal walking, stride length had the highest agreement ($ICC_{2,1} = 1.00$) and stride width the lowest (0.66). Stride length also had the highest consistency across trials ($ICC_{3,1} = 1.00$) and stance percentage the lowest (0.83). For dual-task walking, stride length and step length had the highest agreement (0.99) and stride width the lowest (0.61). Stride length also had the highest consistency across trials (1.00) and stride width and stance percentage the lowest (0.89).

4) Analysis of Stride Width Error: Bland-Altman analysis and ICCs (Table III) both indicated that stride width had low agreement between the Kinect and Zeno Walkway. Further analysis was conducted to identify the source of the error. Fig. 11 shows the relative differences per trial plotted against the Zeno stride widths. The points are coloured by participant. The plot indicates that there is an inverse relationship between the actual (Zeno) stride width and the relative difference. It also shows that the highest differences all emanated from Participant 2, who had the lowest actual stride widths. This visualization suggests that our system overestimates the stride width when the actual value is small.

V. DISCUSSION

Our system is capable of calculating standard spatiotemporal gait parameters starting with multiple joint proposals, which are generated from side-view depth images of walking trials. The joint proposals are represented as a weighted graph, with weights dependent on the difference between expected and measured lengths between body parts. The shortest paths from head to foot find combinations of parts with lengths similar to the expected lengths. We employ a voting process to select the two shortest paths that best represent the actual two sides of the body, in turn providing the best head and feet. By examining the motion of the feet over time, the stance phases (when the foot is contact with the floor) are detected. The stance phases are assigned to left/right sides using the general direction of walking motion. Gait parameters are calculated from the positions and frames of these stance phases.

A potential drawback of tracking from a side view is that one leg occludes the other as it passes. However, our gait measures are intended to be robust to this disturbance, because they are calculated from frames when the feet are mostly or fully apart. Additionally, a side view supports multiple sensors placed along a longer walkway, allowing for analyses of long walking trials. The field of view provided by frontal-view sensors cannot be easily extended in this way, compounded by issues such as interference between sensors facing each other or sharing significant overlap in field of view.

The signals Φ and Ψ (from Sections III-C and III-D) are analogous to the x and z (depth) coordinates of the camera view when the subject is walking in a line perpendicular to the camera. These signals are used instead of the actual x and z coordinates to cover the cases where the line of motion is not directly perpendicular. This helps our method to be generalized to other non-frontal camera views, such as from an upper corner of a room.

We found that few radii were needed for the foot selection algorithm to achieve a high accuracy. The addition of a single radius beyond zero caused the majority of the improvement, from < 0.60 to > 0.92. Furthermore, high accuracies were achieved for the feet individually when compared to expected positions (0.98 for each foot individually, 0.97 for both feet at the same time) when selecting the best proposals available.

We compared our gait parameters to ground truth parameters from the Zeno Walkway, a pressure-sensitive walkway used in clinical practice. Compared to the work of [30], who separated trials by walking type, our agreement ($ICC_{2,1}$) on normal walking pace was higher for step length ($0.98 > 0.93$), and our consistency ($ICC_{3,1}$) was higher as well ($0.98 > 0.94$). For dual-task walking, agreement was again higher for step length ($0.99 > 0.94$), and consistency was much higher ($0.99 > 0.79$). In terms of limits of agreement, we had a similar tolerance (1.96σ) for step length with normal walking ($4.61\% \approx 5\%$)) and dual-task ($4.56\% \approx 5\%$). They did not report results for stance percentage, but did report on step and stance time, which we treat as proxies since stance percentage is the stance time divided by the stride time (Equation 15), and step time is the corresponding time for step length instead of stride length [30]. For normal pace, they achieved higher agreement (0.96 for step time and 0.93 for stance time vs. 0.83 for stance percentage) and consistency (0.90 and 0.92 vs. 0.83). For dual-task walking, they again achieved a higher agreement ($ICC_{2,1} = 0.98$ for step and stance time vs. 0.89 for our stance percentage) but our consistency was similar (0.88 and 0.89 vs. 0.89). Their tolerance was lower for normal pace (4% for step and stance time vs. 7.45% for stance percentage), but the gap was smaller for dual-task walking (6% and 5% vs. 6.26%).

Subjects in [31] completed trials at a comfortable walking speed and their maximum walking speed. We compare our results for normal pace to theirs with comfortable walking. Our agreement was comparable to theirs for stride length ($1.00 \approx 0.999$), step length ($0.98 \approx 0.994$), and step width ($0.66 \approx 0.646$).

Participants in [16] completed walking trials at 100, 75, and 50 percent of comfortable speed. However, they only reported results with all walking types grouped together, so we compare our normal walking and dual-task results to their grouped results. For normal walking, our agreement was higher for stride length ($1.00 > 0.83$) and step length ($0.98 > 0.71$). The same was found for dual-task ($0.99 > 0.83$ and $0.99 > 0.71$). Agreement for our stance percentage was higher than their stance time ($0.83 > 0.74$) but lower than their stride time ($0.83 < 0.89$). However, our agreement for dual-task walking was the same as their stride time agreement (0.89).

VI. LIMITATIONS AND FUTURE WORK

The approximate true positions of the head and feet were derived from two manually labelled walking trials (Section IV-A). As we mentioned in Section IV-B, these labelled trials are not ideal for establishing ground truth positions because some parts occlude each other as the person walks, moving the centres of the pixel segments away from the true centres of the parts. Because of this, we reported a modified accuracy for our selected foot positions, comparing them to the best proposals available rather than directly to the positions from the labelled images. Moving forward, the selected positions should be compared to a ground truth from a gold-standard motion tracking system such as Vicon [13]. Furthermore, our method of selecting from proposals means that we are limited by the accuracy of the proposals themselves. This work could be extended by a post-processing step to improve the accuracy of the selected positions, such as temporal filtering.

As explained in Section III-E, we did not report on stride or stance time because time stamps were not captured during the data collection phase of this study. However, we did report on stance percentage because it can be calculated without knowing the frame rate, under the assumption that the frame rate is constant (Equation 16). Future work should ensure that time stamps are captured for all depth frames, enabling the direct calculation of stride time, stance time, and stance percentage.

Fig. 11 shows that the relative difference of stride width was highest when the actual stride width was lowest. In these trials, our system would overestimate the stride width. We hypothesize that this is caused by different definitions for the positions of the feet. The Zeno gait parameters are calculated using the positions of the heels of the feet, while our system locates the general positions of the feet — typically the centres. The difference is negligible for a large spatial gait parameter like stride length, but may be significant for stride width, especially if the feet are turned outwards. In future work, we propose locating the rear of each foot using our knowledge of the direction of motion, such that our measurement will be consistent with the Zeno heel positions. This could improve the measurement of stride width

and enable the calculation of additional gait parameters such as foot angle (i.e., by similarly locating the front of the feet).

VII. CONCLUSION

We presented a new system for measuring clinical gait parameters with a side-view depth sensor. The use of a non-frontal depth sensor, in comparison to a frontal sensor, adds convenience to clinical trials because only one camera is required for a person walking in both directions along a walkway. Furthermore, a non-frontal sensor extends the range in which a full person is in the field of view. While researchers have previously investigated gait measurement with non-frontal depth sensors, our contribution is the direct calculation of standard gait parameters from individual foot positions from the side perspective.

We first selected human joint positions from multiple proposals generated on depth images. The selected foot positions were further analyzed to detect stance phases, which were used to calculate four gait parameters (stride and step length, stride width, and stance percentage). The results demonstrated that accurate positions were selected from the available proposals. Using a pressure-sensitive walkway as ground truth, we found that the large spatial gait parameters (stride and step length) were the most reliable.

Possible extensions to our system include the use of other body parts to measure gait parameters that are inaccessible to a pressure-sensitive walkway. In order to track the upper body, our foot selection process could be applied to select the two best hand proposals, by finding the shortest paths from the head to the hands. The method could also be adapted into an online algorithm that continuously updates estimates of the body lengths and walking direction.

In conclusion, we envision a vision-based system capable of measuring gait parameters from the full body, that can collect data conveniently and unobtrusively for clinical purposes. To encourage further research, we have made our code open source at github.com/ajhynes7/side-view-depth-gait. The authors can be contacted for the full data set.

REFERENCES

[1] M. W. Whittle, "Clinical gait analysis: A review," *Human Movement Sci.*, vol. 15, no. 3, pp. 369–387, 1996.

[2] J. R. Gage, "Gait analysis. An essential tool in the treatment of cerebral palsy," *Clin. Orthopaedics Related Res.*, vol. 3, no. 288, pp. 126–134, 1993.

[3] M. Benedetti *et al.*, "Gait abnormalities in minimally impaired multiple sclerosis patients," *Mult. Scler. J.*, vol. 5, no. 5, pp. 363–368, 1999.

[4] C. L. Martin *et al.*, "Gait and balance impairment in early multiple sclerosis in the absence of clinical disability," *Mult. Scler. J.*, vol. 12, no. 5, pp. 620–628, 2006.

[5] U. Givon *et al.*, "Gait analysis in multiple sclerosis: Characterization of temporal–spatial parameters using GAITRite functional ambulation system," *Gait Posture*, vol. 29, no. 1, pp. 138–142, 2009.

[6] M. C. Kirkland *et al.*, "Comparing three dual-task methods and the relationship to physical and cognitive impairment in people with multiple sclerosis and controls," *Mult. Scler. Int.*, vol. 2015, 2015, Art. no. 650645.

[7] D. Podsiadlo and S. Richardson, "The timed "Up & Go": A test of basic functional mobility for frail elderly persons," *J. Amer. Geriatrics Soc.*, vol. 39, no. 2, pp. 142–148, 1991.

[8] R. Phan-Ba *et al.*, "Comparison of the timed 25-foot and the 100-meter walk as performance measures in multiple sclerosis," *Neurorehabil. Neural Repair*, vol. 25, no. 7, pp. 672–679, 2011.

[9] R. C. Lynall *et al.*, "Reliability and validity of the protokinetics movement analysis software in measuring center of pressure during walking," *Gait Posture*, vol. 52, pp. 308–311, 2017.

[10] T. Cloete and C. Scheffer, "Benchmarking of a full-body inertial motion capture system for clinical gait analysis," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 4579–4582.

[11] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 4, pp. 413–423, Jul. 2008.

[12] W. Tao *et al.*, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, 2012.

[13] A. Pfister *et al.*, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *J. Med. Eng. Technol.*, vol. 38, no. 5, pp. 274–280, 2014.

[14] S. Czarnuch and M. Ploughman, "Automated gait analysis in people with Multiple Sclerosis using two unreferenced depth imaging sensors: Preliminary steps," in *Proc. IEEE Newfoundland Elect. Comput. Eng. Conf., Newfoundland and Labrador Section*, 2015, pp. 1–5.

[15] M. Zago *et al.*, "3D tracking of human motion using visual skeletonization and stereoscopic vision," *Front. Bioeng. Biotechnol.*, vol. 8, 2020, Art. no. 181.

[16] M. Iwai *et al.*, "The validity of spatiotemporal gait analysis using dual laser range sensors: A cross-sectional study," *Arch. Physiother.*, vol. 9, no. 1, 2019, Art. no. 3. [Online]. Available: https://doi.org/10.1186/s40945-019-0055-6

[17] A. Castelli *et al.*, "A 2D markerless gait analysis methodology: Validation on healthy subjects," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–11, 2015.

[18] X. Gu, F. Deligianni, B. Lo, W. Chen, and G. Z. Yang, "Markerless gait analysis based on a single RGB camera," in *Proc. IEEE 15th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2018, pp. 42–45.

[19] Z. Wang, "A single RGB camera based gait analysis with a mobile tele-robot for healthcare," *ArXiv*, vol. abs/2002.04700, pp. 1–4, 2020.

[20] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Apr. 2012.

[21] A. Hynes and S. Czarnuch, "Human part segmentation in depth images with annotated part positions," *Sensors*, vol. 18, no. 6, 2018, Art. no. 1900.

[22] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with Kinect," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 1964–1967.

[23] E. E. Stone and M. Skubic, "Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2925–2932, Oct. 2013.

[24] G. Baldewijns, G. Verheyden, B. Vanrumste, and T. Croonenborghs, "Validation of the Kinect for gait analysis using the GAITRite walkway," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 5920–5923.

[25] F. Ahmed *et al.*, "Kinect-based gait recognition using sequences of the most relevant joint relative angles," *J. WSCG*, vol. 23, no. 2, pp. 147–156, 2015.

[26] A. A. Chaaraoui, J. R. Padilla-López, and F. Flórez-Revuelta, "Abnormal gait detection with RGB-D devices using joint motion history features," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2015, vol. 7, pp. 1–6.

[27] X. Xu *et al.*, "Accuracy of the Microsoft Kinect$_T$M for measuring gait parameters during treadmill walking," *Gait Posture*, vol. 42, no. 2, pp. 145–151, 2015.

[28] S. Motiian *et al.*, "Automated extraction and validation of children's gait parameters with the Kinect," *Biomed. Eng. Online*, vol. 14, no. 1, 2015, Art. no. 112. [Online]. Available: https://doi.org/10.1186/s12938-015-0102-9

[29] E. Cippitelli *et al.*, "Kinect as a tool for gait analysis: Validation of a real-time joint extraction algorithm working in side view," *Sensors*, vol. 15, no. 1, pp. 1417–1434, 2015.

[30] E. Dolatabadi *et al.*, "Concurrent validity of the Microsoft Kinect for Windows v2 for measuring spatiotemporal gait parameters," *Med. Eng. Phys.*, vol. 38, no. 9, pp. 952–958, 2016.

[31] D. J. Geerse *et al.*, "Kinematic validation of a multi-Kinect v2 instrumented 10-meter walkway for quantitative gait assessments," *PLoS One*, vol. 10, no. 10, 2015, Paper e0139913.

[32] J. Behrens *et al.*, "Using perceptive computing in multiple sclerosis-the Short Maximum Speed Walk test," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, 2014, Art. no. 89. [Online]. Available: https://doi.org/10.1186/1743-0003-11-89

[33] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1297–1304.

[34] C. Amon and F. Fuhrmann, "Evaluation of the spatial resolution accuracy of the face tracking system for Kinect for Windows V1 and V2," in *Proc. 6th Congr. Alps-Adria Acoust. Assosiation*, 2014.

[35] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.

[36] S. Czarnuch and A. Mihailidis, "Development and evaluation of a hand tracker using depth images captured from an overhead perspective," *Disability Rehabil.: Assistive Technol.*, vol. 11, no. 2, pp. 150–157, 2016.

[37] D. Zhang and M. Shah, "Human pose estimation in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2012–2020.

[38] L. Pishchulin *et al.*, "Deepcut: Joint subset partition and labeling for multi person pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4929–4937.

[39] "The Zeno Walkway gait analysis system," 2020. [Online]. Available: https://www.protokinetics.com/zeno-walkway/

[40] F. Huxham *et al.*, "Defining spatial parameters for non-linear walking," *Gait Posture*, vol. 23, no. 2, pp. 159–163, 2006.

[41] A. Hynes and S. Czarnuch, "Combinatorial optimization for human body tracking," in *Proc. Int. Symp. Vis. Comput.*, 2016, pp. 524–533.

[42] E. W. Weisstein, "Complete bipartite graph," From MathWorld–A Wolfram Web Resource. [Online]. Available: https://mathworld.wolfram.com/CompleteBipartiteGraph.html

[43] T. H. Cormen *et al.*, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2009.

[44] M. Ester *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Kdd*, vol. 96, no. 34, pp. 226–231, 1996.

[45] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[46] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial–temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, 2007.

[47] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[48] P. Virtanen *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020.

[49] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, 2007.

[50] W. McKinney, "pandas: A foundational Python library for data analysis and statistics," *Python High Perform. Scientific Comput.*, vol. 14, pp. 1–9, 2011.

[51] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proc. 9th Python Sci. Conf.* Scipy, vol. 57, pp. 92–96, 2010.

[52] S. Van der Walt *et al.*, "scikit-image: Image processing in Python," *PeerJ*, vol. 2, 2014, Paper e453.

[53] S. Hoyer and J. Hamman, "xarray: ND labeled arrays and datasets in Python," *J. Open Res. Softw.*, vol. 5, no. 1, 2017, Art. no. 10. doi: 10.5334/jors.148.

[54] J. Howse, *OpenCV Computer Vision With Python*. Birmingham, U.K.: Packt Publishing Ltd., 2013.

[55] J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[56] D. Giavarina, "Understanding Bland Altman analysis," *Biochemia Medica: Biochemia Medica*, vol. 25, no. 2, pp. 141–151, 2015.

[57] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, 2016.

[58] D. Liljequist *et al.*, "Intraclass correlation—A discussion and demonstration of basic features," *PLoS One*, vol. 14, no. 7, 2019, Art. no. e0219854.