

University of Groningen

2D Gait Skeleton Data Normalization for Quantitative Assessment of Movement Disorders from Freehand Single Camera Video Recordings

Tang, Wei; van Ooijen, Peter M A; Sival, Deborah A; Maurits, Natasha M

Published in:
Sensors

DOI:
[10.3390/s22114245](https://doi.org/10.3390/s22114245)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Tang, W., van Ooijen, P. M. A., Sival, D. A., & Maurits, N. M. (2022). 2D Gait Skeleton Data Normalization for Quantitative Assessment of Movement Disorders from Freehand Single Camera Video Recordings. *Sensors*, 22(11), [4245]. <https://doi.org/10.3390/s22114245>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Article

2D Gait Skeleton Data Normalization for Quantitative Assessment of Movement Disorders from Freehand Single Camera Video Recordings

Wei Tang ^{1,*}, Peter M. A. van Ooijen ², Deborah A. Sival ³ and Natasha M. Maurits ¹

¹ Department of Neurology, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 RB Groningen, The Netherlands; n.m.maurits@umcg.nl

² Data Science Center in Health, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 RB Groningen, The Netherlands; p.m.a.van.ooijen@umcg.nl

³ Department of Pediatric Neurology, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 RB Groningen, The Netherlands; d.a.sival@umcg.nl

* Correspondence: w.tang@umcg.nl

Abstract: Overlapping phenotypic features between Early Onset Ataxia (EOA) and Developmental Coordination Disorder (DCD) can complicate the clinical distinction of these disorders. Clinical rating scales are a common way to quantify movement disorders but in children these scales also rely on the observer's assessment and interpretation. Despite the introduction of inertial measurement units for objective and more precise evaluation, special hardware is still required, restricting their widespread application. Gait video recordings of movement disorder patients are frequently captured in routine clinical settings, but there is presently no suitable quantitative analysis method for these recordings. Owing to advancements in computer vision technology, deep learning pose estimation techniques may soon be ready for convenient and low-cost clinical usage. This study presents a framework based on 2D video recording in the coronal plane and pose estimation for the quantitative assessment of gait in movement disorders. To allow the calculation of distance-based features, seven different methods to normalize 2D skeleton keypoint data derived from pose estimation using deep neural networks applied to freehand video recording of gait were evaluated. In our experiments, 15 children (five EOA, five DCD and five healthy controls) were asked to walk naturally while being videotaped by a single camera in 1280 × 720 resolution at 25 frames per second. The high likelihood of the prediction of keypoint locations (mean = 0.889, standard deviation = 0.02) demonstrates the potential for distance-based features derived from routine video recordings to assist in the clinical evaluation of movement in EOA and DCD. By comparison of mean absolute angle error and mean variance of distance, the normalization methods using the Euclidean (2D) distance of left shoulder and right hip, or the average distance from left shoulder to right hip and from right shoulder to left hip were found to better perform for deriving distance-based features and further quantitative assessment of movement disorders.

Keywords: early onset ataxia (EOA); developmental coordination disorder (DCD); AlphaPose; gait; pose estimation



Citation: Tang, W.; van Ooijen, P.M.A.; Sival, D.A.; Maurits, N.M. 2D Gait Skeleton Data Normalization for Quantitative Assessment of Movement Disorders from Freehand Single Camera Video Recordings. *Sensors* **2022**, *22*, 4245. <https://doi.org/10.3390/s22114245>

Academic Editor: Christoph Hintermüller

Received: 29 April 2022

Accepted: 31 May 2022

Published: 2 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Children with Early-Onset Ataxia (EOA) and Developmental Coordination Disorder (DCD) both show features of motor incoordination, hampering clinical distinction [1]. Gait analysis plays a pivotal role in motor coordination assessment, contributing to early diagnostics and rehabilitation [2]. In the clinical context, ataxic gait can be assessed by validated, semi-quantitative rating scales such as the Scale for the Assessment and Rating of Ataxia (SARA) [3]. Despite the high reliability of the clinical scales, evidence shows they might be less effective in measuring diverse motor control strategies and/or mixed

disorders [4]. Moreover, their application may depend on the evaluation and interpretation of experts, especially in children with mixed movement disorder features [5,6]. Altogether, usage of clinical rating scales such as the SARA can sometimes be regarded as subjective, sensitive to confounding factors and time-consuming [4].

Inertial wearable sensors can provide an objective and feasible alternative for gait assessment [7]. The use of inertial measurement units (IMUs) with accelerometers and gyroscopes linked to the body with elastic straps has been investigated as an aid in the differential diagnosis of early onset ataxia and developmental coordination disorder [8]. Automatic classification based on quantitative gait features from IMUs has been shown to outperform phenotypic diagnosis, implying that movement quantification and subsequent automatic classification might be a useful technique for consistent and repeatable diagnostic evaluation [9]. However, this approach requires special preparation before use, and one of the problems to overcome is how to measure distance-based features such as step width, which is crucial for diagnosing ataxia: ataxia patients walk with reduced step length and increased walking base width [10].

Human pose estimation is one of the fundamental tasks in computer vision. This task can often be subdivided into single-person and multi-person pose estimation, as well as 2D and 3D pose estimation, depending on whether the video or image dataset contains multi-person or 3D depth information. Owing to the fast development of deep learning, many convolutional neural network (CNN) based frameworks have been designed and published, such as Maskrcnn [11] DeeperCut [12], DeepLabCut [13], CPN [14], OpenPose [15], which use convolutional pose machines to first locate human joints in an image and then a part affinity field to complete the human body assembly, and AlphaPose [16], which contains a symmetric spatial transformer network, parametric pose no-maximum-suppression, and a pose-guided proposal generator. Currently, the applications of pose estimation include motion capture for actors [17], assessment of athletes [18], and fall detection [19]. Compared with other frameworks, AlphaPose can achieve higher comprehensive accuracy on the MSCOCO dataset [20] with a map value of 72.3 [16], leading to its application in video behavior detection [21], action recognition [22], and gait analysis [23–26].

In the daily clinical setting, neurologists routinely record patients' movements for further evaluation. The ubiquity of video cameras in outpatient clinics and recent advances in video-based pose estimation have motivated us to design a 2D skeleton-based method for quantitative analysis of movement disorders using video images taken in the coronal plane. In a clinical setting, sagittal plane video recordings are typically not available, due to space limitations. Through pose estimation, the joint coordinates of the patient in single camera video recordings can be obtained in a convenient and low-cost way. The use of 2D video-based motion capture for gait analysis dates back to the last century [27]. It has so far covered a wide range of topics, from disease diagnosis to rehabilitation; several of those contributions relied on AlphaPose [23–26,28]. Mehdizadeh et al. [28] measured gait parameters in elderly adults and discovered that temporal, but not spatial or variability gait measures, derived from AlphaPose correlated with those calculated with a motion capture system in the frontal plane. Sabo et al. [25] made predictions of parkinsonian gait in older adults with dementia and for estimating parkinsonian severity using natural gait videos of older adults [26]. Peng et al. studied differences in gait parameters between the healthy population and patients with lower extremity dyskinesia [24]. Lv et al. found significant differences in the entropy of heel and ankle joint motion signals between healthy people and arthritic patients, which could be used to identify patients with knee arthritis [23]. To allow the calculation of changes in distance-based features, we need to normalize these 2D coordinates first. However, to the best of our knowledge, no prior study has discussed how to perform normalization to derive distance-based features.

Here, we propose and compare seven different normalization methods that enable the calculation of distance-based features derived from pose estimation results. The normalization method needs to address two issues: (1) the skeleton size varies depending on the distance from the camera in every frame, (2) the skeleton position in each frame

might change due to camera instabilities. These problems induce nonphysiological frame-to-frame variability, hindering calculation and comparison of distance-based features. We here initially apply the AlphaPose deep learning model and PoseFlow framework to obtain skeleton keypoint data from ataxia, DCD and the controls. We anticipate that the best normalization method will maintain the constancy of certain distances in the coronal plane, such as the distance between the two shoulders when walking, while maintaining the variability of other distances, such as the distance between the two wrists.

2. Materials and Methods

2.1. Pipeline

The entire pipeline for quantitative assessment of gait from freehand 2D camera recordings is summarized in Figure 1. A monocular camera is placed in front of the participant for recording. After the video data is obtained, a deep neural network model pretrained on the MSCOCO dataset [20] based on AlphaPose [16] is used to extract skeleton keypoints. PoseFlow [29] is applied to match the skeleton to the same participant in a recording. The keypoint data are then normalized using one of the proposed normalization methods, allowing distance-based features derived from the scaled 2D trajectories to be compared across different movement disorders. Finally, these extracted features can be used to perform quantitative analysis or classify participants. The details are explained below.

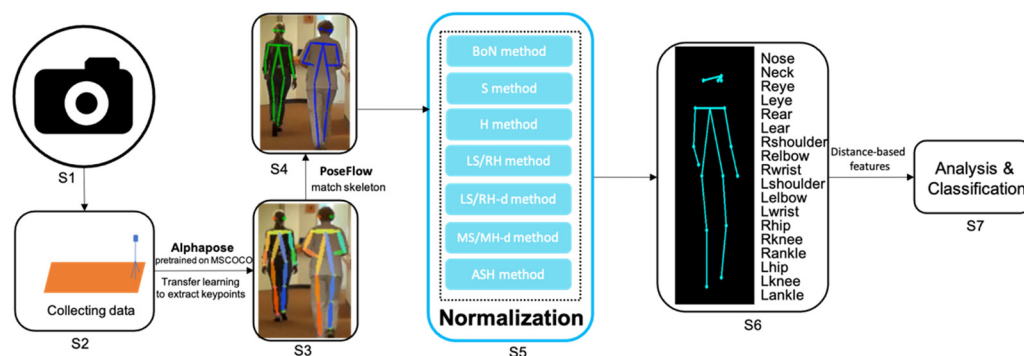


Figure 1. Overview of the proposed pipeline for the quantitative assessment of gait from freehand 2D camera recordings. From left to right: S1: Configure the camera and its settings. S2: Collect data with the camera. S3: Extract keypoints using AlphaPose. S4: Match the skeleton using PoseFlow. S5: Normalize the skeleton data with one of methods given in Section 2.4. (BoN: box normalization; S: shoulder normalization; H: hip normalization; LR/RH: left-shoulder right-hip normalization; LS/RH-d: left-shoulder right-hip distance normalization; MS/MH-d: mid-shoulder mid-hip distance normalization; ASH: average shoulder hip normalization). S6: Obtain the normalized keypoint skeleton sequences. S7: Analyze the (distance-based) features derived from skeleton data for classification.

2.2. Data Preparation

The data for this study were collected at the University Medical Center Groningen (Groningen, The Netherlands), in accordance with local research and integrity codes. Fifteen children, including five with EOA (mean age 12.6 years, SD 1.85 years), five with DCD (mean age 9.6 years, SD 3.21 years), and five healthy controls (mean age 9.4 years, SD 1.81 years) participated in the experiment. All included children older than 12 years of age and all parents of the children included provided informed consent before study enrollment. Children younger than 12 years of age also provided informed assent. EOA patients had radiologic (MRI), metabolic, and laboratory and genetic testing at the department of (pediatric) neurology at the University Medical Center Groningen as part of their diagnostic evaluation. DCD patients were also evaluated at the pediatric neurology outpatient clinic. When applicable, these evaluations included MRI, electromyography, muscle ultrasound, and/or laboratory and sometimes also genetic testing to exclude other underlying neurologic disorders.

Children were asked to walk freely and repeatedly in a straight line while an experimenter stood at the end of the corridor holding a single 2D video camera (1280 × 720, 25 fps) facing the subjects for recording. They were asked to walk at their own pace in a corridor of approximately 15 m, make a 180 degree turn and return to the starting position, in accordance with the SARA regular gait task guidelines. EOA was clinically confirmed in accordance with its definition [30]. For further processing, each video recording was divided into segments belonging to the following categories: (1) walking towards the camera, (2) walking away from the camera, (3) standing still, (4) turning around. We only analyzed the video segments of walking towards and away from the camera. In addition, frames in which body parts were occluded or lost were excluded from further analysis, because the skeletons in these frames could not be detected correctly using the current 2D pose estimation algorithm. In total, 60 segments of video data (15 participants, two types of gait, two segments each) were used for skeleton extraction.

2.3. Pose Estimation

After cleaning the video data, we used AlphaPose, a deep neural network model, to estimate the 2D locations of skeleton keypoints. Here, we utilized the Yolo-v3 detector [31], pretrained on the MSCOCO dataset, to detect people in the video frame. Subsequently, the fast resnet50 [32] model was chosen to obtain the final locations of each keypoint. For each frame, the model returned 2D coordinates (in pixels) as well as a prediction confidence probability for each of the 17 keypoints of each person detected, including the nose, and bilateral eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. Finally, we used a tracker, PoseFlow, to match the skeleton to the same participant in each recording.

2.4. Normalization

We proposed and analyzed different normalization methods. Here, the original keypoint data of each video segment were presented as a sequence of N continuous skeletons ($S_1, S_2, S_3 \dots S_N$), where skeleton S_i of the i -th frame ($i = 1 \dots N$) is composed of 17 pairs of coordinates (in pixels) representing all 17 keypoints.

It should be noted that movement inside the coronal plane (such as up-and-down or side-to-side movement of a limb) has no effect on distance computation, but movement outside the plane (such as walking towards or away from the camera, or body rotation) does. For this reason, we consider both the Euclidean distance between two joints and the distance between two points in the horizontal and vertical directions, respectively, for the different normalization methods. We consider two-step solutions, that include position shifting and size scaling.

Step 1: Position shifting. This method achieves the result that the mid-shoulder point becomes the origin (0, 0) in each frame, by shifting all coordinates according to:

$$X_{S_i}' = X_{S_i} - \frac{x_{S_i}^{Ls} + x_{S_i}^{Rs}}{2}, \quad Y_{S_i}' = Y_{S_i} - \frac{y_{S_i}^{Ls} + y_{S_i}^{Rs}}{2} \quad (1)$$

Here (X_{S_i}', Y_{S_i}') are the position-shifted coordinates in the skeleton, S_i , (X_{S_i}, Y_{S_i}) are the original coordinates and $(x_{S_i}^{Ls}, y_{S_i}^{Ls})$ and $(x_{S_i}^{Rs}, y_{S_i}^{Rs})$ are the coordinates of the left shoulder and right shoulder in the skeleton S_i , respectively.

Step 2: Size scaling. Seven methods were considered that each used different distances to scale every video frame, according to:

$$X_{S_i}'' = w_0 * \frac{X_{S_i}'}{w_{S_i}}, \quad Y_{S_i}'' = h_0 * \frac{Y_{S_i}'}{h_{S_i}} \quad (2)$$

Here (X_{S_i}'', Y_{S_i}'') and (X_{S_i}', Y_{S_i}') are the scaled and position-shifted coordinates, respectively, and w_{S_i} and h_{S_i} are width and height defined by the specific normalization method. Four of the assessed normalization methods use the width and height of the bounding boxes of: (1) all keypoints (Box scale or BoN method), (2) the two-shoulder key-

points (S method), (3) the left-shoulder and right-hip (LS/RH method) and (4) the two-hip keypoints (H method). The other methods use the Euclidean (2D) distance between (5) the left shoulder and right hip for both w and h (LS/RH-d method) and (6) the mid shoulder and mid hip for both w and h (MS/MH-d method). The last method uses the average distance from the left shoulder to right hip and from the right shoulder to left hip for both w and h (ASH method).

Finally, (w_0, h_0) can be obtained by averaging the width w_{S_i} and height h_{S_i} of each video segment, according to:

$$w_0 = \frac{1}{N} \sum_{i=1}^N w_{S_i}, h_0 = \frac{1}{N} \sum_{i=1}^N h_{S_i} \quad (3)$$

To visualize and reconstruct the video after normalization, we finally shift the whole skeleton to the middle of the image:

$$X_{S_i}''' = X_{S_i}'' + x_0, Y_{S_i}''' = Y_{S_i}'' + y_0 \quad (4)$$

Here, x_0, y_0 are constant values; we used $x_0 = 640$ and $y_0 = 200$ as we have a resolution of 1280×720 .

(1), (2) and (4) together can be written as follows:

$$X_{S_i}''' = \frac{w_0}{w_{S_i}} \left(X_{S_i} - \frac{x_{S_i}^{Ls} + x_{S_i}^{Rs}}{2} \right) + x_0 \quad (5)$$

$$Y_{S_i}''' = \frac{h_0}{h_{S_i}} \left(Y_{S_i} - \frac{y_{S_i}^{Ls} + y_{S_i}^{Rs}}{2} \right) + y_0 \quad (6)$$

To illustrate the effect of the different normalization steps on skeleton size and position, we show several frames of an EOA patient walking towards the camera at each step of the normalization process in Figure 2.



Figure 2. Example of several frames of an EOA patient walking towards the camera illustrating the effect of the different normalization steps when using the ASH method.

2.5. Evaluation

Performance was assessed by calculating the mean value of the absolute error of the angle (mean absolute angle error) and the mean variance of the distance between certain keypoints. It is important to note that we cannot obtain any true physical distance

measurements because we lost depth dimension information from the start by using a 2D camera. Therefore, our evaluation metrics are inspired by our clinical task, with the expectation that in the ideal case (1) there should be no changes in angles between vectors in the coronal plane before and after normalization, (2) normalization should maintain the variability of the distance between the two wrists and the two ankles, and (3) the constancy of the distances between the two shoulders and between the two hips is maintained (assuming that out-of-coronal-plane rotation is limited during gait towards and away from the camera). We selected four relative angles (angles between vectors): the angles between the left- and right-wrist–elbow vector and the elbow–shoulder vector, and the angles between the hip–knee vector and the knee–ankle vector. We also selected eight absolute angles (angles between vectors and the horizontal line) related to the left- and right-wrist–elbow, elbow–shoulder, hip–knee, and knee–ankle vectors. For comparison of the mean variance of distance, we selected the distance between the two shoulder, wrist, hip, and ankle keypoints, expecting that after normalization the variance of the distance between the two shoulder and between the two hip keypoints should be close to 0 (due to physical limitations), but the variance of the distance between the two wrist and two ankle keypoints should be as large as possible (expecting most variability in movement from these keypoints).

3. Results

3.1. Likelihood from Pose Estimation

Table 1 provides the mean and standard deviation of the likelihood (confidence probability) of the 17 keypoints for all participants and segments before and after cleaning. The average likelihood of skeleton keypoint locations obtained from pose estimation using raw video data was 0.817 (SD = 0.09). A paired-samples *t*-test was conducted to compare the likelihood of prediction of keypoint locations before data cleaning and after data cleaning. After noisy frames and the standing and turning around segments were removed, there was a significant increase of the probability ($M = 0.889$, $SD = 0.02$, $t(28) = 3.36$, $p < 0.05$).

Table 1. Likelihood of prediction of keypoint locations.

	Mean	STD
before cleaning	0.817	0.09
after cleaning	0.889	0.02

3.2. Mean Absolute Angle Error

The results for the mean absolute angle error are provided in Figure 3. The H method, especially for the eight absolute angles, has the largest error ($M = 14.60$), followed by the S method ($M = 10.39$). These two normalization methods had much higher errors compared with the other methods, hence they were omitted from further consideration. The other normalization methods had similar good performances.

3.3. Mean Variance of Distance

Figure 4 presents the mean variance of the distance between shoulder, wrist, hip, and ankle keypoints for the original data and the BoN, LS/RH, LS/RH-d, MS/MH-d, and ASH methods. As expected, the mean distance variances were high (shoulder: 16.33, wrist: 29.35, hip: 10.29, ankle: 13.74) in the original data. The results of the BoN (shoulder: 5.67, wrist: 8.18, hip: 5.96, ankle: 9.24) and LS/RH (shoulder: 5.26, wrist: 14.08, hip: 5.83, ankle: 9.69) methods were similar, with the exception that LS/RH normalization produces somewhat more fluctuation in the wrist component. The final three normalization methods (LS/RH-d, MS/MH-d, and ASH) provided fairly similar results with minor differences. The results of MS/MH-d (shoulder: 3.08, wrist: 11.52, hip: 5.09, ankle: 8.81) were a little higher than those of LS/RH-d (shoulder: 2.90, wrist: 11.49, hip: 5.06, ankle: 8.79), but these differences were not significant (shoulder: $U = 54$, $p = 0.16$; wrist: $U = 62$, $p = 0.29$; hip: $U = 57$, $p = 0.20$;

ankle: $U = 68, p = 0.41$; Mann–Whitney U-tests). Although the shoulder and hip variances of the ASH method obtained the lowest variances (shoulder: 2.78, wrist: 11.32, hip: 5.05, ankle: 8.77) among all normalization methods, there was no significant difference between the MS/MH-d and ASH methods either (shoulder: $U = 69, p = 0.44$; wrist: $U = 68, p = 0.42$; hip: $U = 65, p = 0.35$; ankle: $U = 67, p = 0.40$; Mann–Whitney U-tests). However, this method balanced using the distance between the left shoulder and right hip and the right shoulder and left hip to achieve a more stable performance.

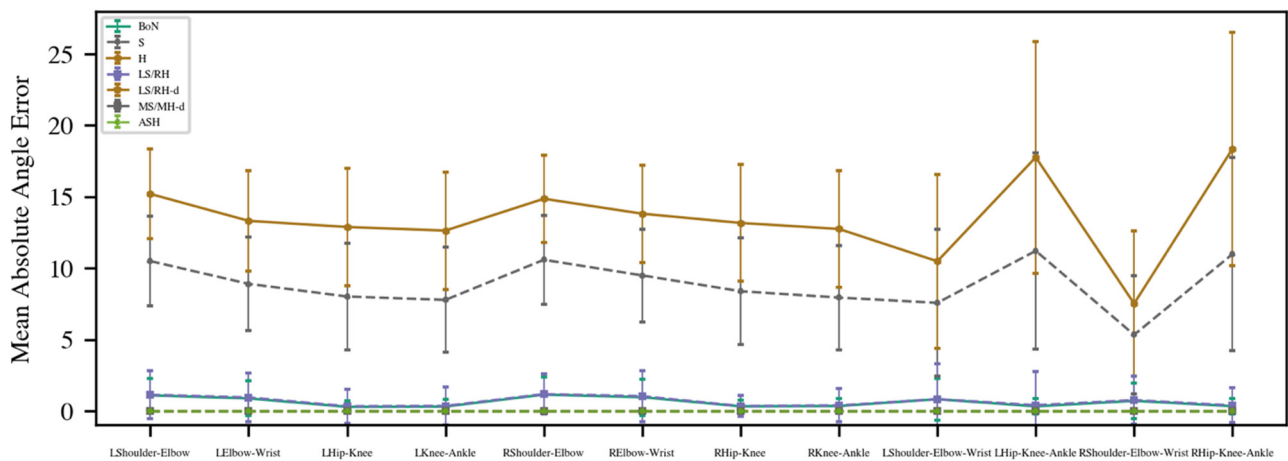


Figure 3. Mean absolute angle error results of proposed methods. (BoN: box normalization; S: shoulder normalization; H: hip normalization; LR/RH: left-shoulder right-hip normalization; LS/RH-d: left-shoulder right-hip distance normalization; MS/MH-d: mid-shoulder mid-hip distance normalization; ASH: average shoulder hip normalization).

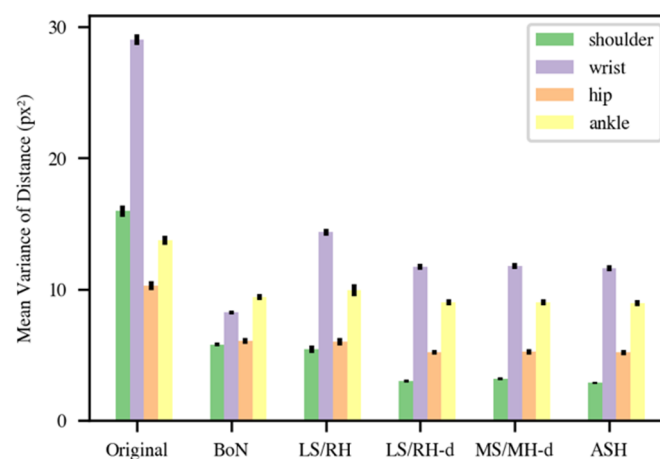


Figure 4. Mean variance of distance between shoulder (green), wrist (purple), hip (orange) and ankle (yellow) keypoints. (BoN: box normalization; LR/RH: left-shoulder right-hip normalization; LS/RH-d: left-shoulder right-hip distance normalization; MS/MH-d: mid-shoulder mid-hip distance normalization; ASH: average shoulder hip normalization).

We subsequently plotted the distance variance for EOA, DCD, and the controls separately in Figure 5 to better comprehend the variability across groups in mean variance of distance. Here, we assume that a good normalization method should (1) preserve the stability of the distance between the shoulders and between the hips within each of the three groups, resulting in minimal shoulder and hip distance variance, and (2) result in larger wrist and ankle distance variance in the EOA group than in the control group [1,2,10,33,34]. The BoN method cannot account for the expected high variability of wrist distance in the EOA group, so we suggest not using this method. The variability of wrist distance in EOA

after LS/RH, LS/RH-d, MS/MH-d, or ASH normalization is considerably larger than in the controls and DCD, as may be expected based on EOA phenomenology. Variance in wrist distance in the DCD and the control children shows different results with different normalization methods, which appears to be consistent with our predictions, given that it is difficult to diagnose DCD clinically. Furthermore, ankle variance in the EOA children is larger than in DCD and the control children after LS/RH-d, MS/MH-d, and ASH normalization, as may also be expected based on EOA phenomenology [1,33,34]. Finally, we performed Mann–Whitney U-tests for all three group pairs for ankle and wrist distance variance for each normalization method. The only significant differences found were for mean variance in ankle distance between the EOA and CON groups for the LS/RH-d and ASH normalization methods (LS/RH-d: $U = 44$, $p = 0.048$; ASH: $U = 42$, $p = 0.045$). As the goal of these tests was only to identify the best performing method(s), we did not correct for multiple comparisons. To conclude, both the LS/RH-d, and ASH normalization methods performed reasonably well and similarly in our task.

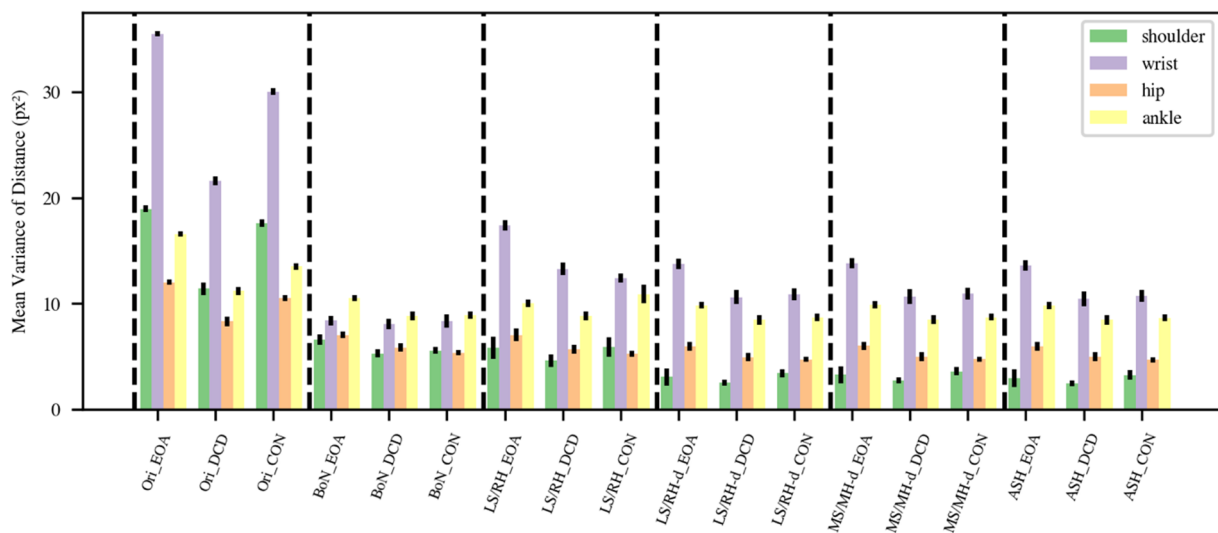


Figure 5. Mean variance of distance between groups. (Ori: original data; BoN: box normalization; LR/RH: left-shoulder right-hip normalization; LS/RH-d: left-shoulder right-hip distance normalization; MS/MH-d: mid-shoulder mid-hip distance normalization; ASH: average shoulder hip normalization).

3.4. Location Distribution Comparison

To better understand what our normalization is accomplishing, we plotted the distribution of the 17 keypoints before and after LS/RH-d normalization for one segment of a DCD child walking towards/away from the camera (See Figure 6). The coordinates on the left (before normalization) are spread widely for each keypoint because when the participant is farther away from the camera, the skeleton's size becomes smaller in the 2D video frame. After we used the LS/RH-d normalization method, each of the keypoints is localized more concentrated in a small region, resulting in a more stable, robust, and less noisy trajectory. This then allows the derivation of distance-based features and their comparison across different movement disorder groups or use for diagnostic classification.

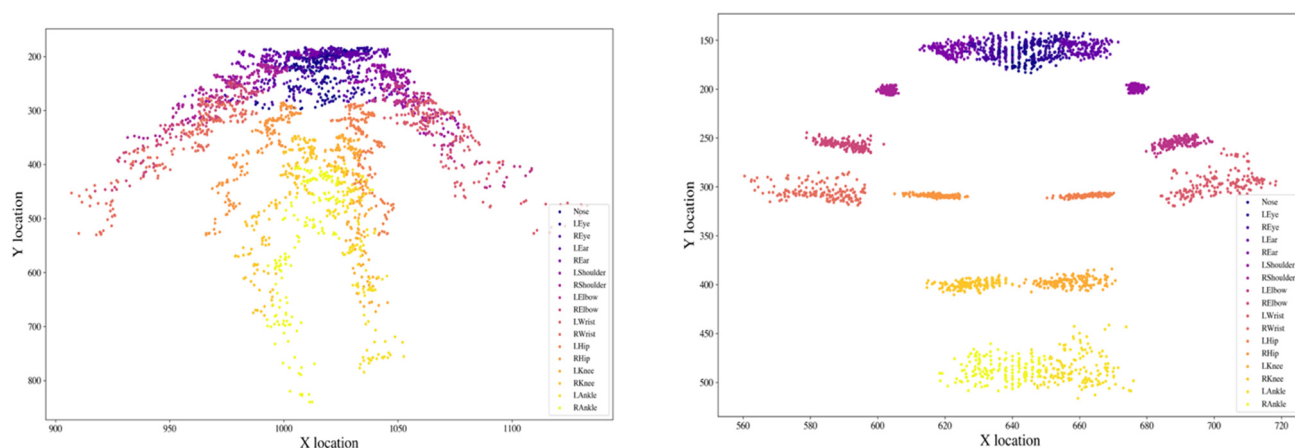


Figure 6. Distribution of the locations of the 17 keypoints before and after LS/RH-d normalization for one segment of a DCD child walking towards/away from the camera. (Left): before scaling; (right): after scaling.

4. Discussion

In this study, we present a framework based on 2D video pose estimation and proposed different normalization methods for skeleton data to enable the quantitative assessment of movement disorders. We found that AlphaPose and PoseTrack based keypoint localization can achieve relatively high confidence in coronal plane 2D recordings, and normalization methods in this framework could be helpful for the calculation of distance-based features.

The high likelihood after cleaning in our dataset from AlphaPose and PoseTrack with pretraining in the MSCOCO dataset supports the potential for quantitative analysis of routine video recordings for assistance in diagnosis and daily monitoring. Our video data, which included 15 children walking freely in a corridor, was captured with a single camera at a resolution of only 1280×720 pixels and a frame rate of 25 frames per second. The major reason the confidence cannot achieve values higher than 0.9 after cleaning is that keypoints on the face, such as the nose, eyes, and ears, cannot be seen or have low spatial resolution as people walk away from the camera. Yet, the likelihood findings show that even at low video resolution, the pose estimation model has a high level of reliability, and thus provides a solid platform for further study and demonstrates that it is promising for clinical application.

Assessing movement disorders objectively and accurately is notoriously difficult. Even though machine learning models have been used to predict gait characteristics [35], current models still rely on data from specialized hardware such as optical motion capture devices [36] and inertial measurement units [9]. Deep learning pose estimation approaches have been investigated in clinics for predicting gait parameters [37] and quantifying parkinsonian gait features [38]. However, the approach investigated in this study has not been adequately illustrated or explored, particularly in terms of how to extract distance-based movement features. It is feasible to reconstruct the three-dimensional kinematics of human movement by recording with a 3D capture system with several different cameras [39,40], or with a single RGB-D camera [41]. Previous research has already compared pose estimation from 2D and 3D video data and demonstrated that the 2D skeleton modality with proper preprocessing performed almost as well as the 3D-based method for an action recognition task [42]. In addition, an earlier study compared the validity of 2D and 3D analyses when recording rearfoot walking [43] and found no significant differences between the two approaches on variables commonly examined in rearfoot motion. Furthermore, the comparison of 2D and 3D video analysis during running [44], for angular measurements [45], and for sagittal plane gait assessment [46] supports that implementing 2D video analysis is more reasonable and feasible due to its convenient features compared with the high equipment cost and time required for 3D analysis, and might thus be clinically applicable. Based on this research conclusion, we investigated several normalization methods in 2D from the perspective of convenience and practicality for clinical use. To the best of our knowledge, this

is the first study to discuss how normalization may be used to extract distance-based features from 2D skeletons to produce interpretable movement features for clinical research.

This study has some limitations. First, the children with EOA were slightly older than the DCD and the control children. However, as healthy and DCD children grow up, the variability in coordinative movements declines [47,48], and thus the difference in variability in movement between EOA and DCD or the control children would be expected to be even larger if the DCD or the control children had been fully age-matched. Second, our long-term goal is to utilize ubiquitous surveillance cameras (such as mobile phones) to collect data and analyze it in a rapid and easy-to-use way; however, the present approach does not address how the camera's height and angle impact the results. When the camera moves to different angles, the size of the human skeleton also varies. This could be solved by using a tripod for recording, but this decreases ease of use. Third, the 2D skeleton data still contains some noise after cleaning, and the trajectories of keypoints through time are still not smooth after normalization; filtering could potentially be applied to deal with this. Further, here we express our distance-based measures in pixels and not (yet) in meters. This does allow consideration of variability in such measures, which, as argued, are clinically relevant parameters, but not yet absolute distances. In principle, such absolute distances could be obtained from 2D video frames if a calibration were performed, for example, by measuring the height of the participant and including a standing frame in the video clip. This would allow the expression of pixel sizes in meters. In the current study this was not yet included as we employed existing clinical footage. In further studies when collecting more data prospectively, this calibration step could be added, which would also allow comparison of distance-based features derived from 2D footage with a gold standard measure, such as those derived from 3D optical measurements.

Our normalization framework applied to 2D freehand single camera video recording data demonstrated a promising application for the quantitative assessment of movement disorders in a convenient and objective manner. For future further evaluation of the proposed normalization methods, after inclusion of a calibration step in prospectively collected video recordings, 3D optical systems employing reflective markers could be utilized to establish ground truth. As such, this study is a preliminary step towards quantitative analysis of clinically observed gait using single camera freehand video recording. In addition, if more data is collected, machine learning algorithms could be developed to help differentiate between different movement disorders.

Author Contributions: Conceptualization, W.T. and N.M.M.; methodology, W.T. and N.M.M.; formal analysis, W.T.; investigation, W.T.; data curation, N.M.M. and D.A.S.; writing—original draft preparation, W.T.; writing—review and editing, N.M.M., P.M.A.v.O. and D.A.S.; supervision, N.M.M., P.M.A.v.O. and D.A.S.; project administration, N.M.M.; funding acquisition, N.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the China Scholarship Council under grant number 202006150040.

Institutional Review Board Statement: The study was conducted in accordance with the principles of the Declaration of Helsinki (2013) and the research and integrity codes of the University Medical Center Groningen (UMCG). The Medical Ethical Committee of the UMCG provided a waiver for ethical approval since the SARA test battery was performed as part of the clinical routine.

Informed Consent Statement: All included children older than 12 years of age and all parents of the children included provided informed consent before study enrollment. Children younger than 12 years of age also provided informed assent.

Data Availability Statement: The original video datasets are not publicly available for privacy reasons, but the skeleton data after pose estimation presented in this study can be found at: <https://drive.google.com/file/d/1hOHnY8qrX4mQ8Lcjc3anOP8qvadCAthR/view?usp=sharing> (accessed on 15 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lawerman, T.F.; Brandsma, R.; Maurits, N.M.; Martinez-Manzanera, O.; Verschuuren-Bemelmans, C.C.; Luning, R.J.; Brouwer, O.F.; Kremer, H.P.; Sival, D.A. Paediatric motor phenotypes in early-onset ataxia, developmental coordination disorder, and central hypotonia. *Dev. Med. Child Neurol.* **2020**, *62*, 75–82. [[CrossRef](#)]
2. Ilg, W.; Timmann, D. Gait ataxia-specific cerebellar influences and their rehabilitation. *Mov. Disord.* **2013**, *28*, 1566–1575. [[CrossRef](#)]
3. Schmitz-Hübsch, T.; Du Montcel, S.T.; Baliko, L.; Berciano, J.; Boesch, S.; Depondt, C.; Giunti, P.; Globas, C.; Infante, J.; Kang, J.-S. Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology* **2006**, *66*, 1717–1720. [[CrossRef](#)]
4. Lawerman, T.F.; Brandsma, R.; Verbeek, R.J.; Van Der Hoeven, J.H.; Luning, R.J.; Kremer, H.P.H.; Sival, D.A. Construct Validity and Reliability of the SARA Gait and Posture Sub-scale in Early Onset Ataxia. *Front. Hum. Neurosci.* **2017**, *11*, 605. [[CrossRef](#)]
5. Bürk, K.; Sival, D.A. Scales for the clinical evaluation of cerebellar disorders. *Handb. Clin. Neurol.* **2018**, *154*, 329–339.
6. Brandsma, R.; Lawerman, T.F.; Kuiper, M.J.; Luning, R.J.; Burger, H.; Sival, D.A. Reliability and discriminant validity of ataxia rating scales in early onset ataxia. *Dev. Med. Child. Neurol.* **2017**, *59*, 427–432. [[CrossRef](#)]
7. Aguilar, V.S.; Manzanera, O.M.; Sival, D.A.; Maurits, N.M.; Roerdink, J.B.T.M. Distinguishing Patients with a Coordination Disorder from Healthy Controls Using Local Features of Movement Trajectories during the Finger-to-Nose Test. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1714–1722. [[CrossRef](#)]
8. Dominguez-Vega, Z.T.; Dubber, D.; Elting, J.W.J.; Sival, D.; Maurits, N.M. Instrumented classification of patients with early onset ataxia or developmental coordination disorder and healthy control children combining information from three upper limb SARA tests. *Eur. J. Paediatr. Neurol.* **2021**, *34*, 74–83. [[CrossRef](#)]
9. Mannini, A.; Martinez-Manzanera, O.; Lawerman, T.F.; Trojaniello, D.; Della Croce, U.; Sival, D.A.; Maurits, N.M.; Sabatini, A.M. Automatic classification of gait in children with early-onset ataxia or developmental coordination disorder and controls using inertial sensors. *Gait Posture* **2017**, *52*, 287–292. [[CrossRef](#)]
10. Pirker, W.; Katzenschlager, R. Gait disorders in adults and the elderly: A clinical guide. *Wien. Klin. Wochenschr.* **2016**, *129*, 81–95. [[CrossRef](#)]
11. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
12. Insafutdinov, E.; Pishchulin, L.; Andres, B.; Andriluka, M.; Schiele, B. Deepcruc: A deeper, stronger, and faster multi-person pose estimation model. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 34–50.
13. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [[CrossRef](#)]
14. Chen, Y.; Wang, Z.; Peng, Y.; Zhang, Z.; Yu, G.; Sun, J. Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7103–7112.
15. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017; pp. 7291–7299.
16. Fang, H.-S.; Xie, S.; Tai, Y.-W.; Lu, C. Rmpe: Regional multi-person pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2334–2343.
17. Joshi, D.; Merler, M.; Nguyen, Q.-B.; Hammer, S.; Kent, J.; Smith, J.R.; Feris, R.S. IBM High-Five: Highlights from Intelligent Video Engine. In Proceedings of the ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1249–1250.
18. Tian, L.; Cheng, X.; Honda, M.; Ikenaga, T. Multi-technology Correction Based 3D Human Pose Estimation for Jump Analysis in Figure Skating. In Proceedings of the Multidisciplinary Digital Publishing Institute Proceedings, West Lafayette, IN, USA, 22–26 June 2020; p. 95.
19. Chen, W.; Jiang, Z.; Guo, H.; Ni, X. Fall detection based on key points of human-skeleton using openpose. *Symmetry* **2020**, *12*, 744. [[CrossRef](#)]
20. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
21. Tian, Y.; Li, N.; Mao, B. Video Behavior Detection based on Optimized Alphapose in Electricity Facility Management. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Maebashi, Japan, 14–17 December 2021; pp. 364–367.
22. Zhang, C.; Yang, X. Bed-Leaving Action Recognition Based on YOLOv3 and AlphaPose. In Proceedings of the 2022 the 5th International Conference on Image and Graphics Processing (ICIGP), Beijing, China, 7–9 January 2022; pp. 117–123.
23. Lv, X.; Ta, N.; Chen, T.; Zhao, J.; Wei, H. Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation. *BioMed Res. Int.* **2022**, *2022*, 7020804. [[CrossRef](#)]
24. Peng, X.; Zhang, Y.; Song, L. Research on Gait Evaluation Method Based on Machine Vision. In Proceedings of the 2021 International Conference on Intelligent Computing, Automation and Applications (ICAA), Nanjing, China, 25–27 June 2021; pp. 357–362.
25. Sabo, A.; Mehdizadeh, S.; Iaboni, A.; Taati, B. Prediction of Parkinsonian Gait in Older Adults with Dementia using Joint Trajectories and Gait Features from 2D Video. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jalisco, Mexico, 26 July 2021; pp. 5700–5703.

26. Sabo, A.; Mehdizadeh, S.; Iaboni, A.; Taati, B. Estimating parkinsonism severity in natural gait videos of older adults with dementia. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 2288–2298. [[CrossRef](#)]
27. Whittle, M.W. Clinical gait analysis: A review. *Hum. Mov. Sci.* **1996**, *15*, 369–387. [[CrossRef](#)]
28. Mehdizadeh, S.; Nabavi, H.; Sabo, A.; Arora, T.; Iaboni, A.; Taati, B. Concurrent validity of human pose tracking in video for measuring gait parameters in older adults: A preliminary analysis with multiple trackers, viewing angles, and walking directions. *J. Neuro Eng. Rehabil.* **2021**, *18*, 139. [[CrossRef](#)]
29. Xiu, Y.; Li, J.; Wang, H.; Fang, Y.; Lu, C. Pose Flow: Efficient online pose tracking. *arXiv* **2018**, arXiv:1802.00977.
30. Schöls, L.; Bauer, P.; Schmidt, T.; Schulte, T.; Riess, O. Autosomal dominant cerebellar ataxias: Clinical features, genetics, and pathogenesis. *Lancet Neurol.* **2004**, *3*, 291–304. [[CrossRef](#)]
31. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Caesars Palace, NV, USA, 26 June–1 July 2016; pp. 770–778.
33. Buckley, E.; Mazzà, C.; McNeill, A. A systematic review of the gait characteristics associated with Cerebellar Ataxia. *Gait Posture* **2018**, *60*, 154–163. [[CrossRef](#)]
34. Baker, J.M. Gait disorders. *Am. J. Med.* **2018**, *131*, 602–607. [[CrossRef](#)]
35. Hannink, J.; Kautz, T.; Pasluosta, C.F.; Gaßmann, K.-G.; Klucken, J.; Eskofier, B.M. Sensor-based gait parameter extraction with deep convolutional neural networks. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 85–93. [[CrossRef](#)]
36. Li, X.X.R.W.M.; Jia-hua, S.C. Accuracy of the Microsoft KinectTM for measuring gait. *Eur. J. Med.* **2012**, *7*, 176–182.
37. Kidziński, Ł.; Yang, B.; Hicks, J.L.; Rajagopal, A.; Delp, S.L.; Schwartz, M.H. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nat. Commun.* **2020**, *11*, 4054. [[CrossRef](#)]
38. Sato, K.; Nagashima, Y.; Mano, T.; Iwata, A.; Toda, T. Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2D pose estimator. *PLoS ONE* **2019**, *14*, e0223549. [[CrossRef](#)]
39. Kwolek, B.; Michalczuk, A.; Krzeszowski, T.; Switonski, A.; Josinski, H.; Wojciechowski, K. Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition. *Multimed. Tools Appl.* **2019**, *78*, 32437–32465. [[CrossRef](#)]
40. Arac, A. Machine learning for 3D kinematic analysis of movements in neurorehabilitation. *Curr. Neurol. Neurosci. Rep.* **2020**, *20*, 29. [[CrossRef](#)]
41. Zhang, Y.; Funkhouser, T. Deep depth completion of a single rgb-d image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 175–185.
42. Elias, P.; Sedmidubsky, J.; Zezula, P. Understanding the limits of 2D skeletons for action recognition. *Multimed. Syst.* **2021**, *27*, 547–561. [[CrossRef](#)]
43. Cornwall, M.W.; McPoil, T. Comparison of 2-dimensional and 3-dimensional rearfoot motion during walking. *Clin. Biomech.* **1995**, *10*, 36–40. [[CrossRef](#)]
44. Dingenen, B.; Staes, F.F.; Santermans, L.; Steurs, L.; Eerdeken, M.; Geentjens, J.; Peers, K.H.; Thysen, M.; Deschamps, K. Are two-dimensional measured frontal plane angles related to three-dimensional measured kinematic profiles during running? *Phys. Ther. Sport* **2018**, *29*, 84–92. [[CrossRef](#)]
45. Daugaard, M.; Brandborg Nielsen, D. Comparison of Angular Measurements by 2D and 3D Gait Analysis. Available online: <https://www.diva-portal.org/smash/get/diva2:113552/FULLTEXT01.pdf> (accessed on 31 May 2022).
46. Saner, R.J.; Washabaugh, E.P.; Krishnan, C. Reliable sagittal plane kinematic gait assessments are feasible using low-cost webcam technology. *Gait Posture* **2017**, *56*, 19–23. [[CrossRef](#)]
47. Brandsma, R.; Lawerman, T.; Burger, H.; Burgerhof, J.; Sival, D. Age-related Reference Values for the Pediatric Scale for Assessment and Rating of Ataxia-A Multicenter Study. *Dev. Med. Child Neurol.* **2017**, *59*, 1077–1082.
48. Golenia, L.; Bongers, R.M.; van Hoorn, J.F.; Otten, E.; Mouton, L.J.; Schoemaker, M.M. Variability in coordination patterns in children with developmental coordination disorder (DCD). *Hum. Mov. Sci.* **2018**, *60*, 202–213. [[CrossRef](#)]