

Article

A Markerless Motion Measurement Method for Sport Climbing Using a Single RGB-D Camera and ICP-Based Model Fitting

Akihiro Kawamura ^{1,2,*} , Wataru Morinaga ³, Tomoro Nakamichi ³ and Ryo Kurazume ¹ ¹ Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan; kurazume@ait.kyushu-u.ac.jp² Department of Sport Science, Japan Institute of Sports Sciences, Tokyo 115-0056, Japan³ Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan

* Correspondence: kawamura@ait.kyushu-u.ac.jp

Abstract

Quantitative motion analysis is important for understanding the characteristics of climbing movement and for providing objective feedback in sport climbing. Although optical motion capture systems can measure three-dimensional body motion with high accuracy, their application to climbing is limited by the need for markers, multiple cameras, and a controlled measurement space, as well as by occlusion caused by the wall, holds, and body segments. A markerless measurement method using a compact sensor configuration would therefore be advantageous for analyzing climbing motion in more practical environments. This study presents a markerless three-dimensional motion measurement method for sport climbing using a single RGB-D camera and iterative closest point (ICP)-based body-part model fitting. The proposed method first separates the human region from the RGB image using image segmentation and then detects two-dimensional body keypoints using OpenPose. The detected keypoints are projected onto the depth image to reconstruct an initial three-dimensional posture. To refine the reconstructed posture, predefined body-part models are fitted to the segmented human point cloud using the ICP algorithm. The joint points included in the fitted body-part models are then used as the refined three-dimensional joint coordinates. Rather than relying solely on the accuracy of the image-based pose estimator, the proposed method introduces a model-based correction process that compensates for uncertainty in the initial keypoint-based reconstruction. This design aims to reduce the influence of background objects, unstable depth measurements, and incorrect or missing two-dimensional keypoints, which are common difficulties in climbing environments. The proposed method was quantitatively evaluated using five successfully completed trials from two participants under a low-occlusion condition and eight successfully completed trials from four participants under an occlusion-prone condition. The proposed method reduced the average three-dimensional RMSE across six representative body points relative to the initial RGB-D reconstruction under both conditions, with a more pronounced reduction under the occlusion-prone condition. The results indicate that ICP-based body-part model fitting is a useful correction step for single-camera RGB-D motion measurement in sport climbing.



Academic Editors: Nuno Mateus and Jose Luis Felipe Hernández

Received: 24 June 2026

Revised: 9 September 2026

Accepted: 10 September 2026

Published: 11 September 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.**Keywords:** sport climbing; markerless motion measurement; RGB-D camera; OpenPose; segmentation; ICP-based model fitting

1. Introduction

Sport climbing requires complex whole-body coordination of the upper limbs, lower limbs, and trunk, together with continuous interaction with the climbing wall. During climbing, athletes repeatedly change hand and foot contacts, transfer body weight, generate propulsive force, and stabilize their posture under constrained environmental conditions. Therefore, quantitative analysis of climbing motion is important for understanding movement strategies, evaluating technical skill, and providing objective feedback for training and coaching [1–3].

Three-dimensional motion capture systems have been widely used for biomechanical analysis because they can measure body motion with high accuracy. In climbing research, such systems can provide detailed information on joint positions, body posture, and movement trajectories [2,4]. However, their application to climbing motion is often limited by several practical constraints. Optical motion capture generally requires reflective markers to be attached to the body and multiple cameras to be arranged around the measurement space. In climbing environments, camera placement is restricted by the wall, and body segments are frequently occluded by the wall, holds, or other body parts. In addition, marker attachment may affect natural climbing motion, and the measurement area is usually limited to a controlled laboratory setting. These constraints make it difficult to apply conventional motion capture systems to more practical climbing environments [5–7].

Markerless motion measurement has attracted attention as an alternative approach for analyzing human movement without attaching markers to the body. Recent advances in image-based pose estimation have enabled two-dimensional body keypoints to be detected from monocular color images, and RGB-D cameras can provide depth information for reconstructing three-dimensional posture [8–10]. Such systems are relatively compact and easy to install compared with optical motion capture systems, suggesting their potential for practical motion analysis in sports environments [6,7,11]. In particular, a single-view RGB-D setup may be useful for climbing analysis because it can reduce the complexity of camera arrangement and measurement preparation. A recent review has highlighted the increasing use of commercial vision sensors and AI-based pose estimation frameworks for markerless motion analysis in sports and exercise [12].

Recently, SAM 3D Body was proposed as a general-purpose method for recovering a full-body human mesh and articulated pose from a single RGB image without requiring a depth sensor [13]. Although it represents an important recent advance in monocular human mesh recovery, its formulation and output representation differ from the metric joint-position measurement considered in this study. The present study instead focuses on correcting RGB-D-based joint coordinates in a climbing-specific environment by fitting body-part models directly to an observed depth point cloud.

RGB-D videos and pose estimation frameworks have also been used for automated segmentation of climbing movements, indicating the potential of vision-based approaches for climbing motion analysis [14]. A recent multimodal climbing motion dataset has addressed the need to capture human–scene interactions during climbing by including LiDAR point clouds, RGB videos, reconstructed scene meshes, and frame-wise annotations of contacted holds [15]. Previous studies have also attempted to apply RGB-D-based markerless measurement to sport climbing. For example, Pandurevic et al. introduced a measurement system for quantitative analysis of force and technique in competitive sport climbing and estimated three-dimensional body posture by combining two-dimensional pose estimation with depth information [16]. In that approach, the depth direction was corrected using offsets based on body thickness. Although this approach reduced the complexity of the measurement setup, fixed-depth correction may not sufficiently account for posture-dependent changes in the relationship between the visible body surface and

the anatomical joint centers. In addition, errors or missing keypoints in two-dimensional pose estimation can directly affect the reconstructed three-dimensional posture.

Several conditions specific to climbing make three-dimensional posture estimation particularly challenging. Unlike many general movements performed in open space, climbing motion involves close interaction between the body and a vertical or inclined wall. The climber's limbs frequently contact the wall and holds, and the body often moves close to the wall surface. As a result, depth measurements may be unstable, and some body segments may be partially hidden from the camera. Furthermore, two-dimensional pose estimation can be affected by unusual postures, self-occlusion, and the limited visibility of limbs [3,7,11]. Therefore, simply applying general-purpose pose estimation methods to climbing motion may not provide sufficiently consistent three-dimensional posture estimates.

To address these issues, this study presents a markerless three-dimensional motion measurement method for climbing using a single RGB-D camera and ICP-based body-part model fitting. The method first extracts the human region from each RGB image by image segmentation and detects two-dimensional body keypoints using OpenPose. The detected keypoints are projected onto the depth image to reconstruct an initial three-dimensional posture. Predefined body-part models are then fitted to the segmented human point cloud using the ICP algorithm, and the joint points included in the fitted models are used as the corrected joint coordinates. The image-based pose estimator itself is not modified; instead, the geometric consistency between the observed point cloud and the body-part models is used to reduce errors caused by incorrect or missing keypoints and by unstable depth measurements.

The contribution of this study is an application-oriented geometric correction framework for sport climbing motion measurement. Unlike a fixed depth-direction offset, the proposed method performs posture-dependent correction by fitting subject-specific body-part models to the observed human point cloud. This enables the discrepancy between the visible body surface and the anatomical joint position to be reduced while retaining a compact, markerless, single-camera RGB-D configuration. The purpose of this study is to develop a practical markerless motion measurement method for sport climbing and to evaluate its applicability in climbing environments.

2. Methods

2.1. Overview of the Proposed Method

This study proposes a markerless three-dimensional motion measurement method for sport climbing using a single RGB-D camera and ICP-based model fitting. The proposed method estimates the three-dimensional posture of a climber through five main processes: RGB-D image acquisition, human region segmentation, two-dimensional pose estimation, initial three-dimensional reconstruction using depth information, and body-part model fitting using the iterative closest point (ICP) algorithm. The overall procedure is designed to improve the reliability of keypoint detection and subsequent RGB-D-based three-dimensional reconstruction in climbing environments, where cluttered backgrounds with multiple holds, partial occlusion, and large variations in body posture can cause false or missing keypoint detections. Figure 1 shows the overall procedure of the proposed method.

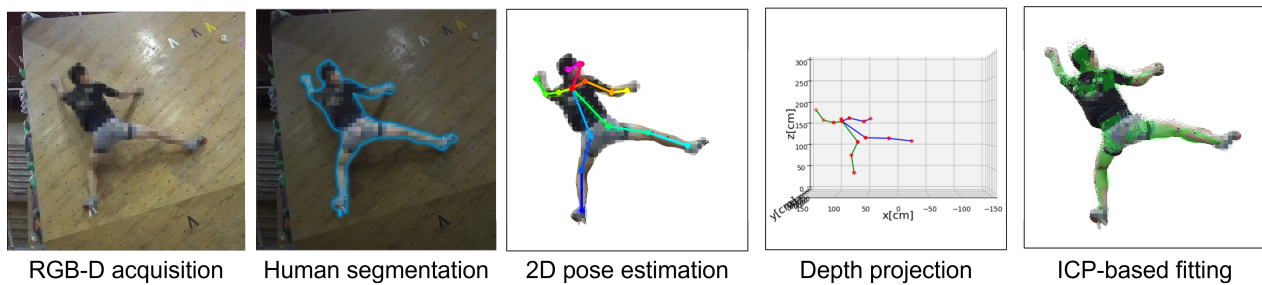


Figure 1. Overview of the proposed markerless motion measurement method. The method consists of RGB-D image acquisition, human region segmentation, OpenPose-based two-dimensional pose estimation, depth projection, and ICP-based body-part model fitting.

2.2. Measurement System

The proposed system uses a single ZED 2i Stereo Camera (Stereolabs, San Francisco, CA, USA) as the RGB-D sensor. The ZED 2i estimates depth from the disparity between its two RGB images. Before selecting the RGB-D sensor, preliminary measurements were conducted using an Azure Kinect DK (Microsoft, Redmond, WA, USA) and an Intel RealSense L515 (Intel Corporation, Santa Clara, CA, USA). In the climbing environment used in this study, these sensors frequently produced missing or unstable depth values, particularly for dark hair, dark clothing, and some climbing holds. Based on these preliminary observations, the ZED 2i stereo camera was selected for the experiments. These preliminary tests were conducted for sensor selection and were not designed as a quantitative comparison of RGB-D sensors.

The RGB-D camera was used to acquire RGB images and depth images of climbing motion. The image resolution and frame rate were set to 1280×720 pixels and 60 fps, respectively. This setting was selected to balance the temporal resolution required for capturing dynamic climbing movements and the image quality required for pose estimation. The camera was placed approximately 3–4 m from the climber during the experiments to capture the entire climbing motion within the field of view while maintaining sufficient image resolution for pose estimation. A representative view of the measurement environment is shown in Figure 2b. Seven Vicon cameras (Vicon Motion Systems Ltd., Oxford, UK) were installed around the measurement area as the reference optical motion capture system. The camera coordinate system of the RGB-D camera was used as the reference coordinate system for the reconstructed three-dimensional posture. The directions of the coordinate axes are shown in Figure 2b, and the same coordinate system was used for the coordinate comparisons presented in the Results section.

2.3. Human Region Segmentation

In climbing environments, two-dimensional pose estimation from RGB images can be affected by the surrounding environment. The climbing wall may have colors similar to human skin, and climbing holds have various colors and shapes. In addition, shadows are often pronounced because the climber moves close to the wall. These factors make it difficult to separate the climber from the background using simple background subtraction.

To reduce the influence of the surrounding environment, the human region was segmented before two-dimensional pose estimation. In this study, Language Segment-Anything (LangSAM) [17], which combines language-guided object detection using Grounding DINO [18] with Segment Anything [19], was used to extract the human region from each RGB image. The text prompt “person” was provided to LangSAM to automatically detect and segment the climber. Therefore, manual point or bounding-box prompts were not provided for individual frames. The segmented image was subsequently used as the input for two-dimensional pose estimation. This preprocessing was intended to

make keypoint detection in climbing scenes less sensitive to the cluttered background produced by the climbing wall and holds. An example of the segmentation process is included in Figure 1. In the present implementation, human region segmentation was performed as a preprocessing step before the OpenPose-based pose estimation and ICP-based fitting processes.

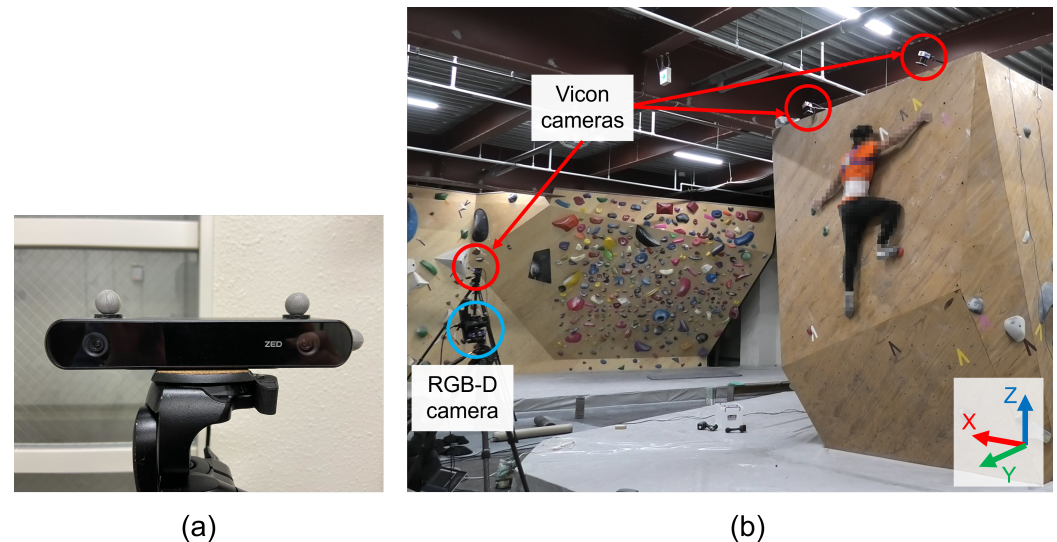


Figure 2. Measurement system and experimental setup. (a) ZED 2i stereo camera used as the RGB-D sensor. (b) Representative view of the measurement environment. The RGB-D camera was placed in front of the climbing wall, and seven Vicon cameras were installed around the measurement area as the reference optical motion capture system. Some of the Vicon cameras are visible in the image. The coordinate axes shown in the lower-right corner indicate the camera coordinate system used for the reconstructed three-dimensional posture and the coordinate comparisons in the Results section.

2.4. Initial 3D Pose Estimation Using OpenPose and Depth Projection

Two-dimensional body keypoints were detected from the segmented RGB image using OpenPose version 1.7.0 with the COCO 18-keypoint model. These two-dimensional keypoints were used as the initial observations for three-dimensional posture reconstruction.

For each detected keypoint, the corresponding depth value was obtained from the depth image. The image coordinates and depth value were then converted into three-dimensional coordinates using the intrinsic parameters of the RGB-D camera. Let (u, v) be the image coordinates of a keypoint, d be the corresponding depth value, and f_x, f_y, c_x , and c_y be the intrinsic parameters of the RGB-D camera. The three-dimensional coordinates (X, Y, Z) in the camera coordinate system are calculated as follows:

$$X = \frac{(u - c_x)d}{f_x}, \quad Y = \frac{(v - c_y)d}{f_y}, \quad Z = d. \quad (1)$$

This projection process provides an initial three-dimensional skeletal posture. However, the reconstructed joint positions do not necessarily correspond to the anatomical joint centers because the depth image represents the visible body surface. In addition, OpenPose may fail to detect some keypoints or may incorrectly estimate keypoint positions when body parts are occluded or when the posture is complex. Therefore, the initial three-dimensional posture was used as the initial value for ICP-based body-part model fitting. For comparison with a previous RGB-D-based approach, a depth-direction correction based on the method reported by Pandurevic et al. [16] was also applied to the initial three-dimensional joint coordinates. In this comparison, the two-dimensional keypoints and depth values were obtained using the same OpenPose-based detection and depth projection procedure as in

the initial RGB-D reconstruction, so that the difference between the compared methods was limited to the correction process after depth projection. The method of Pandurevic et al. [16] describes a depth-direction correction based on body dimensions but does not provide a complete anatomical definition of the body-part thicknesses or a unique procedure for determining the corresponding offsets. In the present implementation, segment-specific thickness parameters corresponding to the body-part models were used for the depth-direction correction. The measurement locations used to determine body-part thicknesses were not defined using standardized anatomical landmarks.

2.5. ICP-Based Body-Part Model Fitting

To refine the initial three-dimensional posture, ICP-based body-part model fitting was performed. The proposed method uses the point cloud obtained from the RGB-D camera and predefined body-part models. The body-part models represent 11 body segments: the head, trunk, and pelvis, together with the left and right upper arms, forearms, thighs, and shanks. Figure 3 shows the body-part models used in this study. Each model includes joint points corresponding to the anatomical endpoints of the segment. The models were scaled and deformed based on the participant's body dimensions. Body dimensions that could be measured directly were measured from each participant using a steel tape measure or digital calipers, depending on the dimension, and were used to scale or deform the corresponding body-part models.

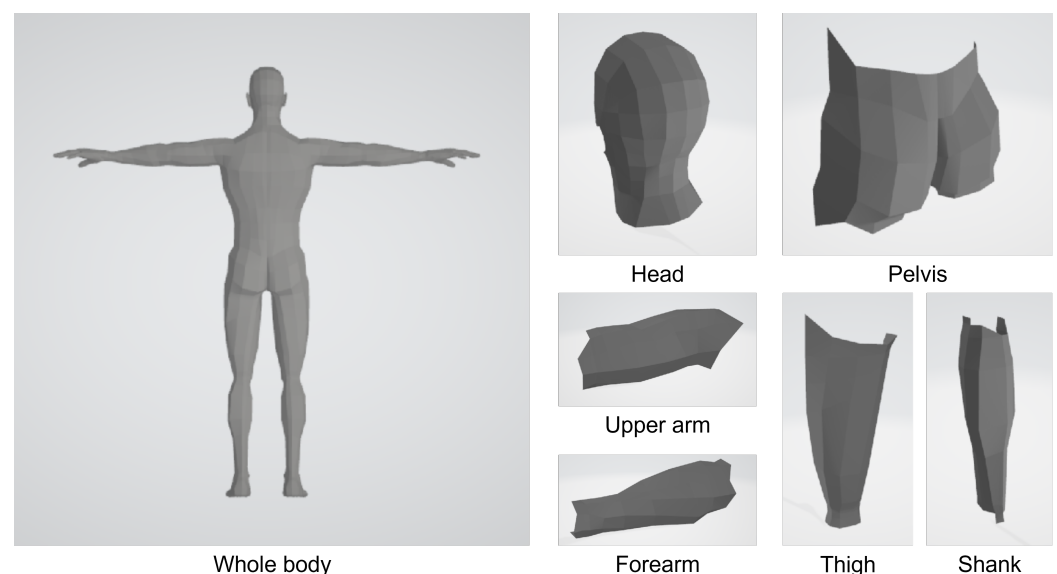


Figure 3. Body-part models used for ICP-based fitting. The predefined body model consisted of 11 body parts: the head, trunk, and pelvis, together with the left and right upper arms, forearms, thighs, and shanks. Each part model was scaled according to the participant's body dimensions and used for fitting to the segmented human point cloud.

Because the proposed system uses a single RGB-D camera, the observed point cloud contains only the body surface visible from the camera. Accordingly, before ICP fitting, the non-visible surface of each body-part model was removed and only the camera-facing surface was retained. This step makes the model geometry consistent with the single-view observation geometry; retaining the non-visible surface would introduce model points for which no corresponding observed surface exists and could produce inappropriate nearest-neighbor correspondences during point-to-point ICP.

The ICP algorithm was applied independently to each body-part model in each frame using the point-to-point formulation implemented in Open3D. The segmented human point

cloud was not explicitly partitioned into body-part-specific target point clouds. Instead, the same full-body human point cloud was used as the target point cloud for all 11 body-part models. The target point cloud had already been restricted to the segmented human region using the LangSAM mask. No additional statistical outlier removal or robust weighting was applied beyond the maximum correspondence-distance threshold.

The trunk model was initialized at the midpoint of the bilateral shoulder keypoints and oriented using the vector connecting the shoulders. The pelvis model was initialized at the midpoint between the bilateral shoulder midpoint and the bilateral hip midpoint and oriented using the vector from the hip midpoint to the shoulder midpoint. The head model was initialized from the bilateral shoulder midpoint with fixed offsets of 0.10 m upward and 0.15 m in the depth direction, without rotational initialization. For each ICP iteration, point correspondences were determined by nearest-neighbor search between the body-part model and the observed human point cloud. Correspondence pairs with a point-to-point distance greater than 0.1 m were rejected. Starting from the initial position and orientation described above, each body-part model was rigidly aligned with the full-body human point cloud by minimizing the point-to-point distances between the remaining corresponding points. After the alignment, the joint points included in the fitted body-part models were used to calculate the refined joint coordinates. For a joint shared by two adjacent body-part models, such as the elbow or knee, the arithmetic mean of the two joint coordinates obtained from the independently fitted models was used as the final joint coordinate. No additional joint-angle or kinematic-chain constraints were imposed during ICP fitting.

ICP requires an initial value that is reasonably close to the actual posture, so the initial three-dimensional posture was checked before fitting. The check tested whether the estimated body-part configuration was anatomically consistent, i.e., whether the left and right limbs had been interchanged, whether a keypoint lay outside the expected range, and whether any body part was undetected. When the initial estimate failed this check, the ICP result from the preceding frame was used as the initial value instead. This procedure allowed posture estimation to continue even when OpenPose failed to detect some keypoints in a frame.

The maximum number of ICP iterations was set to 80 for each body-part model and frame. Numerical convergence was evaluated based on the change in the total RMSE over the corresponding point pairs. At each iteration, the absolute difference between the total RMSE obtained in the current iteration and that obtained in the immediately preceding iteration was calculated. The ICP procedure was terminated when either 80 iterations were completed or the RMSE difference remained below 1.0×10^{-6} m for five consecutive iterations. The threshold was used as a numerical convergence tolerance rather than as a measure of physical positioning accuracy.

2.6. Implementation and Computational Cost

All processing was performed offline on a workstation equipped with an 11th Gen Intel Core i9-11900K CPU operating at 3.50 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU. The computational cost was measured using the same 240-frame segment of a climbing-motion sequence.

Generating the human-region masks using LangSAM required 15 min 30 s, corresponding to 3.875 s per frame. OpenPose required 58 s, corresponding to 0.242 s per frame. Generating PLY point-cloud files containing only the segmented human region from the ZED data required 24 min 36 s, corresponding to 6.150 s per frame. ICP-based fitting of the 11 body-part models required 52 min 53 s, corresponding to 13.221 s per frame.

The total processing time for the 240-frame segment was 93 min 57 s, corresponding to 23.488 s per frame. Therefore, the current implementation is intended for offline analysis and does not operate in real time.

3. Experiments and Results

3.1. Experimental Conditions

The proposed method was evaluated using climbing motion data measured in a climbing environment. An optical motion capture system was used as the reference for evaluating the estimated posture. The proposed RGB-D-based system and the reference motion capture system were synchronized using timestamps. Reflective markers were attached to the RGB-D camera body, and the pose of the RGB-D camera was measured using the Vicon system. The rigid transformation between the Vicon coordinate system and the RGB-D camera coordinate system was then determined from the measured camera pose. The Vicon-based joint coordinates were transformed into the RGB-D camera coordinate system using this transformation, enabling direct comparison with the coordinates estimated from the RGB-D images.

Four male participants took part in the experiments. At the time of measurement, their age was 22.0 ± 2.4 years (range: 19–25 years), height was 171.5 ± 4.4 cm, and body mass was 60.8 ± 5.0 kg. They had approximately 2–14 years of climbing experience. The participants ranged from advanced to elite-level boulderers.

Climbing motion data were collected under two hold configurations that produced different levels of body-part occlusion. In the low-occlusion condition, two participants performed three trials each, resulting in six recorded trials. In the occlusion-prone condition, four participants performed three trials each, resulting in 12 recorded trials.

A trial was considered successfully completed when the participant reached the designated endpoint of the prescribed climbing task. Trials in which the participant did not reach the endpoint were excluded from the quantitative analysis. Completion was determined during the experiment from the observed trial outcome. No trial was excluded on the basis of RGB-D data quality, human-region segmentation performance, or pose-estimation quality. All successfully completed trials were included in the quantitative analysis. Consequently, five of the six recorded trials in the low-occlusion condition and eight of the twelve recorded trials in the occlusion-prone condition were successfully completed and analyzed. In the low-occlusion condition, the two participants contributed three and two retained trials, respectively. In the occlusion-prone condition, each of the four participants contributed two retained trials. The two participants in the low-occlusion condition were included among the four participants in the occlusion-prone condition. All retained trials were processed using the initial RGB-D reconstruction, the method of Pandurevic et al. [16], the proposed method, and the Vicon reference. For each body point and trial, the three methods were compared using the same set of valid frames.

The quantitative evaluation included the center of mass, neck, right and left elbows, and right and left knees. The center of mass obtained from the Vicon software was used as the reference center of mass. For the RGB-D-based methods, the center of mass was calculated as the weighted average of the estimated body segment centers using segmental mass ratios. Representative sequences from the two conditions were additionally used to illustrate the coordinate trajectories and the behavior of the proposed method under different visibility conditions.

3.2. Model Fitting Results

Figure 4 shows an example of ICP-based body-part model fitting. The body-part models were initially placed using the three-dimensional posture reconstructed from

OpenPose and the depth image. After ICP-based fitting, the body-part models were aligned with the segmented human point cloud. The fitted models followed the surface point cloud of the climber, particularly around the lower limbs and trunk, where the depth correction was visually apparent.

In this example, the ICP algorithm aligned the predefined body-part models with the observed human point cloud. The joint points included in the fitted body-part models were then used as the refined joint coordinates. Therefore, the proposed method corrected the initial posture estimated from the RGB-D image and generated a three-dimensional posture more consistent with the observed human point cloud.

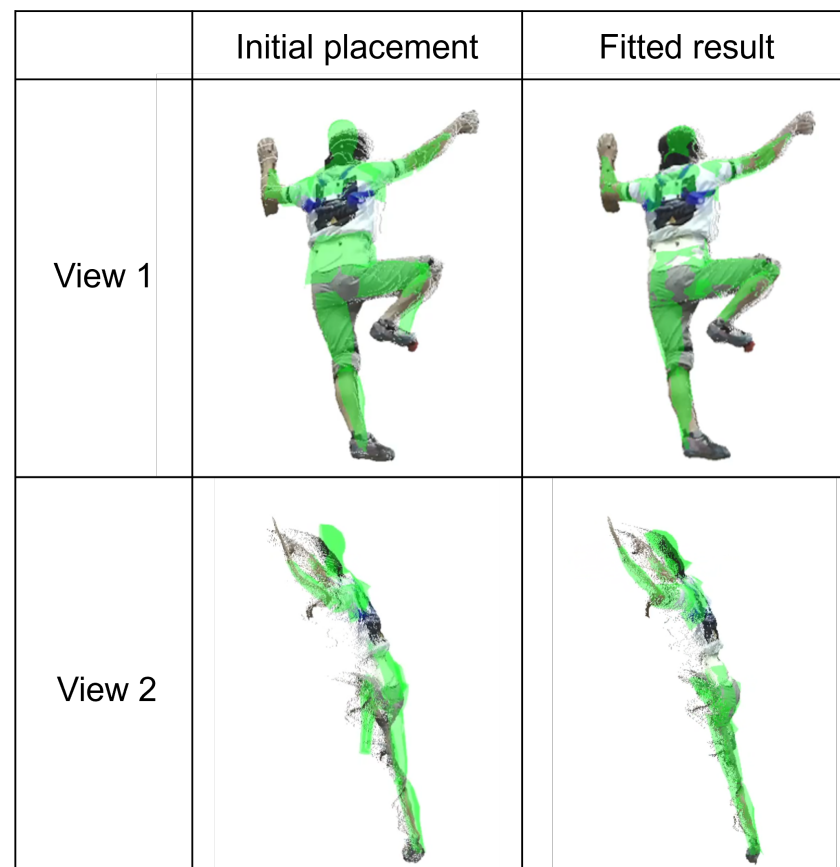


Figure 4. Example of ICP-based body-part model fitting shown from two viewpoints. The initial placement of the body-part models and the fitted result after ICP-based correction are compared. The body-part models are shown in green; the remaining points correspond to the observed human point cloud.

3.3. Quantitative Comparison with Reference Motion Capture Data

The coordinate accuracy of the proposed method was quantitatively evaluated across multiple participants and trials. The three-dimensional RMSE was calculated for six representative body points—the center of mass, neck, right and left elbows, and right and left knees—in five low-occlusion trials and eight occlusion-prone trials. For each condition, the mean and standard deviation of the trial-level RMSE were calculated.

Figure 5 shows a representative climbing motion from one trial in the low-occlusion condition, and Figures 6 and 7 show the corresponding coordinate trajectories. These figures are presented to illustrate the differences among the initial RGB-D reconstruction, the depth-direction correction method of Pandurevic et al. [16], and the proposed method. The aggregated quantitative results across all 13 analyzed trials are presented in Table 1.

Figure 6 shows the coordinate comparison of the right elbow obtained using Vicon, the initial RGB-D reconstruction based on OpenPose keypoints and depth projection, the

method reported by Pandurevic et al. [16], and the proposed method. In this example, the previous depth-direction correction method mainly changed the depth-direction coordinate, reflecting its correction strategy based on body-part thickness. The proposed method reduced the deviation from the Vicon data by fitting the body-part model to the observed point cloud. This example indicates that the model-based correction reduced the coordinate deviation of an individual joint.

Figure 7 shows the comparison of center-of-mass coordinates obtained using Vicon, the initial RGB-D reconstruction, the previous depth-direction correction method, and the proposed method. The gray regions indicate frames in which OpenPose keypoints were missing, and therefore the initial RGB-D-based center-of-mass estimate could not be reliably obtained. Because the center of mass was calculated from the estimated body segment centers, missing or unreliable keypoints in any body segment can affect the reliability of the center-of-mass estimation, although the degree of influence differs depending on the segmental mass ratio. These intermittent gray regions therefore highlight a practical limitation of RGB-D reconstruction based on detected keypoints in climbing motion. In contrast, the proposed method uses ICP-based body-part model fitting and can continue posture refinement based on the segmented point cloud and the previous fitting result when the initial keypoint-based reconstruction is unreliable. In the center-of-mass coordinates, the proposed method reduced the deviation from the Vicon data compared with the initial RGB-D reconstruction and the previous depth-direction correction method.

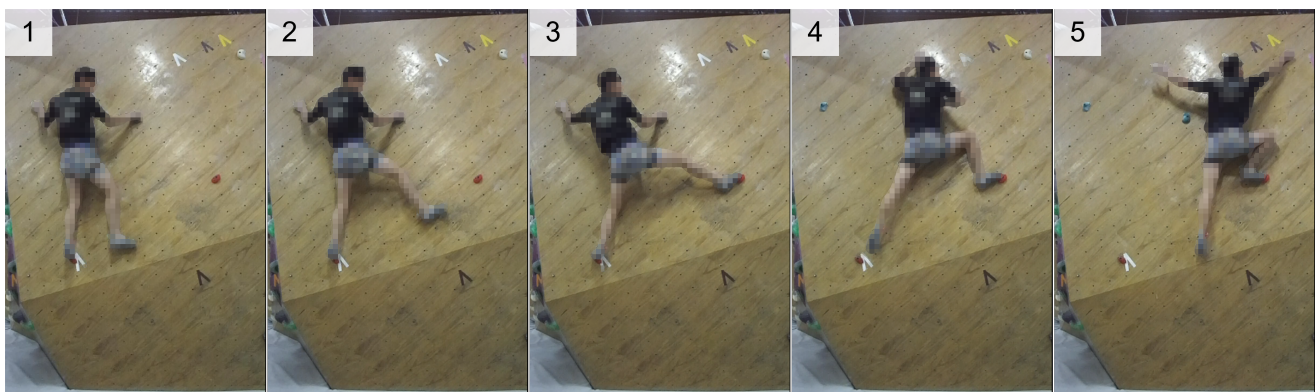


Figure 5. Representative climbing motion from the low-occlusion condition used to illustrate the coordinate trajectories. Panels 1–5 show successive frames of the representative climbing motion.

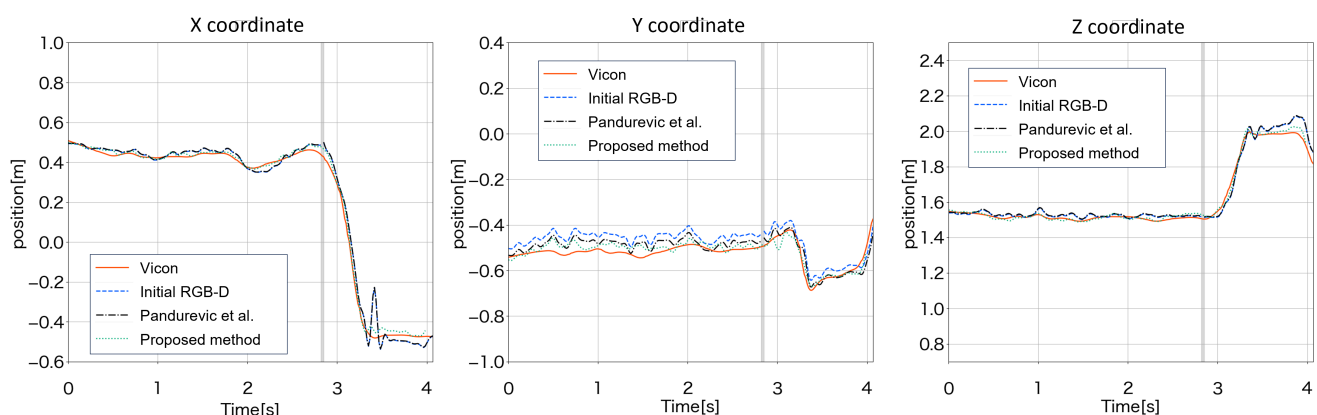


Figure 6. Coordinate comparison of the right elbow in a representative low-occlusion trial obtained using Vicon, the initial RGB-D reconstruction based on OpenPose keypoints and depth projection, the method reported by Pandurevic et al. [16], and the proposed method. The right elbow was selected as a representative joint point for evaluating local joint-coordinate estimation. The gray shaded regions indicate frames in which OpenPose keypoints were missing.

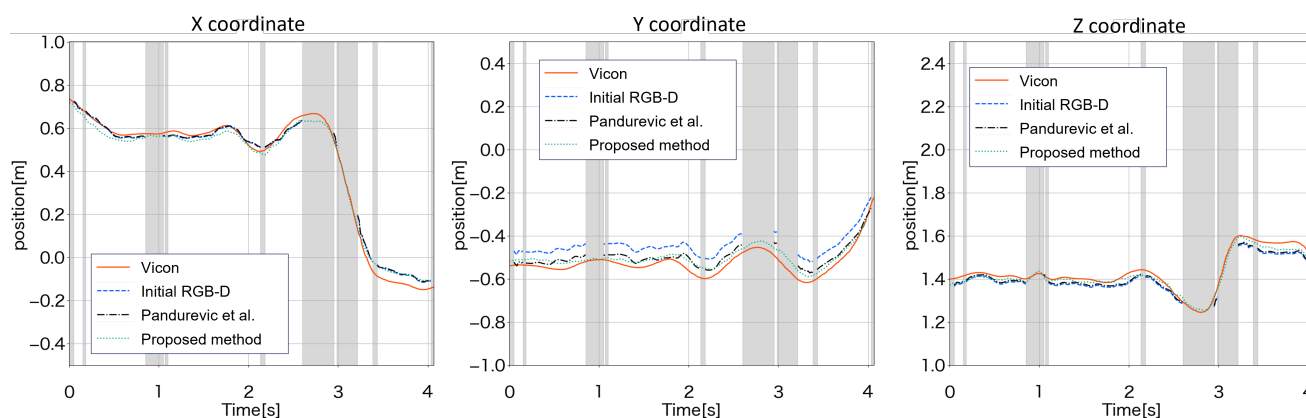


Figure 7. Coordinate comparison of the center of mass in a representative low-occlusion trial, obtained using Vicon, the initial RGB-D reconstruction based on OpenPose keypoints and depth projection, the method of Pandurevic et al. [16], and the proposed method. The gray shaded regions indicate frames in which OpenPose keypoints were missing. The center of mass was used to evaluate the consistency of the estimated whole-body posture.

Table 1. Three-dimensional RMSE of six representative body points under the low-occlusion and occlusion-prone conditions. Values are presented as the mean $\pm$ standard deviation across trials.

Body Point	Initial RGB-D [m]	Pandurevic et al. [16] [m]	Proposed Method [m]
(a) Low-occlusion condition: five successfully completed trials from two participants			
Center of mass	0.107 $\pm$ 0.017	0.062 $\pm$ 0.015	0.061 $\pm$ 0.013
Neck	0.220 $\pm$ 0.021	0.174 $\pm$ 0.022	0.110 $\pm$ 0.009
Right elbow	0.087 $\pm$ 0.014	0.078 $\pm$ 0.016	0.102 $\pm$ 0.107
Left elbow	0.096 $\pm$ 0.019	0.085 $\pm$ 0.014	0.086 $\pm$ 0.049
Right knee	0.100 $\pm$ 0.031	0.073 $\pm$ 0.020	0.065 $\pm$ 0.034
Left knee	0.101 $\pm$ 0.028	0.065 $\pm$ 0.023	0.075 $\pm$ 0.035
Mean across six body points	0.119 $\pm$ 0.010	0.089 $\pm$ 0.007	0.083 $\pm$ 0.027
(b) Occlusion-prone condition: eight trials from four participants			
Center of mass	0.171 $\pm$ 0.169	0.139 $\pm$ 0.184	0.085 $\pm$ 0.042
Neck	0.257 $\pm$ 0.136	0.218 $\pm$ 0.148	0.116 $\pm$ 0.077
Right elbow	0.194 $\pm$ 0.197	0.196 $\pm$ 0.199	0.161 $\pm$ 0.144
Left elbow	0.198 $\pm$ 0.180	0.195 $\pm$ 0.183	0.144 $\pm$ 0.156
Right knee	0.161 $\pm$ 0.101	0.143 $\pm$ 0.112	0.084 $\pm$ 0.055
Left knee	0.200 $\pm$ 0.139	0.177 $\pm$ 0.151	0.128 $\pm$ 0.144
Mean across six body points	0.197 $\pm$ 0.150	0.178 $\pm$ 0.158	0.120 $\pm$ 0.098

For each body point, values indicate the mean and standard deviation of the trial-level three-dimensional RMSE. The “Mean across six body points” was first calculated for each trial as the arithmetic mean of the RMSE values for the six evaluated body points; the table reports the mean and standard deviation of these trial-level values. Bold values indicate the lowest mean RMSE in each row.

In the representative coordinate trajectories shown in Figures 6 and 7, the effect of the proposed method was particularly apparent in the depth direction. Because the initial RGB-D-based estimation assigns the depth value of the visible body surface to each detected keypoint, the estimated coordinates may deviate from the reference motion capture data. The ICP-based body-part model fitting reduced this deviation by fitting the body-part models to the observed point cloud and using the joint points of the fitted models as corrected coordinates.

The gray regions in Figure 7 indicate frames in which the initial RGB-D reconstruction could not be reliably obtained because of missing OpenPose keypoints. The RMSE for each body point and trial was calculated using the common valid frames available for all three methods and the Vicon reference.

Table 1 summarizes the three-dimensional RMSE across the analyzed trials. In the low-occlusion condition, the mean RMSE across the six body points was 0.119 ± 0.010 m for the initial RGB-D reconstruction, 0.089 ± 0.007 m for the method of Pandurevic et al. [16], and 0.083 ± 0.027 m for the proposed method. The proposed method reduced the trial-level mean RMSE relative to the initial RGB-D reconstruction in all five trials and relative to the method of Pandurevic et al. [16] in four of the five trials. At the individual-body-point level, the proposed method showed the lowest mean RMSE for the center of mass, neck, and right knee. However, the improvement was not uniform across all body points, and relatively large inter-trial variation remained for the elbow estimates.

In the occlusion-prone condition, the mean RMSE across the six body points was 0.197 ± 0.150 m for the initial RGB-D reconstruction, 0.178 ± 0.158 m for the method of Pandurevic et al. [16], and 0.120 ± 0.098 m for the proposed method. The proposed method reduced the trial-level mean RMSE relative to the initial RGB-D reconstruction in all eight trials and relative to the method of Pandurevic et al. in seven of the eight trials. Furthermore, the proposed method showed the lowest mean RMSE for all six evaluated body points under this condition.

These descriptive results indicate that the benefit of ICP-based body-part model fitting was more evident under the occlusion-prone condition. At the same time, the inter-trial variation and the absence of improvement for some body points under the low-occlusion condition show that the effectiveness of the correction depends on the visibility of the target body part and the quality of the ICP initialization.

3.4. Effect of the Initialization Strategy

To examine the contribution of frame-wise OpenPose-based initialization, an additional comparison was performed using the five low-occlusion trials. In the standard proposed method, the initial posture was obtained from OpenPose and depth projection in each frame, while the previous-frame ICP result was used when the initial estimate was judged unreliable. In the additional condition, OpenPose and depth projection were used only for the first analyzed frame, and the ICP result from the preceding frame was used as the initial value for all subsequent frames.

The first analyzed frame was defined as the earliest frame in which a complete initial posture could be reconstructed using OpenPose and depth projection. In some trials, the analysis was started several frames after the beginning of the recorded sequence because the full-body posture was not available in the first recorded frame. For each trial, the two initialization strategies were evaluated over the same frame interval.

Figure 8 compares the right-elbow coordinates obtained using the two initialization strategies. With OpenPose-based initialization in each frame, the estimated coordinates remained close to the Vicon reference throughout the representative sequence. In contrast, when OpenPose and depth projection were used only in the first analyzed frame, the estimated coordinates gradually deviated from the reference and did not follow the rapid coordinate change at approximately 3 s, particularly in the X and Z directions. These results indicate that propagation of the preceding-frame ICP result alone can lead to accumulated tracking error, whereas frame-wise OpenPose-based initialization contributes to stable model fitting over the sequence.

Table 2 summarizes the comparison of the initialization strategies. The trial-level mean RMSE across the six evaluated body points was lower with OpenPose-based initialization

in each frame than with OpenPose-based initialization only in the first analyzed frame in all five trials. Across the five trials, the mean and standard deviation of the trial-level mean RMSE were 0.083 ± 0.027 m with frame-wise initialization and 0.173 ± 0.040 m with initial-frame-only initialization.

As illustrated by the representative trajectory in Figure 8, using only the preceding-frame ICP result caused the estimated joint position to deviate progressively from the Vicon reference. These results indicate that the preceding-frame ICP result can support temporary continuation when OpenPose-based initialization is unavailable, but its repeated use as the sole initialization can lead to accumulated tracking error.

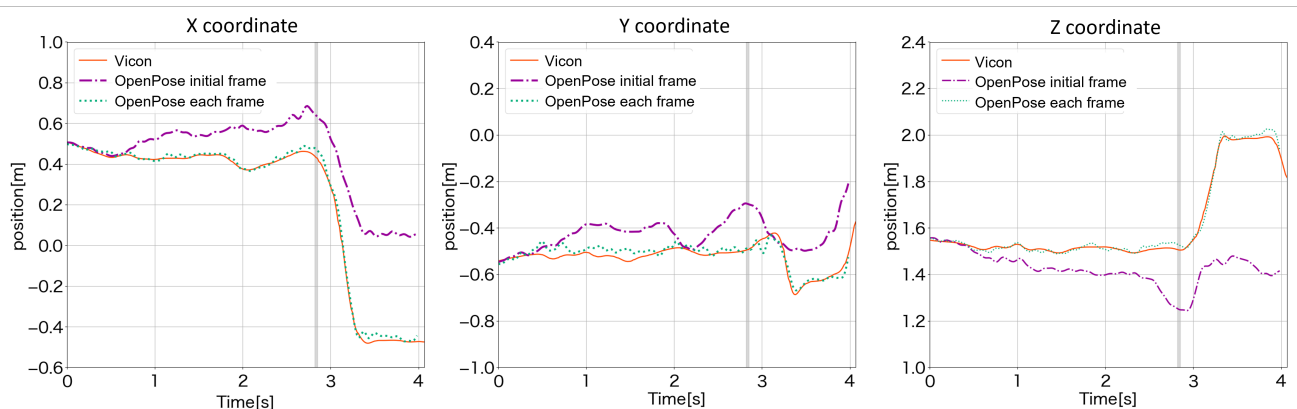


Figure 8. Coordinate comparison of the right elbow in a representative low-occlusion trial using Vicon, OpenPose-based initialization in each frame, and OpenPose-based initialization only in the first analyzed frame. The right elbow was selected as a representative joint point for examining the effect of the initialization strategy. The gray shaded regions indicate frames in which OpenPose keypoints were missing.

Table 2. Comparison of the initialization strategies using the five low-occlusion trials. The values represent the arithmetic mean of the three-dimensional RMSE values for the six evaluated body points in each trial.

Trial	OpenPose Each Frame [m]	OpenPose Initial Frame [m]
1	0.054	0.153
2	0.079	0.135
3	0.088	0.240
4	0.126	0.170
5	0.068	0.169
Mean $\pm$ SD	0.083 ± 0.027	0.173 ± 0.040

3.5. Effect of ICP-Based Correction Under Occlusion-Related Keypoint Loss

In addition to the multi-trial quantitative evaluation, this section examines a representative sequence that was intentionally selected because it included initial keypoint loss, in order to clarify the behavior and limitations of the proposed method under an unfavorable measurement condition. Figure 9 shows the representative climbing motion used to examine the effect of occlusion-related keypoint loss. In this motion, the left arm was temporarily hidden by the trunk, which likely caused missing or unreliable OpenPose keypoints for the target body segment. Figure 10 shows the left elbow coordinate trajectory during this motion. At the beginning of the motion, the OpenPose keypoints for the left arm were missing or unreliable, and therefore a reliable initial RGB-D reconstruction could not be obtained for ICP-based fitting. After the left arm became visible again, the proposed method updated the estimate using the observed point cloud.

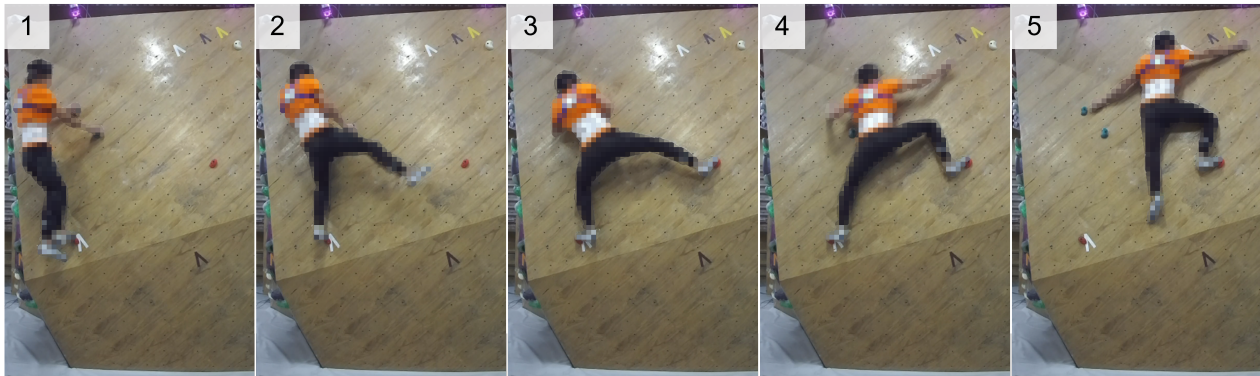


Figure 9. Climbing motion used to examine occlusion-related keypoint loss. Panels 1–5 show successive frames of the representative climbing motion. During the motion, the left arm was temporarily hidden by the trunk, which likely caused missing or unreliable OpenPose keypoints for the target body segment.

These results indicate that the proposed method can continue posture estimation when keypoint-based reconstruction becomes temporarily unreliable, for example, because of occlusion or limited visibility, provided that a reliable initial posture is available from the previous-frame fitting result. However, when the target body segment is completely hidden and sufficient point cloud data are not available for model fitting, the accuracy of the estimated joint position is limited. Therefore, both the availability of a reliable initial posture and the visibility of the target body segment in the point cloud are important factors affecting the accuracy of the proposed method.

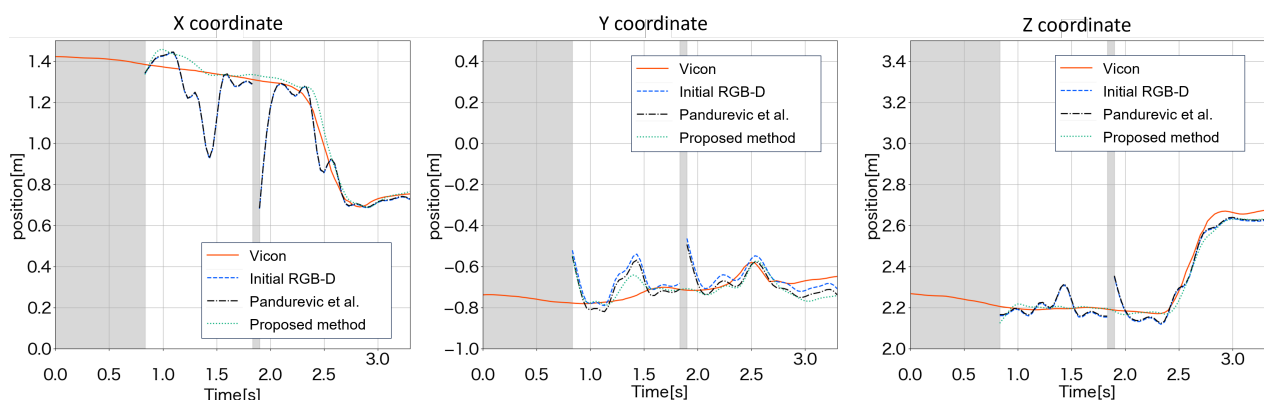


Figure 10. Left elbow coordinate trajectory during the climbing motion involving occlusion-related keypoint loss, obtained using Vicon, the initial RGB-D reconstruction, the method of Pandurevic et al. [16], and the proposed method. The gray shaded regions indicate frames in which the OpenPose keypoints for the left arm were missing or unreliable. At the beginning of the motion, the OpenPose keypoints for the left arm were missing or unreliable, and therefore a reliable initial RGB-D reconstruction could not be obtained for ICP-based fitting. After the left arm became visible again, the proposed method updated the estimate using the observed point cloud.

3.6. Effect of Human-Region Segmentation

To examine the contribution of human-region segmentation, an additional ablation analysis was performed on the five low-occlusion trials. These were the same five trials used for the initialization-strategy comparison in Section 3.4. In the ablation condition, the LangSAM-based human-region segmentation step was omitted, while the remaining processing procedure was unchanged.

Table 3 summarizes the trial-level mean RMSE across the six evaluated body points and the availability of OpenPose-based initialization. With human-region segmentation, the mean RMSE across the five trials was 0.083 ± 0.027 m, whereas without segmentation

it was 0.122 ± 0.070 m. The segmented condition showed a lower RMSE in four of the five trials.

Human-region segmentation also had a pronounced effect on the availability of OpenPose-based initialization. Across the five trials, an OpenPose-based initial posture was successfully obtained in 1005 of 1351 frames (74.4%) with segmentation, compared with 492 of 1351 frames (36.4%) without segmentation.

Human-region segmentation also substantially affected the computational cost of the subsequent ICP fitting. Without segmentation, ICP fitting of the 11 body-part models required 11 h 42 min 30 s for a 230-frame sequence, corresponding to 183.3 s per frame. For comparison, the ICP fitting time for the standard segmented pipeline reported in Section 2.6 was 13.221 s per frame. Because these timing measurements were obtained from different frame sequences, they should not be interpreted as a direct speed ratio. Nevertheless, the large difference indicates that restricting the target point cloud to the human region substantially reduces the computational burden of ICP.

Table 3. Ablation analysis of human-region segmentation using the five low-occlusion trials. (a) Trial-level mean RMSE across the six evaluated body points. (b) Availability of OpenPose-based initialization.

(a) Estimation accuracy		
Trial	With segmentation [m]	Without segmentation [m]
1	0.054	0.077
2	0.079	0.067
3	0.088	0.094
4	0.126	0.131
5	0.068	0.239
Mean $\pm$ SD	0.083 ± 0.027	0.122 ± 0.070
(b) OpenPose-based initialization availability		
Trial	With segmentation	Without segmentation
1	191/245 (78.0%)	154/245 (62.9%)
2	238/316 (75.3%)	215/316 (68.0%)
3	252/316 (79.7%)	57/316 (18.0%)
4	141/226 (62.4%)	50/226 (22.1%)
5	183/248 (73.8%)	16/248 (6.5%)
Overall	1005/1351 (74.4%)	492/1351 (36.4%)

Note: In part (a), bold values indicate the lower RMSE in each row. In part (b), the bold value indicates the higher overall OpenPose-based initialization availability.

4. Discussion

The proposed method estimates three-dimensional climbing posture using a single RGB-D camera, human region segmentation, OpenPose-based two-dimensional keypoint detection, depth projection, and ICP-based body-part model fitting. The main advantage of this approach is that it does not require markers to be attached to the climber and uses only a single RGB-D camera. This feature is useful for climbing motion measurement because marker attachment may interfere with natural movement, and the installation of multiple cameras is often difficult in climbing environments.

The present implementation used OpenPose for two-dimensional keypoint detection. The proposed method operates as a post-processing framework for RGB-D-based posture reconstruction and could potentially be combined with other pose-estimation models. However, its performance with more recent or task-specific pose estimators was not evaluated in this study. Therefore, the extent to which the observed improvement depends on the

accuracy and characteristics of the OpenPose-based initialization remains to be investigated. In the present framework, keypoint detection errors and missing keypoints are treated as sources of uncertainty in climbing measurements, and the reconstructed posture is refined using the geometric consistency between the segmented point cloud and the predefined body-part models.

SAM 3D Body, which was published after the implementation and experimental stages of the present study, provides a recent approach for recovering a full-body mesh and articulated pose from a single RGB image [13]. Because its monocular mesh-recovery formulation and output representation differ from the metric joint-position measurement considered in this study, a direct comparison of absolute three-dimensional joint coordinates is not straightforward. Future work may investigate whether its articulated-pose estimates, such as joint rotations or derived joint angles, can complement the proposed RGB-D-based framework or provide an alternative initialization for body-part model fitting.

Human region segmentation was introduced to reduce the influence of the climbing environment on OpenPose-based initialization and ICP fitting. The ablation analysis showed that removing segmentation reduced the availability of OpenPose-based initialization from 74.4% to 36.4% and increased the mean RMSE from 0.083 ± 0.027 m to 0.122 ± 0.070 m. The effect varied among trials. In the fifth trial, for example, OpenPose-based initialization was available in 183 of 248 frames (73.8%) with segmentation but in only 16 of 248 frames (6.5%) without segmentation, while the trial-level mean RMSE increased from 0.068 m to 0.239 m. These results indicate that human-region segmentation contributes to both reliable frame-wise initialization and estimation accuracy when the climbing environment strongly interferes with pose estimation. In addition, removing segmentation greatly increased the computational cost of ICP because the wall, holds, and other background surfaces remained in the target point cloud. Thus, human-region segmentation also substantially reduces the computational burden of ICP.

The initial three-dimensional posture reconstructed from OpenPose keypoints and depth values provides a simple way to obtain joint positions from RGB-D images. However, this initial reconstruction has inherent limitations. The depth value assigned to each keypoint represents the visible body surface rather than the anatomical joint center. Therefore, the reconstructed joint coordinates can deviate from the true joint positions, especially in the depth direction. In addition, OpenPose may fail to detect some keypoints or may incorrectly estimate keypoint positions when body parts are occluded, when limbs overlap, or when the climber takes postures that are uncommon in ordinary standing or walking motions.

To address these limitations, ICP-based body-part model fitting was introduced. By fitting predefined body-part models to the segmented human point cloud, the proposed method refines the initial posture obtained by depth projection. The fitted body-part models provide corrected joint coordinates based on the geometric relationship between the observed point cloud and the body-part models. This process is effective for reducing errors caused by assigning surface depth values directly to joint positions. The results also suggest that the use of the previous frame as the initial value can improve continuity when OpenPose fails to detect some keypoints in a frame.

The initialization-strategy comparison further showed that the preceding-frame ICP result is useful as a temporary fallback but is not sufficient as the sole initialization throughout a complete motion sequence. When OpenPose-based initialization was provided only in the first analyzed frame, tracking errors accumulated, and in the representative trial, a rapid posture change was not followed accurately. Thus, frame-wise OpenPose-based initialization remains important for preventing long-term drift, whereas the previous-frame ICP result supports continuity during temporary keypoint loss.

The multi-trial quantitative results further demonstrate the overall performance of the proposed method. In the low-occlusion condition, the proposed method reduced the mean RMSE across the six body points relative to the initial RGB-D reconstruction in all five analyzed trials, although the improvement over the method of Pandurevic et al. [16] was small on average and was not observed for every body point. In contrast, under the occlusion-prone condition, the proposed method showed the lowest mean RMSE for all six evaluated body points and reduced the trial-level mean RMSE relative to the initial RGB-D reconstruction in all eight trials.

These results suggest that ICP-based body-part model fitting is particularly useful when keypoint-based RGB-D reconstruction is degraded by limited visibility or occlusion. However, substantial inter-trial variation remained for some body points, particularly the right elbow in the low-occlusion condition. This variation was primarily attributable to a single trial, in which the right-elbow RMSE was 0.292 m, whereas the RMSEs in the other four trials ranged from 0.033 to 0.071 m. Although the low-occlusion condition produced less overall occlusion than the occlusion-prone condition, the visibility of individual body segments still varied with posture. Local occlusion or a reduction in the visible surface of an upper limb can provide fewer geometric constraints for ICP, making the fitting result more sensitive to the initial model placement and neighboring body surfaces. Thus, the effectiveness of ICP correction is not uniform across body parts and depends on both the quality of the initial posture and the amount of visible surface available for fitting.

The magnitude of the remaining error should also be considered when interpreting the practical applicability of the proposed method. The trial-level mean RMSE across the six evaluated body points was approximately 0.08 m under the low-occlusion condition and 0.12 m under the occlusion-prone condition. Errors of this magnitude limit the suitability of the present method for applications requiring precise localization of anatomical joint centers or evaluation of small differences in joint position. However, in sports-science applications, absolute joint-position accuracy is not always the only quantity of interest; temporal and relative changes in whole-body posture and movement trajectories can also provide useful information for characterizing movement. Therefore, the proposed method may be applicable to analyses of relatively large postural changes and movement patterns, particularly in situations where marker attachment or multi-camera motion capture is difficult. Further evaluation is required to determine the accuracy with which such relative changes can be quantified.

Several limitations of the proposed method remain. First, the number and diversity of participants were limited. All four participants were male and ranged from advanced to elite-level boulderers, and the low-occlusion condition included only two participants. Therefore, the present results should be interpreted as a feasibility evaluation rather than as evidence of population-level generalizability or state-of-the-art performance. Evaluation with additional participants with different body characteristics and skill levels, as well as different climbing styles and wall configurations, is required. Because multiple trials were obtained from the same participants, the trial-level standard deviations reported in Table 1 reflect both within-participant and between-participant variability and should not be interpreted as variability across independent participants. Second, the performance of ICP-based fitting depends on the quality of the segmented human point cloud. Because the same full-body point cloud was used as the target for all body-part models, inaccurate initialization may also cause a model to converge toward the surface of an adjacent body part. If the segmentation result includes background regions or misses parts of the body, the fitting result may become unstable. Third, the use of a single RGB-D camera limits the observable surface of the body. When a body segment is completely occluded by the trunk, another limb, or the wall, the corresponding point cloud cannot be obtained, and

the proposed method cannot accurately fit the body-part model. Fourth, the ICP algorithm requires an appropriate initial value. Although the proposed method uses the projected three-dimensional posture or the previous frame result as the initial value, large errors in the initial posture may still lead to incorrect fitting. In practical measurements, this situation can be mitigated by acquiring an initial posture in which all target body segments are visible before the start of the climbing motion.

Finally, the experiments were conducted using a single indoor climbing wall and under one environmental condition. Although preliminary observations motivated the selection of the ZED 2i instead of the Azure Kinect DK and Intel RealSense L515 in this environment, the robustness of the proposed method under different walls, background textures, illumination conditions, and outdoor or semi-outdoor environments was not quantitatively evaluated.

These limitations suggest several directions for future improvement. The robustness of the method could be improved by incorporating temporal constraints across consecutive frames, because climbing motion changes continuously over time. In addition, improving the body-part models and the segmentation process may increase fitting accuracy under complex postures. The use of multiple RGB-D cameras or additional environmental information may also reduce the influence of occlusion. Further quantitative evaluation with a larger number of participants, climbing motions, and body points will be necessary to clarify the conditions under which the proposed method can be reliably applied to sport climbing motion analysis.

5. Conclusions

This study proposed a markerless three-dimensional motion measurement method for sport climbing using a single RGB-D camera and ICP-based body-part model fitting. The method reconstructs an initial three-dimensional posture from segmented RGB-D images and OpenPose keypoints and then refines the posture by fitting predefined body-part models to the observed human point cloud.

The quantitative evaluation included five successfully completed low-occlusion trials from two participants and eight successfully completed occlusion-prone trials from four participants. The proposed method reduced the mean three-dimensional RMSE across six representative body points relative to the initial RGB-D reconstruction under both conditions. The trial-level mean RMSE of the proposed method was 0.083 ± 0.027 m under the low-occlusion condition and 0.120 ± 0.098 m under the occlusion-prone condition. The improvement was particularly evident under the occlusion-prone condition, in which the proposed method showed the lowest mean RMSE for all six evaluated body points.

The additional analyses showed that frame-wise OpenPose-based initialization is important for preventing accumulated tracking error, while the previous-frame ICP result can support temporary continuation when OpenPose-based initialization is unavailable. Human-region segmentation increased the availability of OpenPose-based initialization from 36.4% to 74.4% in the evaluated five trials and reduced the mean trial-level RMSE from 0.122 ± 0.070 m to 0.083 ± 0.027 m, while also substantially reducing the computational burden of subsequent ICP fitting.

The remaining estimation error limits the use of the present method for applications requiring highly accurate anatomical joint positions. However, the method may still be useful for characterizing relatively large changes in whole-body posture and movement trajectories in sport-science applications, particularly when marker attachment or multi-camera motion capture is difficult. Further evaluation is required to determine the accuracy of relative changes in posture and to assess the method across a wider range of participants, climbing movements, and measurement environments.

Author Contributions: Conceptualization, A.K.; methodology, A.K.; software, A.K. and W.M.; validation, A.K. and W.M.; formal analysis, A.K., W.M. and T.N.; investigation, A.K. and W.M.; resources, A.K. and R.K.; data curation, A.K. and W.M.; writing—original draft preparation, A.K.; writing—review and editing, A.K. and T.N.; visualization, A.K. and W.M.; supervision, A.K. and R.K.; project administration, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Number JP23K11351 and the Enhancement of HPSC Infrastructure through Technology Innovation Project of the Japan Sports Agency.

Institutional Review Board Statement: The study was approved by the Institutional Review Board of the Faculty of Information Science and Electrical Engineering, Kyushu University (approval number: 2024-15; date of approval: 23 July 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. The consent form for participation in the experiment was distributed to all participants and signed before the experiment.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical considerations but are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank Bare Hands Climbing Gym and its owner, Toru Iseki, for providing the experimental site and for their cooperation in conducting the experiments.

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

References

1. Saul, D.; Steinmetz, G.; Lehmann, W.; Schilling, A.F. Determinants for Success in Climbing: A Systematic Review. *J. Exerc. Sci. Fit.* **2019**, *17*, 91–100. [[CrossRef](#)] [[PubMed](#)]
2. Richter, J.; Beltrán, R.B.; Köstermeyer, G. Human Climbing and Bouldering Motion Analysis: A Survey on Sensors, Motion Capture, Analysis Algorithms, Recent Advances and Applications. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Valletta, Malta, 27–29 February 2020; pp. 751–758. [[CrossRef](#)]
3. Beltrán Beltrán, R.; Richter, J.; Köstermeyer, G.; Heinkel, U. Climbing Technique Evaluation by Means of Skeleton Video Stream Analysis. *Sensors* **2023**, *23*, 8216. [[CrossRef](#)] [[PubMed](#)]
4. Iguma, H.; Kawamura, A.; Kurazume, R. A New 3D Motion and Force Measurement System for Sport Climbing. In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 1002–1007. [[CrossRef](#)]
5. Mündermann, L.; Corazza, S.; Andriacchi, T.P. The Evolution of Methods for the Capture of Human Movement Leading to Markerless Motion Capture for Biomechanical Applications. *J. Neuroeng. Rehabil.* **2006**, *3*, 6. [[CrossRef](#)] [[PubMed](#)]
6. Colyer, S.L.; Evans, M.; Cosker, D.P.; Salo, A.I.T. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sport. Med. Open* **2018**, *4*, 24. [[CrossRef](#)] [[PubMed](#)]
7. Wade, L.; Needham, L.; McGuigan, P.; Bilzon, J. Applications and Limitations of Current Markerless Motion Capture Methods for Clinical Gait Biomechanics. *PeerJ* **2022**, *10*, e12995. [[CrossRef](#)] [[PubMed](#)]
8. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [[CrossRef](#)] [[PubMed](#)]
9. Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, A. Real-Time Human Pose Recognition in Parts from Single Depth Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1297–1304. [[CrossRef](#)]
10. Desmarais, Y.; Mottet, D.; Slangen, P.; Montesinos, P. A Review of 3D Human Pose Estimation Algorithms for Markerless Motion Capture. *Comput. Vis. Image Underst.* **2021**, *212*, 103275. [[CrossRef](#)]
11. Nakano, N.; Sakura, T.; Ueda, K.; Omura, L.; Kimura, A.; Iino, Y.; Fukashiro, S.; Yoshioka, S. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Front. Sport. Act. Living* **2020**, *2*, 50. [[CrossRef](#)] [[PubMed](#)]

12. Edriss, S.; Romagnoli, C.; Caprioli, L.; Bonaiuto, V.; Padua, E.; Annino, G. Commercial Vision Sensors and AI-Based Pose Estimation Frameworks for Markerless Motion Analysis in Sports and Exercises: A Mini Review. *Front. Physiol.* **2025**, *16*, 1649330. [[CrossRef](#)] [[PubMed](#)]
13. Yang, X.; Kukreja, D.; Pinkus, D.; Fan, T.; Park, J.; Shin, S.; Cao, J.; Liu, J.W.; Ugrinovic, N.; Sagar, A.; et al. SAM 3D Body: Robust Full-Body Human Mesh Recovery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, CO, USA, 3–7 June 2026; pp. 7209–7219.
14. Beltrán B., R.; Richter, J.; Heinkel, U. Automated Human Movement Segmentation by Means of Human Pose Estimation in RGB-D Videos for Climbing Motion Analysis. In Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Virtual, 6–8 February 2022; Volume 5, pp. 366–373. [[CrossRef](#)]
15. Yan, M.; Wang, X.; Dai, Y.; Shen, S.; Wen, C.; Xu, L.; Ma, Y.; Wang, C. CIMI4D: A Large Multimodal Climbing Motion Dataset Under Human-Scene Interactions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 12977–12988.
16. Pandurevic, D.; Sutor, A.; Hochradel, K. Introduction of a Measurement System for Quantitative Analysis of Force and Technique in Competitive Sport Climbing. In Proceedings of the 8th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2020), Budapest, Hungary, 5–6 November 2020; pp. 173–177. [[CrossRef](#)]
17. Medeiros, L. Language Segment-Anything. 2023. Available online: <https://github.com/luca-medeiros/lang-segment-anything> (accessed on 27 July 2026).
18. Liu, S.; Zeng, Z.; Ren, T.; Li, F.; Zhang, H.; Yang, J.; Jiang, Q.; Li, C.; Yang, J.; Su, H.; et al. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. *arXiv* **2023**, arXiv:2303.05499.
19. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 4015–4026.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.