aerospace

MDPI

Article

A Multi-View Vision System for Astronaut Postural Reconstruction with Self-Calibration

Shuwei Gan [1], Xiaohu Zhang [1], Sheng Zhuge [1], Chenghao Ning [1], Lijun Zhong [1] and You Li [2,*]

[1] School of Aeronautics and Astronautics, Sun Yat-sen University, Shenzhen 518107, China
[2] National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing 100094, China
* Correspondence: liyou@nudt.edu.cn

Abstract: Space exploration missions involve significant participation from astronauts. Therefore, it is of great practical importance to assess the astronauts' performance via various parameters in the cramped and weightless space station. In this paper, we proposed a calibration-free multi-view vision system for astronaut performance capture, including two modules: (1) an alternating iterative optimization of the camera pose and human pose is implemented to calibrate the extrinsic camera parameters with detected 2D keypoints. (2) Scale factors are restricted by the limb length to recover the real-world scale and the shape parameters are refined for subsequent postural reconstruction. These two modules can provide effective and efficient motion capture in a weightless space station. Extensive experiments using public datasets and the ground verification test data demonstrated the accuracy of the estimated camera pose and the effectiveness of the reconstructed human pose.

Keywords: multi-view system; astronaut performance capture; extrinsic camera calibration; human pose estimation

Citation: Gan, S.; Zhang, X.; Zhuge, S.; Ning, C.; Zhong, L.; Li, Y. A Multi-View Vision System for Astronaut Postural Reconstruction with Self-Calibration. *Aerospace* **2023**, *10*, 298. https://doi.org/10.3390/aerospace10030298

Academic Editors: Paolo Tortora and Pierre Rochus

Received: 29 November 2022
Revised: 15 February 2023
Accepted: 14 March 2023
Published: 17 March 2023

1. Introduction

As a scientific research laboratory in a microgravity environment, low-Earth-orbit space stations conduct a wide range of experimental tasks in many domains, such as biology, physics, and astronomy. To ensure the physical and mental well-being of the astronauts and the efficient completion of tasks, astronauts must carry out a number of complex scientific experiments while in orbit, monitor and analyze their workload, arrange in-orbit tasks reasonably, and increase their work efficiency.

Attitude and pose are the external manifestations of astronauts' human performance when performing the in-orbit missions. By measuring and analyzing the postural data of astronauts in orbit, it is possible to discover their long-term postural characteristics in weightlessness, which is of great value for improving the design of orbital modules, regulating experimental tasks in accordance with workload and enhancing human operative capabilities during prolonged space voyages [1–3]. Therefore, it is necessary to precisely estimate the body pose of astronauts when performing different tasks to quantify their postures and movements.

The capture of human motion data relies on different sensors, which can be divided into two categories according to working mode: contact methods and non-contact methods [4]. Contact human motion measurement is performed by installing inertial sensors in key parts of the human body (e.g., limbs, torso) and capturing information, such as the acceleration and angular velocity of the wearing part in real-time for postural measurement. These inertial measurement unit (IMU)-based motion capture systems (MOCAP), such as Xsens, can provide robust motion capture with occlusion, which can only be alleviated by increasing camera numbers for optical-based MOCAP. IMU-based MOCAP is widely used in biomedical applications for gait analysis of patients [5] and astronaut postural analysis in extravehicular operations [6]. Non-contact methods mainly rely on optical

sensors, including infrared cameras, color cameras, and depth cameras. Optical motion capture generally uses a multi-camera configuration, and can be divided into two categories according to whether markers are required. Marker-based optical motion capture, such as Vicon, OptiTrack, etc., can provide the high-precision position of optical markers placed on the surface of the human body to measure human motion, while markerless optical motion capture [7] mainly uses computer vision algorithms for human pose estimation, which is less accurate than the marker-based human motion capture systems. However, with the continuous development of computer vision, its accuracy continues to improve and the advantages of convenience gradually emerge. The space station requires more convenience due to its small space and limited resources; therefore, designing a markerless optical human motion capture system for in-orbit is the best way to achieve astronaut postural measurements.

With the development of deep learning in computer vision, multiple convolutional neural network (CNN) models have been proposed to solve the human pose estimation problem [8,9]. At present, 2D human pose estimation has been explored to provide effective and efficient human body joint detection [10–12]. However, 3D human pose estimation is still problematic due to the depth ambiguity and occlusion problem in single-view camera setups [13–15]. Thus, multi-view 3D human pose estimation methods are the most robust solutions to infer the 3D location of human joints at present. However, the multi-camera system relies on extrinsic calibration before deployment, which is cumbersome in weightless space stations. To reduce the complexity of calibration, several calibration-free human pose recovery methods have been proposed. As the human body moving in the scene is captured by these multi-view systems, the most intuitive strategy is to take advantage of the human body that is presented to provide common viewpoints [16–18]. The workflow usually contains three parts. First, an initial calibration of the multi-camera systems is obtained with traditional fundamental matrix estimation. Then, bundle adjustment is implemented to optimize the rotation and translation among each cameras with re-projection error and other specifically designed priors. Finally, the human body is reconstructed by optimizing an appropriate cost function. However, the initial calibration is usually far from the ground truth and the bundle adjustment struggles to deal with the deviated rotation matrix and translation vectors together with the erroneous 3D keypoints because the camera pose and human pose are highly coupled and influence each other.

In this paper, we propose a multi-view system to recover the astronaut's postural performance without cumbersome extrinsic calibration. More precisely, the first part of our pipeline consists of multi-view camera pose estimation. The self-calibration technique is based on the insight that the inaccuracy of the camera pose and human pose are coupled and will influence each other. We proposed to alternatively estimate the camera pose and recover the human pose with confidence-weighted iterative perspective-n-point (PnP) and triangulation. First, the fundamental matrix estimation between each camera pair is estimated and the rotation matrix and translation vector are recovered by matrix decomposition. Then, the re-projection error is calculated to assess the estimation results for the selection of the first two base cameras and the human skeletons are triangulated to provide 3D–2D correspondences for the other cameras. Third, the remaining cameras are incorporated into the multi-camera systems with PnP-based camera pose estimation and the human skeletons are alternatively refined with the gradually incorporated cameras. In this way, the triangulation error caused by the erroneous camera's extrinsic parameters and partially occluded human skeletons will alternatively decrease and the iteration is terminated until the difference between consecutive triangulated points is lower than the pre-defined tolerance. Furthermore, during the procedure of incorporating an extra available camera, the scale of the translate vectors among each cameras will be consistent with the first chosen camera pair.

After the multi-camera system is calibrated with the presented human body, the posture of the astronaut can be recovered with a pre-scanned mesh model. The length of the limbs is regarded as a reference when calculating the scale, as the microgravity can barely

influence the bone length of the limbs compared to the trunk [19]. With the scale-modified camera's extrinsic parameters, the shape parameters of the skinned multi-person linear model (SMPL) are refined to describe the spinal lengthening and anthropometric changes in microgravity. Finally, the posture of the astronauts are reconstructed with SMPL parameters fitting to the detected 2D keypoints. The schematic workflow is shown in Figure 1.

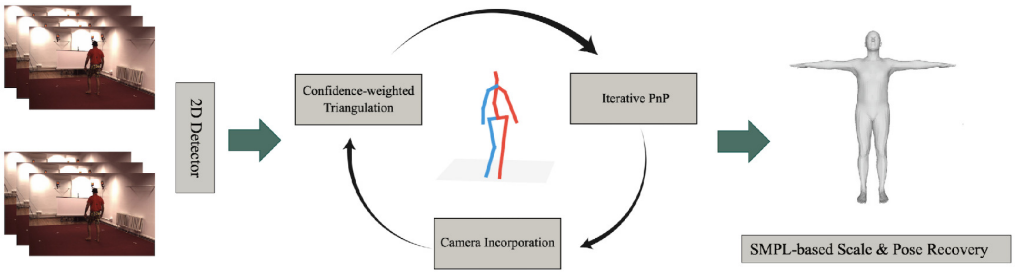


Figure 1. The schematic workflow of the proposed systems.

To summarize, the novelties of this work are twofold:

- First, an alternating iterative optimization of the camera pose parameter estimation and human pose parameters is proposed to provide a convenient and accurate estimation of the camera's extrinsic parameters.
- Second, a shape optimization is implemented based on the pre-scanned astronaut body model to refine the shape parameters for long-term space exploration missions and the astronaut's postural performance is reconstructed with non-linear optimization.

The rest of the paper is organized as follows. Related works in astronaut performance analysis and motion capture and calibration-free human body reconstruction are discussed in Section 2. The principles and methods are described in Section 3. The experiment details and an extensive evaluation of the proposed method for further applications in space stations are revealed in Section 4. Finally, the advantages, disadvantages and future works are discussed in Section 5.

2. Related Works

In this section, we review related works on astronaut performance capture, calibration-free systems and 3D human shape reconstruction.

2.1. Astronaut Performance Capture

Astronaut performance capture is of great importance to the research on manned space exploration to investigate the influence of weightlessness on the human body. To capture the astronaut's performance when collecting and analyzing human body characteristic in weightlessness, different kinds of sensors have been developed and tested. The marker-based motion capture system ELITE-S2 [20,21] was developed and brought into the International Space Station in August 2007. The goal was to research the motor control strategies of astronauts in a weightless environment. The system comprised eight TVC boxes containing a camera and processor, by which the astronauts' motion could be measured using four lasers to illuminate up to 100 markers placed on the body surface with an accuracy of less than a millimeter. This is quite similar to the optical motion capture system, while passive markers and retro-reflective markers are both attached to the body and captured by the camera. The system is complicated regarding calibration and marker preparation for astronauts performing in-orbit experiments, despite its high precision. NASA applied a compact stereo-vision-based motion capture system—ESPRIT [22]—to monitor microgravity exercises using marker detection and 3D kinematic posture recovery.

Despite the marker-based system, markerless computer vision methods are also proposed to monitor the kinematic states of astronauts. In the Human Research Program [23], NASA has developed a machine-learning based method to automatically extract the astronaut's posture from video data. Considering the limited computational resources in a

space station, ROpenPose [2] was proposed to solve the posture detection problem with optimized neural networks. The markerless-based vision system is more friendly for astronauts due to its senselessness and unaffectedness during postural monitoring, while the accuracy of the reconstructed posture still needs to be continuously improved. At present, the multi-view motion capture system can provide the most reliable and accurate 3D human pose [24].

2.2. Calibration-Free Multi-View System

Practically, the camera intrinsic parameters are considered invariant and can be calibrated using the classic chessboard method. Therefore, extrinsic calibration is required for each camera position and the orientation is unknown with each installation. The fundamental of multi-camera extrinsic calibration is obtaining enough point correspondences for several camera perspectives, with or without specially designed calibration tools. Calibration tools refer to precisely manufactured calibrators, such as one-dimensional objects [25], chessboard [26], and spheres [27], which are required to move around the target area to ensure there are enough points correspondences for high-precision calibration. Thus, this intensive manual operation makes the calibration cumbersome, especially when the multi-view vision system is orientation-adjusted, which makes it unfeasible in the weightless space station.

Alternatively, calibration with geometric methods using the structure from motion (SfM) pipelines requires no special calibration tools. These methods consider extrinsic calibration as a camera pose estimation problem and solve the problem with a two-stage framework by 2D–2D matching points' extraction and successive camera pose estimation. Keypoints, such as Harris and FAST [28], are commonly detected from images of various viewpoints. Then, the feature descriptors of these keypoints are exploited to match points in different views. The most-used are SIFT [29] and ORB [30]. Afterwards, the fundamental matrix between each camera pair is estimated with different methods, such as the RANSAC-based N-point algorithm [31,32] or convex optimization [33]. Successively, the camera rotation and translation are obtained by decomposing the essential matrix. To further improve the 3D positions of all cameras, they utilize bundle adjustment [34] as a last step. Despite the convenience of camera pose estimation, this pipeline is severely limited, as a large amount of camera power is required to provide a sufficient common field of view, which is usually not an issue for SfM reconstruction with a moving camera. In addition, the estimated camera translation is up-to-scale between each camera pair and requires a constant-sized object to calculate the scale.

To solve these problems, researchers have taken advantage of the human body to provide 2D–2D point correspondence, which is meaningful and available to those multi-view systems aiming for human body reconstruction. Reference [17] proposed detecting the bounding box of each person in the wide-baseline multi-camera and using a CNN with person re-identification ability to match these detected person centers. These 2D–2D correspondence are then transmitted to the typical process pipeline with bundle adjustment. It is noticeable that the center of the bounding box in different views cannot be trusted to provide accurate 2D–2D correspondences, as the human body is complicated and can form different poses. With utilization of off-the-shelf 2D human pose estimators, [16] calibrated and synchronized the multi-view system with OpenPose and proposed constructing an object function with relaxed reprojection errors to avoiding optimization in noisy observations. Reference [18] proposed taking advantage of a moving person in a different way by first lifting the 2D human pose to a 3D human pose with VideoPose3D [35], and then solving the camera extrinsic calibration as an absolute orientation problem. In these paradigms, the multi-view reconstructed human poses are regarded as a ground truth to fine-tune the CNN module of 2D or 3D human pose estimation.

3. Proposed Method

The details of our method are presented in this section. In this paper, we used an off-the-shelf 2D pose detector as the 2D–2D correspondence extractor and treated the astronaut performance capture as a iteratively refined problem. The key idea is to optimize the camera pose and human pose alternatively by incorporating multi-view cameras into the postural assessment system.

First, the fundamental matrix estimation between each camera pair was estimated, and the rotation matrix and translation vector were recovered by matrix decomposition. Then, the re-projection error was calculated to assess the estimation results for the selection of the first two base cameras. Third, the other cameras were gradually incorporated into the camera systems with PnP-based camera pose estimation until all the camera poses were estimated and could be used for the reconstruction of 3D human poses. Furthermore, while incorporating the extra available camera, the detected human body keypoints in the extra camera gradually improved the intersected 3D human pose. Thus, the estimated camera pose can be iteratively refined during the 2D–3D PnP calculation. In this way, the scale ambiguity problem of each camera can be reduced to be related to the scale of the first chosen camera pair. The schematic workflow is shown in Figure 2.

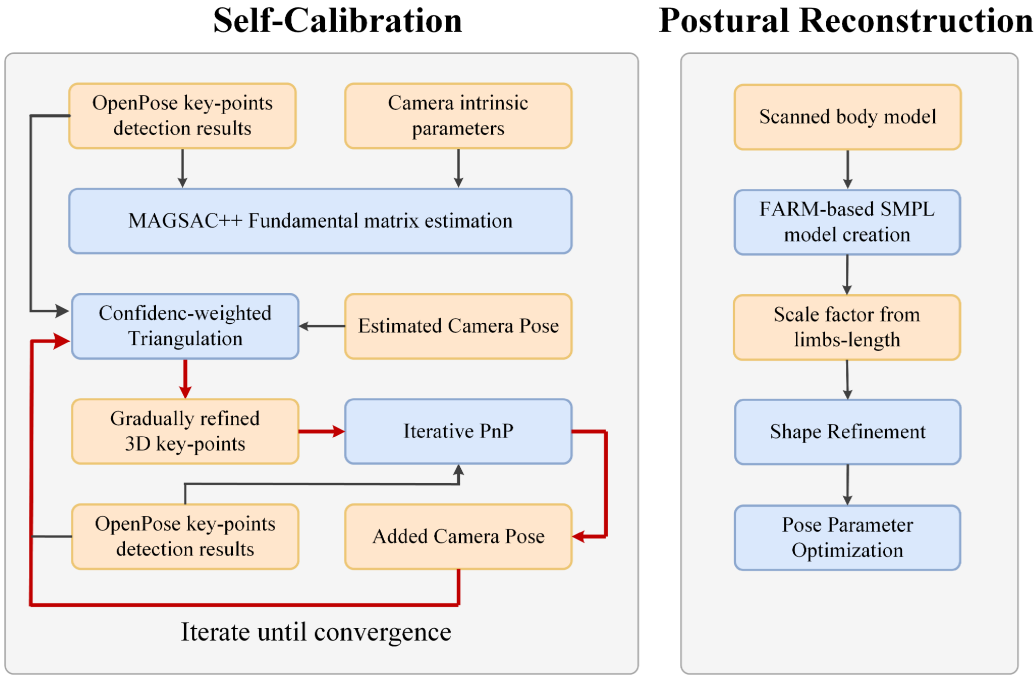


Figure 2. The workflow of our proposed calibration-free postural reconstruction. The self-calibration module iteratively estimates the multi-camera extrinsic parameters from the detected 2D human keypoints. The iteration flows in the red arrow direction to alternatively optimize the 3D keypoints and the camera pose. The postural reconstruction module works after the self-calibration and estimates the scale factor from the pre-scanned human models. The shape parameters are then refined and posture is reconstructed.

3.1. MAGSAC-Based Fundamental Matrix Estimation

Given a set of keypoint detection results with correspondences in homogeneous coordinates $\{\mathbf{x}_i \Longleftrightarrow \mathbf{x}_i'\}, i = 1, 2, \ldots, n, n \geq 8$, between two extrinsic uncalibrated cameras, the objective is to estimate a 3×3 fundamental matrix F satisfying the epipolar constraints:

$$\mathbf{x}_i F \mathbf{x}_i' = 0, i = 1, 2, \ldots, n. \tag{1}$$

The standard eight-point method can be applied to estimate the fundamental matrix. However, as the human movements in the target space can provide more keypoint cor-

respondence, at $N \times F$, N is used as the keypoint number of the skeleton definition and F as the frames captured during self-calibration. To select the best correspondent pairs that can minimize the error according to the epipolar constraints Equation (1), the most well-known RANSAC-like fitting model method can iteratively explore the space of model parameters by random sampling and estimate the most reliable model by maximizing the inliers. In the recently proposed MAGSAC++ [36], the model quality calculation is formulated as a marginalization over a range of noise scales. The inlier residuals are assumed to have χ^2 distribution. This allows for MAGSAC++ to be significantly less sensitive to the inlier–outlier threshold than other robust estimators. In this paper, we adopted the recently proposed MAGSAC++ as the fundamental matrix estimator with numerous keypoint correspondences.

3.2. Confidence-Weighted Camera Pose Refinement

In this paper, cameras were iteratively incorporated into the multi-camera system by alternatively estimating the camera pose and refining the 3D human pose. The proposed iteration can alternatively improve the 3D joints triangulation and camera pose estimation, which are coupled to affect the final postural measurement. First, the 3D joints were triangulated with the confidence output of the 2D pose detector, and the re-projection error was combined with its corresponding confidence to provide the weights for subsequent 2D–3D camera pose estimations. Second, the remaining cameras were incorporated into the multi-camera system with the confidence-weighted perspective-n-points methods from OPnP [37].

Given the detected image points of each joint in multi-view cameras and the camera parameters, the triangulation problem aims to recover the best estimation of the 3D point $\mathbf{p}$. With C cameras, the midpoint method [38] minimizes the following cost function:

$$E(\mathbf{p}) = \sum_{i=1}^{C} \|(\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T)(\mathbf{p} - \mathbf{o}_i)\|^2 \tag{2}$$

where $\mathbf{o}_i$ is the optical center of the *i-th* camera, $\mathbf{b}_i$ is the *i-th* unit vector pointing from the optical center to the image points and $\mathbf{p}$ is the point to be triangulated. To reduce the influence of the mis-detected keypoints during triangulation, we multiple the confidence of the detected keypoint to make the 3D point closer to the more reliable image point direction.

$$E(\mathbf{p}) = \sum_{i=1}^{C} \omega_i \|\mathbf{B}_i(\mathbf{p} - \mathbf{o}_i)\|^2 \tag{3}$$

where ω_i is the confidence of the keypoints detected in the *i-th* camera images and C is the number of cameras.

$$\left(\sum_{i=1}^{C} \omega_i \mathbf{B}_i\right)\mathbf{p} = \left(\sum_{i=1}^{C} \omega_i \mathbf{B}_i \mathbf{o}_i\right) \tag{4}$$

given the 3D point $\mathbf{p}$ triangulated with the previous incorporated cameras' parameters. The multi-view camera pose can be re-estimated with the current 3D points using the PnP method. The cost function used to estimate the rotation matrix and translation vectors is the sum of the confidence-weighted squared measurement errors, as follows:

$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} \omega_{i,j} \|\mathbf{u}_i - \frac{1}{z_i}(\mathbf{R}_i \mathbf{p}_i + \mathbf{t}_i)\|^2 \tag{5}$$

where ω_i is the confidence of the detected 2D keypoints and N is the number of points detected from the camera that is to be incorporated. We initialized the camera pose parameters with direct linear transform (DLT) and iteratively minimized the re-projection error with the Levenberg–Mardquart algorithm.

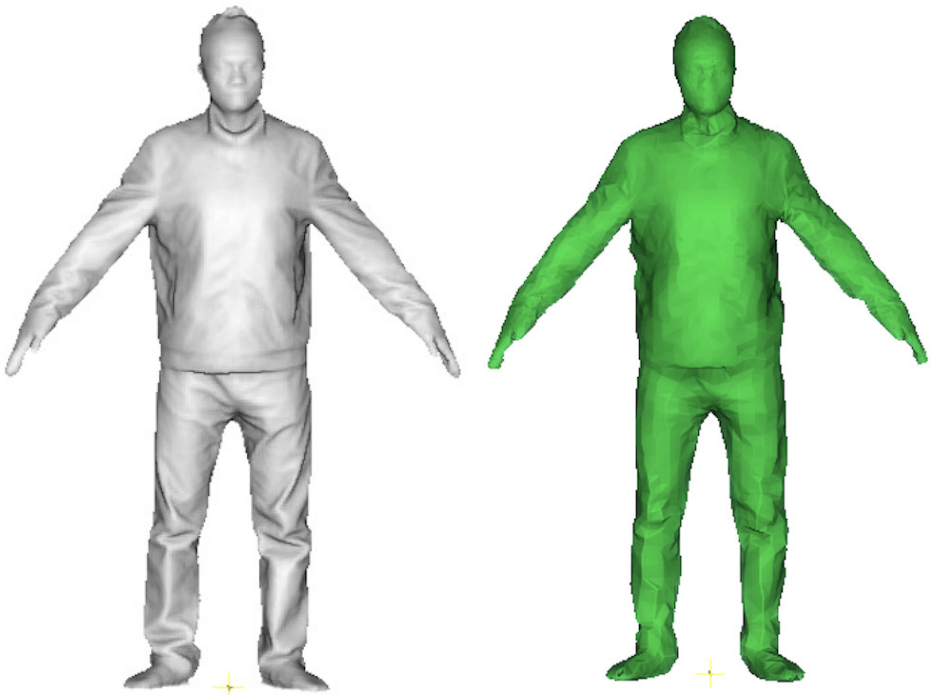
3.3. SMPL-Based Postural Reconstruction

The SMPL model [39] is a skinned vertex-based model, which parametrizes a triangulated mesh using pose and shape parameters. The shape parameters β are coefficients of a low-dimensional shape space, learned from a training set of thousands of registered 3D human body scans. The pose parameters θ represent the joint angle in an axis-angle representation of the relative rotation between body parts. The posed body model $\mathcal{M}(\beta, \theta)$ is formulated as below, given the shape and pose parameters,

$$\mathcal{M}(\beta, \theta) = W(T_P(\beta, \theta), J(\beta), \theta, \Omega) \tag{6}$$

3.3.1. Personalized SMPL Model

The SMPL model provides a convenient way to edit the human mesh model with shape parameters controlling the anthropometric parameters of different person. However, the astronauts are usually scanned before missions and the scanned model can be used to personalize the SMPL model with fixed shape parameters for each person as shown in Figure 3. In this paper, we used the FARM algorithm to reconstruct the personalized SMPL model from a scanned unstructured human point cloud. Afterwards, the limb lengths of the model were calculated and regarded as a reference to scale the estimated translation vectors.



(**a**) Scanned human body model. (**b**) Reconstructed SMPL model.

Figure 3. Personalized SMPL model for 3D posture reconstruction.

3.3.2. Shape Parameters Refinement

We observed that the spin lengths gradually increase during long-term space missions by around 5–10%, while the lengths of arms and legs were almost invariant due to the structural property of the bones. Thus, we used the personalized SMPL model to scale the triangulated human body keypoints with the bone length of the arms and legs. The shape parameters of the personalized SMPL model were refined with the re-scaled 3D keypoints.

The length of the bones was first calculated from the triangulated keypoints with the connecting relationship, and the cost function E_{bone} was minimized with multi-frame data to refine the shape parameters β

$$E_\beta = \lambda_{bone} E_{bone}(\beta) + \lambda_{prior} E_{prior}(\beta) \tag{7}$$

$$E_{bone}(\beta) = \sum_{i \in \mathcal{B}} \| f(J_i \cdot \mathcal{M}(\beta), \mathcal{C}_i) - B_i \|^2 \tag{8}$$

$$E_{prior}(\beta) = -log(\mathcal{N}(\beta; 0, \Sigma_\beta)) \tag{9}$$

where J_i is denoted as the rows of the joint regressor, used to calculate the joint position from the structured vertex of SMPL model, C_i as the connection relationship of the *i-th* bone for calculation of the bone length from joint position, and B_i is the *i-th* bone length from triangulated 3D keypoints.

3.3.3. Postural Reconstruction

The 10 shape parameters were fixed after shape parameters' refinement, and the postures of the performing human were reconstructed with the recovery of the pose parameters of the SMPL model. In the final step, the energy function was constructed to recover the pose parameters as follows:

$$E_{jtr}(\theta) = \lambda_{j2d} \sum_{v=1}^{C} \| \Pi(\mathcal{M}(\beta, \theta), T_v) - J_v^{2d} \|^2 + \lambda_{j3d} \| \mathcal{M}(\beta, \theta) - J^{3d} \|^2 + \lambda_{prior} R(\theta) \tag{10}$$

The first two energy terms are the data terms that fit the SMPL parameters to the detected 2D joints' location and the triangulated 3D joints. $\Pi(p, T)$ is the projection of the 3D point p with the projection matrix T. To prevent the parameters from falling into wired postures during the non-linear optimization process, the regularization term $R(\theta)$ was applied to force the pose parameters to satisfy the normal distribution of the pose dictionary, as follows:

$$E_{prior}(\theta) = \min_i(-log(g_i \mathcal{N}(\theta; \mu_{\theta,i}, \Sigma_{\theta,i}))) \tag{11}$$

4. Experimental Evaluation

We carried out numerous experiments on synthetic data and real scene data from open datasets and ground verification test data to confirm the calibration-free postural monitoring system. To assess the precision and accuracy of camera pose estimate and human body reconstruction, the human3.6m datasets were used for quantitative evaluation with the proposed method, focusing on three aspects:

1. Camera pose differences with available ground truth.
2. 2D reprojection error of the reconstructed joints.
3. 3D reconstruction error of the reconstructed joints.

We also qualitatively compared the calibration result of the ground verification test data with Zhang's Chessboard calibration method and evaluated the postural reconstruction results with reprojected human body meshes for a visual comparison.

4.1. Extrinsic Parameter Calibration

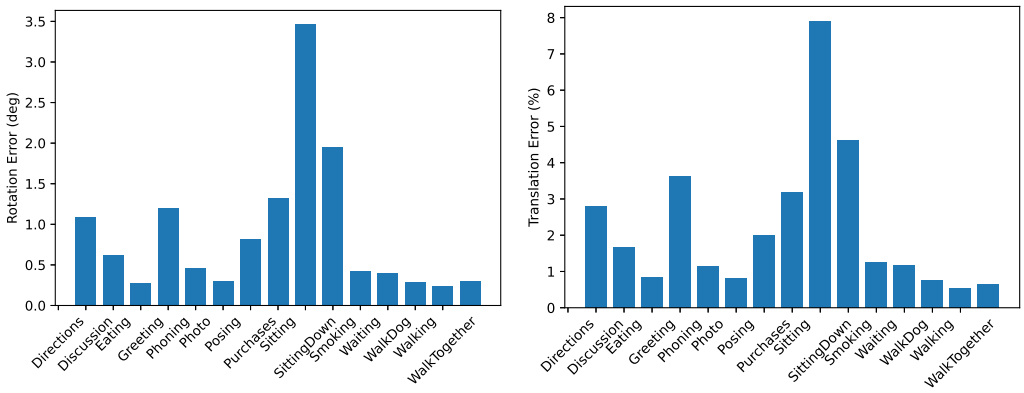
In this section, we evaluated the self-calibration accuracy with the detected human keypoints.

4.1.1. Human 3.6M Evaluation

The Human 3.6M dataset is a dataset for human pose estimation with annotated 3D human poses. People performing 15 activities were captured with four RGB cameras and the ground truth human keypoints data were obtained by the marker-based motion capture system Vicon. In this experiment, the calibration result with a T-shaped wand in the capture space was treated as the ground truth, and the images with different activities were input for the self-calibration and postural reconstruction.

We evaluated our self-calibration on the 15 activities of S9 subjects, from which 15 sequences of 400 frames were sent into the self-calibration module, and each sequence

was processed 100 times. To measure the errors in the poses estimated from the ground truth, the rotation difference with Riemannian distance [40] and the translation difference with root mean squared error (RMSE) in meters were calculated for each calibration result. The mean and standard deviations of the rotation error and translation error are shown in Figure 4. The calibration accuracy is clearly highly effected by the activity and the main influencing factors are twofold. First, the human body moving in each scenario could cover different space volumes with different types of actions. Second, the human keypoints' detection accuracy is related to the postural complexity.



(**a**) Rotation error in degrees.

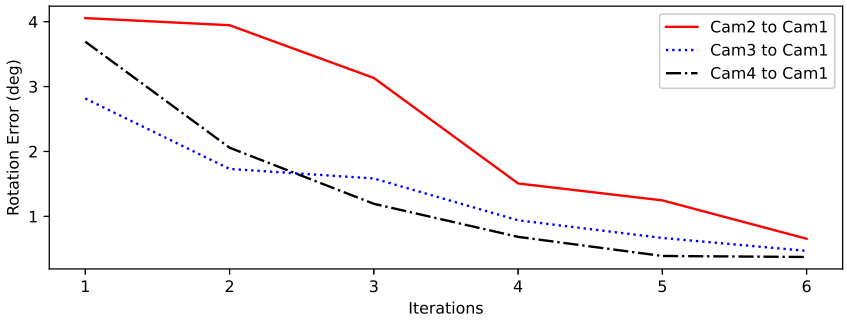
(**b**) Translation error in percentage.

Figure 4. The rotation error and translation error of the 15 activities.

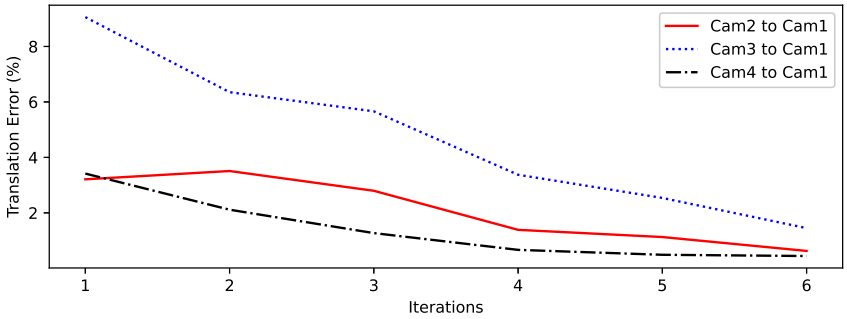
Conventionally, the detected keypoints are only used as an initial calibration and a subsequent optimization module with bundle adjustment, and different priors are implemented to obtain a better solution. To validate the iterative refinement of the camera pose estimation with the gradually improved 3D triangulation of the body joints, derived from the additional keypoint information. The calibration error after each round of alternative optimization is shown in Figure 5.

4.1.2. Ground Verification Test Data

To examine the actions required for in-orbit camera self-calibration, we also conducted a ground verification test experiment with limited activities by constraining the foot on imaginary foot restrictors. Our ground verification experiments were conducted with the same camera configurations compared to the in-orbit environment, with four cameras settled in a 3×2 m scene. The ground truth of extrinsic parameters was obtained with Zhang's Chessboard stereo camera calibration methods. We conducted three experiments with different actions and persons, as shown in the figure. In the experiments, three human subjects are captured with four cameras located at the corner of the 3×2 m scene. Subjects conducting actions such as walking, squatting, and sitting are captured with four synchronized cameras. The resolution of the camera is 640×480 and the frequency of the recording is 30 frames per second. During self-calibration and reconstruction, we use the pinhole camera model to project the 3D point into image coordinates. In the three ground verification test, the first 400 frames of the images are chosen to self-calibrate the camera extrinsic parameter for equal comparison with the Human 3.6M test data. The chessboard calibration was conducted with 30 chessboards, placed in different positions and orientations, occupying the common viewpoints of each camera pair. The re-projection error of the chessboard corners was 0.06–0.08 px for all camera pairs. The images captured for chessboard calibration are shown in Figure 6, and the comparison results are shown in Table 1. The first two columns show the rotation error and translation error of the proposed iterative refining camera pose, and the re-projection error of the chessboard corners with our calibration results are shown in the third column.



(**a**) Rotation error curve with iteration.



(**b**) Translation error curve with iterations.

Figure 5. The translation error and rotation error curves with iterations.

Figure 6. The images samples captured for chessboard calibration. More than 50 pairs of images are captured to guarantee the accuracy of the calibration results.

Table 1. Self-calibration results compared with the chessboard calibration. The re-projection error is computed by projecting the 3D position of the chessboard corners with the proposed calibration results.

Actions	Rotation Error (deg)	Translation Error (%)	Re-Projection Error (px)
No. 1	0.8234	1.2453	1.34
No. 2	0.7422	0.8264	0.98
No. 3	0.9231	1.534	1.65

4.2. Postural Reconstruction

In this section, we carried out SMPL-based shape refinement and postural reconstruction and evaluated the reconstruction accuracy by focusing on two aspects, namely the re-projection error of the recovered human joints and the 3D joint error when the ground truth joint positions are available. First, the quantitative validation on the Human 3.6M dataset was analyzed with the provided ground-truth 3D keypoints. We reported the results for subjects S9. We used all four views and the ground-truth camera parameters. Table 2 compares our approach with other state-of-the-art multi-view methods. The methods proposed by Trumble et al. [41] and Pavlakos et al. [42] only optimize the joint location, and the rest consider the shape and pose combined. As most of the multi-view methods regard the extrinsic camera parameters as known parameters, we compared the postural reconstruction process with the ground-truth camera extrinsic parameters as Our^1 and the estimated camera pose with proposed self-calibration procedure as Our^2. The MPJPE of our method was 41.52 mm and 42.75, respectively, which demonstrates that our approach is effective in recovering the camera pose and reconstructing the shape and pose of moving persons.

Table 2. Quantitative comparison on Human 3.6M (subject 9). "Shape" indicates if the method estimates the shape parameters. "PA" indicates if Procrustes analysis was applied before computing the MPJPE (mm). Calibration indicates if the ground-truth camera's extrinsic parameters were used to integrate multi-view data. Ours^1 is the postural reconstruction result with ground-truth extrinsic parameters. Ours^2 shows the results with proposed self-calibration parameters.

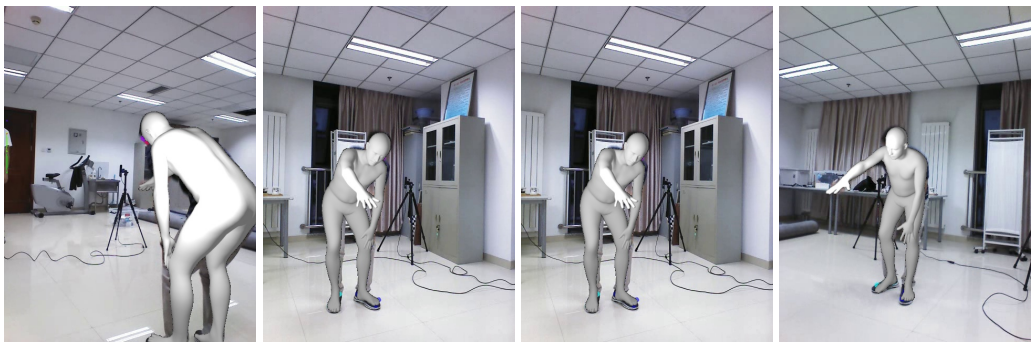
Methods	Shape	PA	Calibration	MPJPE
Trumble et al. [41]	No	No	Yes	62.50
Pavlakos et al. [42]	No	Yes	Yes	56.89
Huang et al. [43]	Yes	Yes	Yes	47.09
Ours^1	Yes	Yes	Yes	41.52
Ours^2	Yes	Yes	No	42.75

In order to quantify the time consumption of the proposed method, we analyze each module of the workflow. The experiment is conducted on the computer with an i9-9900K processor and Nvidia 2080ti. The 2D pose detection using OpenPose costs around 25 ms for each frame with the resolution of 640×480. Afterwards, the detected 2D keypoints of 400 frames are sent to the self-calibration module. In this experiment, we iterate the human pose and camera pose with 5 iterations, and the total time is around 30 s. Finally, the human postural of each frame can be reconstructed with the optimization-based method and the run-time of each frame is around 120 ms.

For convenient qualitative evaluation of the ground verification test data, we implemented the self-calibration module and the postural reconstruction module on three different action sets. Figure 7 shows examples of the detection results of the OpenPose with Body25 definition. The reconstructed SMPL models were re-projected and rendered on the corresponding images, as shown in Figure 8. The reconstructed SMPL models are shown in Figure 9. More rendered re-projected data are shown in Figure 10.

(**a**) Cam 1. (**b**) Cam 2. (**c**) Cam 3. (**d**) Cam 4.

Figure 7. The OpenPose detection results of four camera views with body25 definitions.



(**a**) Cam 1. (**b**) Cam 2. (**c**) Cam 3. (**d**) Cam 4.

Figure 8. The SMPL parameters' fitting results with projected vertex on the original images.

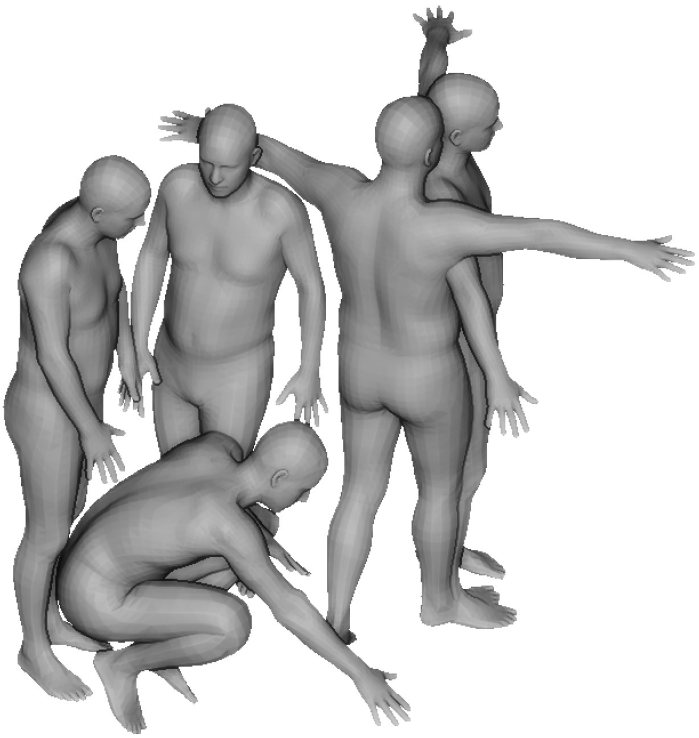


Figure 9. The reconstructed 3D posture.

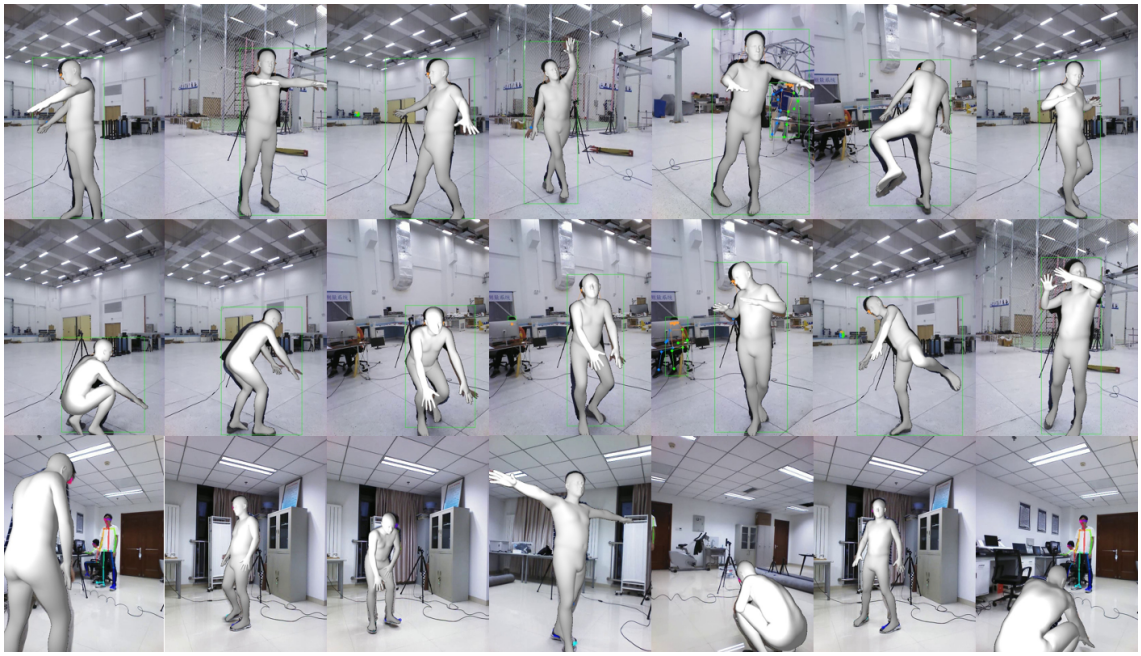


Figure 10. Qualitative results of the ground verification test data.

5. Discussion

We presented a calibration-free multi-view system for convenient in-orbit astronaut performance capture. We introduced several improvements to our proposed approach. First, we used an alternating iterative optimization instead of the two-step methods containing an initial calibration and bundle adjustment. We achieved better calibration results with different actions, which is meaningful due to the limited space and variety in actions due to microgravity. However, the proposed calibration process was restricted to only one person performing actions in the center of the scene for the effective selection of the detected 2D keypoints. Second, scanned models on ground are used to provide a scale factor to recover the actual distances between each camera according to limb length, which is considered invariant with the long-term weightlessness impact. The shape parameters were then refined with the detected 2D keypoints. The posture of the astronaut when performing different actions was then reconstructed with the shape parameters' fixed optimization. An evaluation of benchmarks and ground verification test data demonstrated the effectiveness and efficiency of our approach. Usually, two or more astronauts work together in the space-limited stations. Thus, we plan to add a person re-ID module to identify the different astronaut working and explore the temporal continuity property of the action series to remove the limitations of the self-calibration process. With the reconstructed posture data, the astronaut ergonomics for space flight in the microgravity environment can be assessed in future work.

Author Contributions: Conceptualization, S.G.; Methodology, S.G.; Validation, S.G.; Data curation, S.G.; Writing—review & editing, S.Z., C.N. and L.Z.; Visualization, S.G.; Supervision, X.Z.; Project administration, X.Z.; Funding acquisition, X.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The present work received the support of the Foundation of Key Laboratory of National Defense Science and Technology of Human Factors Engineering (614222210401) and the Foundation of China Astronaut Research and Training Center (2022SY54B0605).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. This work involved human subjects in this research. Approval of all ethical and experimental procedures and protocols was granted by the Biomedical Research Ethics Committee of Sun Yat-Sen University.

Data Availability Statement: Not applicable.

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

References

1. Amir, A.; Baroni, G.; Pedrocchi, A.; Newman, D.; Ferrigno, G.; Pedotti, A. Measuring astronaut performance on the ISS: Advanced kinematic and kinetic instrumentation. In Proceedings of the IMTC/99, 16th IEEE Instrumentation and Measurement Technology Conference (Cat. No. 99CH36309), Venice, Italy, 24–26 May 1999; Volume 1, pp. 397–402.
2. Wu, E.Q.; Tang, Z.R.; Xiong, P.; Wei, C.F.; Song, A.; Zhu, L.M. ROpenPose: A rapider OpenPose model for astronaut operation attitude detection. *IEEE Trans. Ind. Electron.* **2021**, *69*, 1043–1052. [CrossRef]
3. Wang, W.; Zhang, W.; Feng, W. The astronaut ergonomics assessment methodology in microgravity environment. In Proceedings of the 2017 Second International Conference on Reliability Systems Engineering (ICRSE), Beijing, China, 10–12 July 2017; pp. 1–7.
4. Xia, S.; Gao, L.; Lai, Y.K.; Yuan, M.Z.; Chai, J. A survey on human performance capture and animation. *J. Comput. Sci. Technol.* **2017**, *32*, 536–554. [CrossRef]
5. Mihcin, S. Simultaneous validation of wearable motion capture system for lower body applications: Over single plane range of motion (ROM) and gait activities. *Biomed. Eng. Tech.* **2022**, *67*, 185–199. [CrossRef] [PubMed]
6. McGrath, T.M. IMU-Based Estimation of Human Lower Body Kinematics and Applications to Extravehicular Operations. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2021.
7. 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]
8. Gall, J.; Rosenhahn, B.; Brox, T.; Seidel, H.P. Optimization and filtering for human motion capture. *Int. J. Comput. Vis.* **2010**, *87*, 75–92. [CrossRef]
9. Liu, Y.; Stoll, C.; Gall, J.; Seidel, H.P.; Theobalt, C. Markerless motion capture of interacting characters using multi-view image segmentation. In Proceedings of the CVPR 2011, Washington, DC, USA, 20–25 June 2011; pp. 1249–1256.
10. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.
11. 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.
12. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5693–5703.
13. Dong, J.; Jiang, W.; Huang, Q.; Bao, H.; Zhou, X. Fast and robust multi-person 3d pose estimation from multiple views. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 7792–7801.
14. Qiu, H.; Wang, C.; Wang, J.; Wang, N.; Zeng, W. Cross view fusion for 3d human pose estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 4342–4351.
15. Zhang, Y.; Wang, C.; Wang, X.; Liu, W.; Zeng, W. Voxeltrack: Multi-person 3d human pose estimation and tracking in the wild . *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2613–2626. [CrossRef] [PubMed]
16. Takahashi, K.; Mikami, D.; Isogawa, M.; Kimata, H. Human pose as calibration pattern; 3D human pose estimation with multiple unsynchronized and uncalibrated cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1775–1782.
17. Xu, Y.; Li, Y.J.; Weng, X.; Kitani, K. Wide-baseline multi-camera calibration using person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13134–13143.
18. Lee, S.E.; Shibata, K.; Nonaka, S.; Nobuhara, S.; Nishino, K. Extrinsic Camera Calibration from a Moving Person. *IEEE Robot. Autom. Lett.* **2022**, *7*, 10344–10351. [CrossRef]
19. Winnard, A.; Nasser, M.; Debuse, D.; Stokes, M.; Evetts, S.; Wilkinson, M.; Hides, J.; Caplan, N. Systematic review of countermeasures to minimise physiological changes and risk of injury to the lumbopelvic area following long-term microgravity. *Musculoskelet. Sci. Pract.* **2017**, *27*, S5–S14. [CrossRef] [PubMed]
20. Ferrigno, G.; Pedrocchi, A.; Baroni, G.; Bracciaferri, F.; Neri, G.; Pedotti, A. ELITE-S2: The multifactorial movement analysis facility for the International Space Station. *Acta Astronaut.* **2004**, *54*, 723–735. [CrossRef] [PubMed]
21. Neri, G.; Mascetti, G.; Zolesi, V. ELITE S2–A Facility for Quantitative Human Movement Analysis on Board the ISS. *Microgravity Sci. Technol.* **2014**, *26*, 271–278. [CrossRef]
22. Lee, M.W. Exercise Sensing and Pose Recovery Inference Tool (ESPRIT)-A Compact Stereo-based Motion Capture Solution For Exercise Monitoring. In *An Overview of SBIR Phase 2 Physical Sciences and Biomedical Technologies in Space*; NASA: Washington, DC, USA, 2015.
23. Available online: https://humanresearchroadmap.nasa.gov/Tasks/task.aspx?i=1557 (accessed on 29 July 2022)

24. Wang, J.; Tan, S.; Zhen, X.; Xu, S.; Zheng, F.; He, Z.; Shao, L. Deep 3D human pose estimation: A review. *Comput. Vis. Image Underst.* **2021**, *210*, 103225. [CrossRef]
25. Wang, L.; Duan, F.; Lu, K. An adaptively weighted algorithm for camera calibration with 1D objects. *Neurocomputing* **2015**, *149*, 1552–1559. [CrossRef]
26. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
27. Guan, J.; Deboeverie, F.; Slembrouck, M.; Van Haerenborgh, D.; Van Cauwelaert, D.; Veelaert, P.; Philips, W. Extrinsic calibration of camera networks using a sphere. *Sensors* **2015**, *15*, 18985–19005. [CrossRef] [PubMed]
28. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 430–443.
29. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
30. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
31. Zheng, Y.; Sugimoto, S.; Okutomi, M. A practical rank-constrained eight-point algorithm for fundamental matrix estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 23–28 June 2013; pp. 1546–1553.
32. Barath, D. Five-point fundamental matrix estimation for uncalibrated cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 235–243.
33. Cheng, Y.; Lopez, J.A.; Camps, O.; Sznaier, M. A convex optimization approach to robust fundamental matrix estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2170–2178.
34. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In Proceedings of the International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999; pp. 298–372.
35. Pavllo, D.; Feichtenhofer, C.; Grangier, D.; Auli, M. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 7753–7762.
36. Barath, D.; Noskova, J.; Ivashechkin, M.; Matas, J. MAGSAC++, a fast, reliable and accurate robust estimator. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1304–1312.
37. Zheng, Y.; Kuang, Y.; Sugimoto, S.; Astrom, K.; Okutomi, M. Revisiting the pnp problem: A fast, general and optimal solution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2344–2351.
38. Yang, K.; Fang, W.; Zhao, Y.; Deng, N. Iteratively reweighted midpoint method for fast multiple view triangulation. *IEEE Robot. Autom. Lett.* **2019**, *4*, 708–715. [CrossRef]
39. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL: A skinned multi-person linear model. *ACM Trans. Graph. (TOG)* **2015**, *34*, 1–16. [CrossRef]
40. Moakher, M. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.* **2002**, *24*, 1–16. [CrossRef]
41. Trumble, M.; Gilbert, A.; Hilton, A.; Collomosse, J. Deep autoencoder for combined human pose estimation and body model upscaling. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 784–800.
42. Pavlakos, G.; Zhou, X.; Derpanis, K.G.; Daniilidis, K. Harvesting multiple views for marker-less 3d human pose annotations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6988–6997.
43. Huang, Y.; Bogo, F.; Lassner, C.; Kanazawa, A.; Gehler, P.V.; Romero, J.; Akhter, I.; Black, M.J. Towards accurate marker-less human shape and pose estimation over time. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 421–430.