


Article

A Novel Method of Human Joint Prediction in an Occlusion Scene by Using Low-Cost Motion Capture Technique

Jianwei Niu ^{1,*} , Xiai Wang ¹, Dan Wang ¹ and Linghua Ran ²

¹ School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China; wangxai_wang@163.com (X.W.); 13426023594@163.com (D.W.)

² China National Institute of Standardization, Beijing 100191, China; ranlh@cnis.ac.cn

* Correspondence: niujw@ustb.edu.cn

Received: 25 January 2020; Accepted: 16 February 2020; Published: 18 February 2020



Abstract: Microsoft Kinect, a low-cost motion capture device, has huge potential in applications that require machine vision, such as human-robot interactions, home-based rehabilitation and clinical assessments. The Kinect sensor can track 25 key three-dimensional (3D) “skeleton” joints on the human body at 30 frames per second, and the skeleton data often have acceptable accuracy. However, the skeleton data obtained from the sensor sometimes exhibit a high level of jitter due to noise and estimation error. This jitter is worse when there is occlusion or a subject moves slightly out of the field of view of the sensor for a short period of time. Therefore, this paper proposed a novel approach to simultaneously handle the noise and error in the skeleton data derived from Kinect. Initially, we adopted classification processing to divide the skeleton data into noise data and erroneous data. Furthermore, we used a Kalman filter to smooth the noise data and correct erroneous data. We performed an occlusion experiment to prove the effectiveness of our algorithm. The proposed method outperforms existing techniques, such as the moving mean filter and traditional Kalman filter. The experimental results show an improvement of accuracy of at least 58.7%, 47.5% and 22.5% compared to the original Kinect data, moving mean filter and traditional Kalman filter, respectively. Our method provides a new perspective for Kinect data processing and a solid data foundation for subsequent research that utilizes Kinect.

Keywords: Kinect; Kalman filter; occlusion; human joint prediction

1. Introduction

The development of robotics technology is driving the application of robots from industrial production to the military, medical, and service fields [1–3]. In industrial production lines, industrial robots can replace workers in various tasks, such as assembly, handling, pick-up and welding, which can greatly improve work efficiency [4]. In the military, robots can be operated to perform dangerous tasks, such as bomb and mine defusing [5]. However, in the service field, robots are often used to handle more complex tasks that require people’s involvement [6]. Therefore, the combination of robot control technology and human-computer interaction technology can effectively improve the working ability and intelligence of civil robots [7,8].

At present, the control of civilian robots has been transformed from the traditional manual control mode, such as remote control and operation handling, to the vision-based robot control mode [9]. Visual-based robotic somatosensory control methods are gaining increasing applications, such as the treatment of children with autism, robot classroom teaching, and assisting robots [10–12]. This type of control mode is simple to operate, more in line with the human mindset and easy to perform by even

children and elderly people. However, how to obtain human motion information as a control signal for somatosensory operation of the robot is an urgent problem that needs to be solved. Marker-based motion capture settings (such as VICON, <https://www.vicon.com>) are a potential solution in this area because of their proven accuracy [13,14], but they are very expensive and cumbersome to use. Robot control based on color image information is the current main somatosensory robot control method [15]. This control method is simple, natural and convenient, but it is subject to environmental lighting, background complexity and human skin color. Thus, an inexpensive, environmentally unaffected system is essential for robot control.

Kinect is a somatosensory sensor from Microsoft that is low-cost and mainly used in the civilian field [16]. Kinect uses an infrared (IR) projector, an IR sensor and an RGB (Red Green Blue) camera to track human joints in three-dimensional (3D) space, which enables it to analyze joint kinematics [17–19]. However, the skeleton data obtained from the Kinect exhibit a high level of jitter due to noise and estimation error. This jitter worsens when there is occlusion or a subject moves slightly out of the field of view of the sensor for a short period of time [20]. Nonetheless, researchers have shown great interest in Kinect and applied it to home-based rehabilitation, clinical assessments and ergonomics. Wochatz et al. [21] think that the Kinect system can reliably assess lower limb joint angles and positions during simple rehabilitation exercises. Sarsfield et al. [22] present a clinical qualitative and quantitative analysis of the pose estimation algorithms of Kinect to assess its suitability for technology-supervised rehabilitation and to guide the development of future pose estimation algorithms for rehabilitation applications. Manghisi et al. [23], Xu et al. [24] and Plantard et al. [25] suggest a Rapid Upper Limb Assessment (RULA) assessment using the Kinect v2 sensor, where an ergonomic assessment is performed by computer processing and skeleton tracking. However, if the data obtained by Kinect are inaccurate, it will seriously affect these studies, thus, it is necessary to process Kinect skeleton data.

Various approaches are employed to stabilize joint coordinates. The main approaches are filter algorithms, such as the amplitude-limited filter, moving mean filter and Kalman filter. Edwards and Green [26] compared four different filter-based approaches to obtain smooth joint coordinates: the Kinect SDK's built-in Holt double exponential smoothing filter, an averaging filter, a Kalman filter with a constant-value model, and a Kalman filter with a Wiener Process Acceleration (WPA) model. Du and Zhang [27] proposed an innovative amplitude-limited algorithm of over-damping to solve the problem of error extraction and dithering due to the noncontact measure. Rosado et al. [28] improved the accuracy of the motions captured by Kinect from both static and dynamic aspects. Static calibration was used to obtain the average static distance of adjacent joints, and the joint position was optimized in the dynamic calibration using this static distance. Wang et al. [29] proposed a kinematic filtering algorithm based on the Unscented Kalman Filter and kinematic model of the human skeleton. The proposed algorithm can obtain a smooth kinematic parameter with reduced noise compared to the kinematic parameter generated from the raw motion data from Kinect. The traditional time series filter method has real-time performance and low algorithm complexity, which can partially remove jitter and noise from the Kinect joint data. However, abnormal joint data with large errors cannot be completely eliminated.

Researchers have adopted various approaches for dealing with abnormal joint data with large errors, for example, the joint estimation algorithm. Shen et al. [30] proposed an exemplar-based method to learn to correct the initially estimated joint-based skeleton, and observed a significant improvement compared to the approaches delivered by the current Kinect system. Shum et al. [31] proposed a set of erroneous data identification methods and established a human joint posture database to find the best substitute data from erroneous data. Liu et al. [32] proposed a posture reconstruction method based on a local mixture of Gaussian process models that Plantard et al. [33] adopted to filter pose graphs for efficient Kinect pose reconstruction. Approaches for abnormal joint data with large errors mostly use the real joint as a reference to learn the relationship between Kinect joint data and real joint data through machine learning, thus, they have high complexity and require real joint data as a reference. Moreover, different models have been developed for different types of motion learning, so they are not suitable for practical applications.

In the present paper, the advantages of the two methods are combined. We proposed a reliability index to identify abnormal joint data with large errors. Then, we improved the traditional Kalman filter according to various human movement constraints to realize the low-complexity joint correction algorithm. In the rest of the paper, Section 2 describes the proposed method. In Section 3, the experimental setup is explained, which also includes the experiment results. The conclusions and scope of future work are discussed in Section 4.

2. Methodology

2.1. Reliability Measurement

An incorrect skeleton joint in a motion capture system is even more damaging than a missed joint since it incorrectly guides the system to infer posture. Therefore, we applied an index called the vibration degree to evaluate the reliability of the joint [28].

When Kinect cannot accurately track a joint, there is a high-frequency vibration of the joint. Assuming $p_i(f) = (x_1, y_1, z_1)$ and $p_i(f + 1) = (x_2, y_2, z_2)$ to be the 3D position of skeleton i in two successive frames, we can calculate the displacement vectors as:

$$d_i(f) = p_i(f + 1) - p_i(f) = (x_2 - x_1, y_2 - y_1, z_2 - z_1) \quad (1)$$

The angle between continuous displacement vectors can be described as:

$$\theta_i(f) = \begin{cases} \arccos\left(\frac{d_i(f) \cdot d_i(f+1)}{|d_i(f)| |d_i(f+1)|}\right) & \text{if } |d_i(f)| > d_{\min}, |d_i(f+1)| > d_{\min} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where d_{\min} is the minimum distance value of an acceptable displacement vector. d_{\min} is used to avoid a large change in angle caused by small changes when the joint position is basically stationary. In our experiment, the distance value of a displacement vector is approximately 0.01 m when the joint position is basically stable. By contrast, when the joint position is unstable, the distance of the displacement vector increases. Therefore, the d_{\min} value is set to 0.02 m in our experiment.

The vibration degree reliability is defined as:

$$R_i(f) = 1 - \frac{\max(\min(\theta_i(f), \theta_{\max}) - \theta_{\min}, 0)}{\theta_{\max} - \theta_{\min}} \quad (3)$$

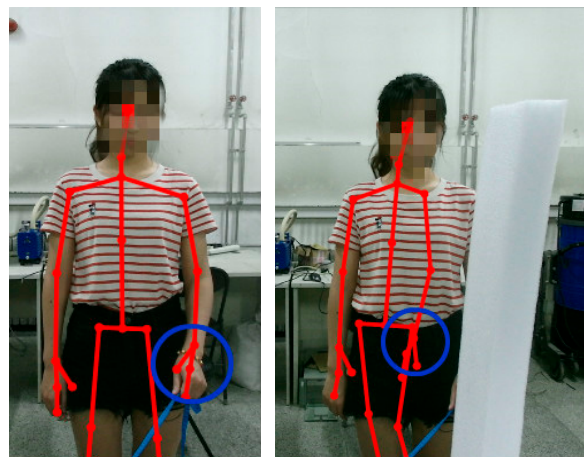
where θ_{\max} and θ_{\min} are the extremities of human body movement. θ_{\min} is the lower limit of the angle change when there is jitter between each frame, and θ_{\max} is the upper limit of the angle change that we consider. Based on Morasso [34], which is concerned with kinesiology, we set $\theta_{\min}=45^\circ$ and $\theta_{\max}=135^\circ$. However, the setting of the threshold values here are empirically determined, and this limitation is expected to be overcome in our future research.

2.2. Reliability Threshold

The main advantages of Kinect sensors are their low price, ease of use and adaptability to the environment. However, all sensors produce measurement errors and noise when measuring physical quantities. The Kinect sensor is an inaccurate system that provides joint measurement data with certain measurement errors and noise [35]. These errors and noise are generated by various factors, which can be classified into two main types. The first type is the lack of joint position information caused by occlusion and the part of the human body that leaves the measurement range. The Kinect sensor estimates the missing joint using the estimation algorithm and can obtain erroneous data. The second type is the systematic error introduced by quantization noise and sensor stability. The first type of data error may cause joint data to significantly deviate from the true value, which affects the accuracy of the joint data; the second type of error has a small amplitude but appears more frequently, which results in

uneven joint data [36]. Therefore, this paper classifies the two types of joint data and performs the corresponding processing after classification. We applied the vibration degree introduced in Section 2.1 to evaluate the reliability of the joint and determine the reliability threshold to divide the two types of joint data. Joint data with lower reliability than the threshold are recognized as abnormal data and are called erroneous data, and data with higher reliability than the threshold are identified as data to be optimized and are called noise data. This paper used the common approximation method in mathematics to obtain the joint reliability threshold, as follows.

First, the occlusion marker was artificially set in the experimental scene. Second, we obtained the motion data through occlusion from the Kinect and calculated the joint reliability. Third, we simultaneously collected the human motion color image information to manually mark wrong joint data frame by frame, as shown in Figure 1. Finally, we used the approximation idea to determine the joint reliability threshold. When the current threshold identification error data are lower than the manual labeling, the current threshold is set to the lower threshold. When the current threshold identification error data are higher than the manual labeling, the current threshold is set to the upper threshold. The approximation algorithm stops when the threshold judgment and manual labeling error are within 10 percent of each other.



(a) Correct joints

(b) Wrong joints.

Figure 1. Visualization of skeleton tracking.

In the present paper, wrist joint motion data of five subjects were collected. Each subject repeated five experiments, and each experiment collected 150 frames of data. We manually marked the number of frames of wrong joint and used the approximation algorithm to determine the reliability threshold. The results are shown in Table 1. Generally speaking, the data difference is not big enough, which may lead to doubts about the rationality of classification. However, in our opinion, the difference is a relative concept. Whether the difference is significant or not depends on the specific application. For example, if the proposed method in this paper is applied in the simulation of physical exercise such as table tennis playing, the data difference we provided is not big enough since the amplitude of the arm of the player in such kind of motion is quite big. In contrast, if the proposed method in this paper is applied in the simulation of rehabilitation training of patients with Parkinson's, the data difference we provided is very big since the amplitude of the arm of the patients in such kind of motion is quite small. Therefore, we determined the reliability threshold is the average value of the experimental data from 25 groups of 0.70 based on the results in Table 1 eventually. We defined erroneous data as joint data with a reliability threshold below 0.70 and noise data as joint data with a reliability threshold above 0.70.

Table 1. Reliability threshold results.

Experiment	Reliability Threshold				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
1	0.69	0.68	0.74	0.72	0.79
2	0.70	0.73	0.65	0.77	0.75
3	0.72	0.69	0.64	0.76	0.66
4	0.75	0.74	0.63	0.73	0.73
5	0.66	0.65	0.66	0.69	0.69

2.3. Algorithm to Handle Noise Data

Joint data with a reliability threshold above 0.70 are defined as noise data, and a Kalman filter is used to smooth the noise of the data. Except for separately obtaining each joint coordinate, we used Kinect to collect the sound source angle of the subject. Therefore, the state vector is taken to be the true 3D coordinates of the skeleton joint and their velocities and is written as $X = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$. The measurement vector is taken to be the true 3D coordinates of the skeleton joint and sound source angle and is written as $Y = [x, y, z, \arctan(x/z)]^T$. The state transition process is modeled as a linear dynamic system, and the measurement is modeled as a nonlinear dynamic system, where the next state at time instance $k+1$ is expressed in terms of the previous state at the k th instance and mathematically represented as:

$$X_{k+1} = FX_k + Q_k \quad (4)$$

$$Y_k = h(X_k) + R_k \quad (5)$$

where X_k and Y_k are the state vector and measurement vector, respectively, at time instant k ; Q_k and R_k are the process noise and measurement noise, respectively; F is the state transition matrix; and h is the state transformation function.

Matrix F is given in block form by:

$$F = \begin{pmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

For state transformation function h , we adopted the extended Kalman Filter to linearize h and replace matrix H in the filter with the Jacobian of h , which is evaluated at the current state estimate as:

$$H_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1/\hat{z}_k^-}{1+(\hat{x}_k^-/\hat{z}_k^-)^2} & 0 & \frac{\hat{x}_k^-/(\hat{z}_k^-)^2}{1+(\hat{x}_k^-/\hat{z}_k^-)^2} & 0 & 0 & 0 \end{pmatrix} \quad (7)$$

Kalman filter estimates \hat{X}_k from X_k with the knowledge of measurement vector Y_k in two steps: prediction and update. The standard Kalman filtering prediction step can be written as:

$$\hat{X}_k^- = F\hat{X}_{k-1} \quad (8)$$

$$P_k^- = FP_{k-1}^-F^T + Q_k \quad (9)$$

where P_k^- is the covariance matrix associate with prediction \hat{X}_k^- for an unknown true state X_k and is expressed as:

$$P_k^- = E[(X_k - \hat{X}_k^-)(X_k - \hat{X}_k^-)^T] \quad (10)$$

The updated state based on the measurement is expressed as:

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (11)$$

$$\hat{X}_k = \hat{X}_k^- + K_k (Y_k - H \hat{X}_k^-) \quad (12)$$

$$P_k = (I - K_k H) P_k^- \quad (13)$$

where K_k is the Kalman gain matrix. The Kalman filter minimizes the mean square error between the estimated \hat{X}_k and true X_k , providing smoother coordinates.

2.4. Algorithm to Handle Erroneous Data

We define erroneous data as joint data with a reliability threshold below 0.70, and a Kalman filter with human model constraints is used to correct the error of the data. To illustrate the algorithm to handle erroneous data, we assume that the wrong joint is wrist joint B at the k th frame and that its parent joint is elbow joint A (X_1, Y_1, Z_1), as shown in Figure 2.

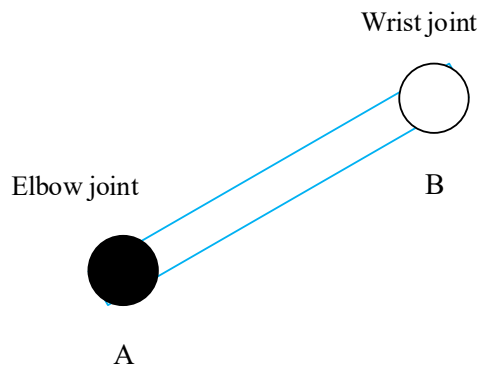


Figure 2. Schematic diagram of the error joint and its parent joint.

First, the Kalman filter algorithm was used to estimate the motion trend between frames to obtain the error joint position estimate $P(\tilde{X}, \tilde{Y}, \tilde{Z})$. Then, we established the constraint equation. Since the length of the human skeleton is constant, it is estimated that the error joint should be on the spherical surface with radius l_{AB} at the center of the parent node. The constraint equation is as follows:

$$(X - X_1)^2 + (Y - Y_1)^2 + (Z - Z_1)^2 = l_{AB}^2 \quad (14)$$

Finally, the estimated joint position $(\tilde{X}, \tilde{Y}, \tilde{Z})$ is optimized. By establishing a spatial linear equation between $P(\tilde{X}, \tilde{Y}, \tilde{Z})$ and $A(X_1, Y_1, Z_1)$, we can acquire optimized joint position $B(\hat{X}, \hat{Y}, \hat{Z})$, which is on the constraint equation and closest to the estimated joint position $P(\tilde{X}, \tilde{Y}, \tilde{Z})$, as shown in Figure 3.

The constraint equation intersects the linear equation at two points. The solution with the smallest coordinate distance from the joint estimated position point P is selected as the final estimated error of the joint optimization estimated position.

$$\hat{X} = \pm \sqrt{\frac{l_{AB}^2 (\tilde{X} - X_1)}{(\tilde{X} - X_1)^2 + (\tilde{Y} - Y_1)^2 + (\tilde{Z} - Z_1)^2}} + \tilde{X} \quad (15)$$

$$\hat{Y} = \pm \sqrt{\frac{l_{AB}^2(\tilde{Y} - Y_1)}{(\tilde{X} - X_1)^2 + (\tilde{Y} - Y_1)^2 + (\tilde{Z} - Z_1)^2}} + \tilde{Y} \quad (16)$$

$$\hat{Z} = \pm \sqrt{\frac{l_{AB}^2(\tilde{Z} - Z_1)}{(\tilde{X} - X_1)^2 + (\tilde{Y} - Y_1)^2 + (\tilde{Z} - Z_1)^2}} + \tilde{Z} \quad (17)$$

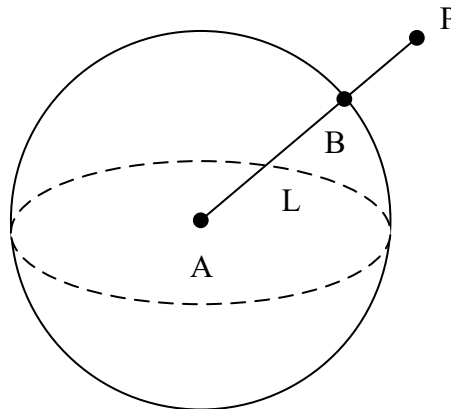


Figure 3. Schematic diagram of the position of error joint B.

3. Experimental Setup

Our experiment is based on Kinect version 2.0, which provides pose estimations for 25 “skeleton” joints at 30 Hz and enables the tracking of a user’s skeleton on a subset of joints [21]. A schematic of the Kinect, its sensor locations and its right-handed coordinate system is shown in Figure 4. The Kinect base sits parallel to the (x, z) plane, and the origin of the coordinate is at the center of the infrared camera. The X-axis runs parallel through the video and audio sensor arrays, the Y-axis runs perpendicular to the Kinect base, and the Z-axis defines the illumination direction. The coordinate unit is meter (m).

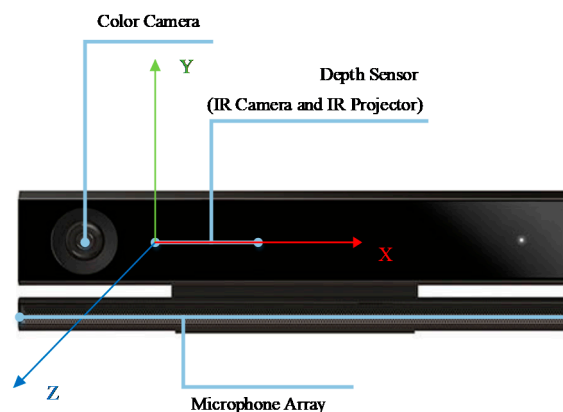


Figure 4. Kinect sensor and its coordinate system (IR means infrared).

In general, the accuracy of Kinect is evaluated by comparing data collected by Kinect with data acquired by optical motion capture devices (such as VICON). However, as described in Section 2.2, all sensors produce a few measurement errors when measuring physical quantities. Thus, we may not be able to obtain the most accurate joint position trajectory. Since the precise trajectory is difficult to measure, this paper abandoned the use of an optical motion capture instrument to obtain human skeleton joint positions as the ground truth. Instead, we adopted the trajectory acquisition method presented in [19], which first set the fixed point on the ground as the center of the special motion trajectory in the (X, Z) plane. The present paper selected the quarter circular trajectory. First, we determined a point as the

center of the quarter circular trajectory, which implies that we fixed the Y-direction coordinate of the human joint position. Then, we took a piece of a tape measure and attached it to the fixed point. Finally, we instructed the subject to face the Kinect at all times and move along the quarter circular path while holding the other end of the tape at the skeleton wrist joint. The obtained quarter circular trajectory of the wrist joint is considered to be the ground truth. Unlike [19], we added an obstruction to the joint trajectory to generate incorrect data. In this experiment, a total of five subjects' upper limb movement data were collected, and the experiment was repeated five times for each subject with 120 frames of experimental data. The experimental scene is shown in Figure 5.



Figure 5. Occlusion experimental setup.

4. Results and Discussion

Figure 6 below shows the performance of tracking the wrist trajectory using the original Kinect, the moving mean filter algorithm, the traditional Kalman filter algorithm and our method compared to the ground truth. The trajectory shown in the black circle in Figure 6 is erroneous data caused by occlusion.

From Figure 6, it is observed that the algorithm proposed in this paper is superior to the other algorithms. The idea of our method is to separate erroneous data from noise data, perform targeted processing of the identified erroneous data, discard the original erroneous data and estimate the new joint position by combining the human constraint and filtering prediction as the current joint position. Therefore, the algorithm presented in this paper is less affected by external measurement data and maintains a similar trend to the real trajectory near the erroneous data. The ordinary filtering method applies the erroneous data to the smoothing process, which is greatly affected by the external measurement data. Therefore, the movement trend of the measurement data will remain in the vicinity of the erroneous data, and the deviation is large.

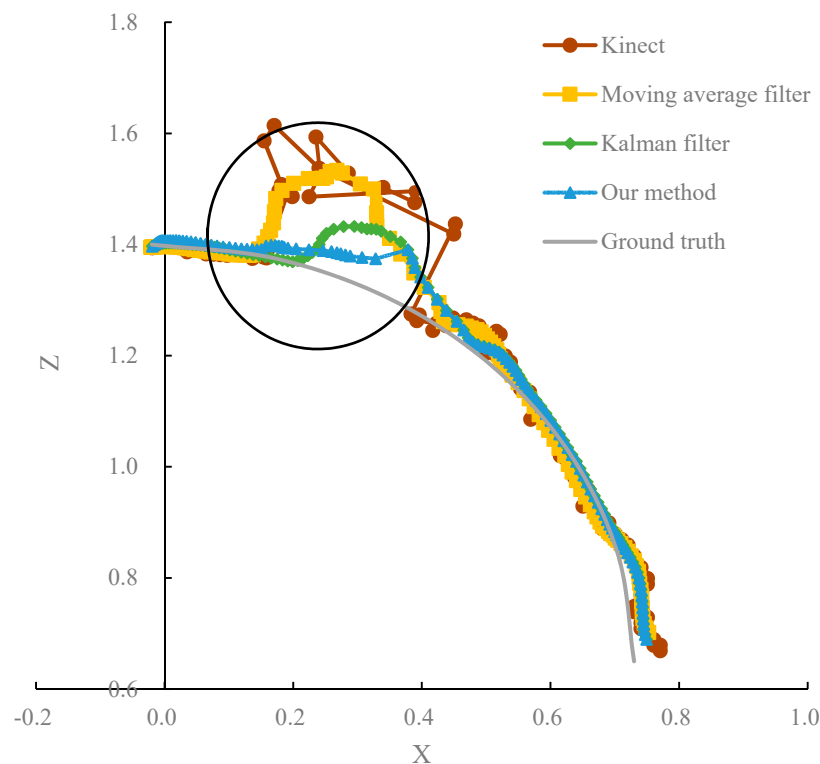


Figure 6. Accuracy comparison of different algorithms and the true trajectory.

To measure accuracy, the average error of the estimated joint position and the true trajectory were calculated using the following formula:

$$E = \frac{\sum_{i=1}^n \sqrt{(x(i) - x_0(i))^2 + (z(i) - z_0(i))^2}}{n} \quad (18)$$

where $x(i)$ and $z(i)$ are the x and z components of the human joint position coordinate of the i th frame processed by different algorithms, respectively; $x_0(i)$ and $z_0(i)$ are the x and z components of the human joint position coordinate of the i th frame in the true trajectory, respectively.

Table 2 shows the error of the original joint movement trajectory acquired by Kinect; the joint trajectories processed by the moving mean filter algorithm, the traditional Kalman filter algorithm and the algorithm proposed in this paper; and the true geometric trajectory. Table 2 shows that the joint data processing algorithm proposed in this paper is superior to the other algorithms in regard to the overall average error comparison. Based on the original Kinect data, the data accuracy was improved by 21.3% after moving mean filter. After the traditional Kalman filter processing, the data accuracy is increased by 46.7%, and after the algorithm proposed in this paper processing, the data accuracy is increased by 58.7%.

As for computational efficiency, though the algorithm complexity of our method is higher than other algorithm like moving mean filter algorithm, it is not obvious in terms of the difference. Since the moving mean filter algorithm is simple, we observed that it takes roughly 0.975 s to process one experiment sample, whereas it takes roughly 1.248 s for traditional Kalman filter to process one sample. Our proposed method adds classification algorithm before extend Kalman filter so that it takes roughly 1.592 s to adapt one sample. All the algorithms are performed on the MATLAB 2016b platform with 3.1 GHz Intel Core i5 Processor.

Table 2. Comparison of the error of different algorithms (unit: m).

	Kinect		Moving Mean Filter		Kalman Filter		Our Method	
	Error	SD	Error	SD	Error	SD	Error	SD
1	0.081	0.008	0.065	0.007	0.043	0.003	0.032	0.003
2	0.076	0.004	0.061	0.005	0.041	0.002	0.031	0.003
3	0.071	0.004	0.054	0.010	0.036	0.003	0.028	0.005
4	0.069	0.007	0.051	0.009	0.039	0.002	0.030	0.002
5	0.078	0.005	0.062	0.004	0.042	0.004	0.036	0.004
Mean	0.075	0.006	0.059	0.007	0.040	0.003	0.031	0.003

SD=Standard Deviation.

5. Conclusions and Future Work

Regarding the accuracy of Kinect, few studies have focused on improving the inherent skeleton tracking accuracy of Kinect. These studies simply intended to show that applications based on Kinect could be significantly improved by applying optimal techniques. In this process, the researchers ignored the generalization of methods to improve the accuracy of Kinect. This tendency is susceptible to the embarrassing situation that the method is suitable for posture assessment but not home rehabilitation. We proposed a novel algorithm to improve the accuracy of Kinect skeletal joint coordinates for improving the inherent accuracy of Kinect. Our method introduced a skeletal joint data classification algorithm to divide noise data and erroneous data. Furthermore, we proposed two different algorithms to smooth noise data and correct erroneous data to accurately track dynamic trajectory joint center location over time. Our method can potentially expand the way to process Kinect data in applications based on Kinect because we separate Kinect data processing from applications. Thus, our method is suitable for most applications related to Kinect.

The present paper evaluated the algorithm in an occlusion experiment. The results of these experiments are significant. The results show that the algorithm substantially smooths the skeleton joint position estimates of Kinect; more importantly, the experiments demonstrate that the tracking accuracy is significantly increased. In this study, we compared the results of our method with the original Kinect data, the moving mean filter algorithm and the traditional Kalman filter algorithm and obtained an accuracy improvement of 58.7%, 47.5% and 22.5%, respectively. As a result, using the skeletal joint data classification algorithm and two different data-processing algorithms to smooth noise data and correct erroneous data reduce the average estimation error for tracking human dynamic skeleton joints.

However, there are limitations to this study. Our proposed algorithm for Kinect skeletal joint data classification only considers the vibration between frames. However, there is also a limited relationship between the coordinates of each joint point in the same frame. For future work, we plan to enrich the skeletal joint data classification algorithm by incorporating the limit relationship. Furthermore, the setting of the values like reliability threshold is a shortcoming. We should consult with expertise from physiology, rehabilitation or even neuroscience to determine the reference threshold, and we should have conducted some preliminary experiments to verify the rationality of the data difference. In addition, we only considered the tracking of the (x, z) coordinate of the wrist joint of the subject with a known quarter circular trajectory. The results must be verified based on more complex motions. However, the true trajectories are difficult to measure and may require more sophisticated and expensive equipment, which will be conducted in possible future research.

Data Availability: The data used to support the findings of this study are included within this paper. It is also available from the corresponding author upon request.

Author Contributions: Conceptualization: J.N.; Formal analysis: J.N.; Funding acquisition: L.R.; Investigation: X.W. and D.W.; Methodology: J.N., X.W.; D.W. and L.R.; Validation: X.W.; Writing – original draft: X.W.; Writing – review & editing: J.N. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2017YFF0206602).

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

References

1. Lasota, P.A.; Shah, J.A. A Multiple-Predictor Approach to Human Motion Prediction. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, Singapore, 29 May–3 June 2017.
2. Haidegger, T.; Barreto, M.; Gonçalves, P.; Habib, M.K.; Ragavan, S.K.V.; Li, H.; Vaccarella, A.; Perrone, R.; Prestes, E. Applied ontologies and standards for service robots. *Rob. Autom. Syst.* **2013**, *61*, 1215–1223. [[CrossRef](#)]
3. Habib, M.K.; Baudoin, Y.; Nagata, F. Robotics for rescue and risky intervention. In Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society (IECON 2011), Melbourne, VIC, Australia, 7–10 November 2011.
4. Michalos, G.; Makris, S.; Papakostas, N.; Mourtzis, D.; Chryssolouris, G. Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP J. Manuf. Sci. Technol.* **2010**, *2*, 81–91. [[CrossRef](#)]
5. Prakash, P.; Dhanasekaran, C.; Surya, K.; Pius, V.K.; Vishal, A.S.; Kumar, S.V. Gesture controlled dual six axis robotic arms with rover using MPU. *Mater. Today Proc.* **2019**, *21*, 547–556. [[CrossRef](#)]
6. Kuhner, D.; Fiederer, L.D.J.; Aldinger, J.; Burget, F.; Völker, M.; Schirrmeister, R.T.; Do, C.; Boedecker, J.; Nebel, B.; Ball, T.; et al. A service assistant combining autonomous robotics, flexible goal formulation, and deep-learning-based brain–computer interfacing. *Rob. Autom. Syst.* **2019**, *116*, 98–113. [[CrossRef](#)]
7. Tseng, S.H.; Chao, Y.; Lin, C.; Fu, L.C. Service robots: System design for tracking people through data fusion and initiating interaction with the human group by inferring social situations. *Rob. Autom. Syst.* **2016**, *83*, 188–202. [[CrossRef](#)]
8. Paulius, D.; Sun, Y. A Survey of Knowledge Representation in Service Robotics. *Rob. Autom. Syst.* **2019**, *118*, 13–30. [[CrossRef](#)]
9. Ramoly, N.; Bouzeghoub, A.; Finance, B. A Framework for Service Robots in Smart Home: An Efficient Solution for Domestic Healthcare. *IRBM* **2018**, *39*, 413–420. [[CrossRef](#)]
10. Tseng, R.Y.; Do, E.Y.L. Facial Expression Wonderland: A Novel Design Prototype of Information and Computer Technology for Children with Autism Spectrum Disorder. In Proceedings of the 1st ACM International Health Informatics Symposium (IHI 2010), Arlington, VA, USA, 11–12 November 2010.
11. Blanchard, S.; Freiman, V.; Lirrete-Pitre, N. Strategies used by elementary schoolchildren solving robotics-based complex tasks: innovative potential of technology. *Procedia Soc. Behav. Sci.* **2010**, *2*, 2851–2857. [[CrossRef](#)]
12. Nergui, M.; Imamoglu, N.; Yoshida, Y.; Gonzalez, J.; Sekine, M.; Kawamura, K.; Yu, W.W. Human Behavior Recognition by a Mobile Robot Following Human Subjects. In *Evaluating AAL Systems Through Competitive Benchmarking*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 159–172.
13. Carse, B.; Meadows, B.; Bowers, R.; Rowe, P. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3d motion analysis system. *Physiotherapy* **2013**, *99*, 347–351. [[CrossRef](#)]
14. Pfister, A.; West, A.; Bronner, S.; Noah, J.A. Comparative Abilities of Microsoft Kinect and Vicon 3D Motion Capture for Gait Analysis. *J. Med. Eng. Technol.* **2014**, *38*, 274–280. [[CrossRef](#)]
15. Shen, W.; Bai, X.; Hu, R.; Wang, H.Y.; Latecki, L.J. Skeleton Growing and Pruning with Bending Potential Ratio. *Pattern Recognit.* **2011**, *44*, 196–209. [[CrossRef](#)]
16. Kean, S.; Hall, J.; Perry, P. *Meet the Kinect: An Introduction to Programming Natural User Interfaces*; Apress: New York, NY, USA, 2011.
17. Tripathy, S.R.; Chakravarty, K.; Sinha, A.; Chatterjee, D.; Saha, S.K. Constrained Kalman Filter For Improving Kinect Based Measurements. In Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, USA, 28–31 May 2017.
18. Das, P.; Chakravarty, K.; Chowdhury, A.; Chatterjee, D.; Sinha, A.; Pal, A. Improving joint position estimation of Kinect using anthropometric constraint based adaptive Kalman filter for rehabilitation. *Biomed. Phys. Eng. Express* **2018**, *4*, 035002. [[CrossRef](#)]

19. Clark, R.A.; Mentiplay, B.F.; Hough, E.; Pua, Y.H. Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. *Gait Posture* **2019**, *68*, 193–200. [CrossRef] [PubMed]
20. Shu, J.; Hamano, F.; Angus, J. Application of extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates. *J. Eng. Math.* **2014**, *88*, 161–175. [CrossRef]
21. Wochatz, M.; Tilgner, N.; Mueller, S.; Rabe, S.; Eichler, S.; John, M.; Völler, H.; Mayer, F. Reliability and validity of the Kinect V2 for the assessment of lower extremity rehabilitation exercises. *Gait Posture* **2019**, *70*, 330–335. [CrossRef]
22. Sarsfield, J.; Brown, D.; Sherkat, N.; Langensiepen, C.; Lewis, J.; Taheri, M.; McCollin, C.; Barnett, C.; Selwood, L.; Standen, P.; et al. Clinical assessment of depth sensor based pose estimation algorithms for technology supervised rehabilitation applications. *Int. J. Med. Informatics* **2019**, *121*, 3038. [CrossRef]
23. Manghisi, V.M.; Uva, A.E.; Fiorentino, M.; Bevilacqua, V.; Trotta, G.F.; Monno, G. Real time RULA assessment using Kinect v2 sensor. *Appl. Ergon.* **2017**, *65*, 481–491. [CrossRef]
24. Xu, X.; Robertson, M.; Chen, K.B.; Lin, J.H.; McGorry, R.W. Using the Microsoft Kinect™ to assess 3-D shoulder kinematics during computer use. *Appl. Ergon.* **2017**, *65*, 418–423.
25. Plantard, P.; Shum, H.P.H.; Pierres, A.S.L.; Multon, F. Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Appl. Ergon.* **2017**, *65*, 562–569. [CrossRef]
26. Edwards, M.; Green, R. Low-latency filtering of kinect skeleton data for video game control. In Proceedings of the 29th International Conference on Image and Vision Computing New Zealand, Hamilton, New Zealand, 19–21 November 2014.
27. Du, G.L.; Zhang, P. Markerless Human-Robot Interface for Dual Robot Manipulators Using Kinect Sensor. *Rob. Comput. Integr. Manuf.* **2014**, *30*, 150–159. [CrossRef]
28. Rosado, J.; Silva, F.; Santos, V. A Kinect-Based Motion Capture System for Robotic Gesture Imitation. In *ROBOT 2013: First Iberian Robotics Conference*; Springer: Cham, Switzerland, 2014; Volume 1, pp. 585–595.
29. Wang, Q.F.; Kurillo, G.; Ofli, F.; Bajcsy, R. Remote Health Coaching System and Human Motion Data Analysis for Physical Therapy with Microsoft Kinect. Available online: <https://arxiv.org/abs/1512.06492> (accessed on 1 January 2020).
30. Shen, W.; Deng, K.; Bai, X.; Leyvand, T.; Guo, B.N.; Tu, Z.W. Exemplar-Based Human Action Pose Correction and Tagging. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.
31. Shum, H.P.H.; Ho, E.S.L. Real-Time Posture Reconstruction for Microsoft Kinect. *IEEE Trans. Cybern.* **2013**, *43*, 1357–1369. [CrossRef] [PubMed]
32. Liu, Z.G.; Zhou, L.Y.; Leung, H.; Shum, H.P.H. Kinect Posture Reconstruction Based on a Local Mixture of Gaussian Process Models. *IEEE Trans. Visual Comput. Graphics* **2016**, *22*, 2437–2450. [CrossRef] [PubMed]
33. Plantard, P.; Shum, H.P.H.; Multon, F. Filtered Pose Graph for Efficient Kinect Pose Reconstruction. *Multimedia Tools Appl.* **2017**, *76*, 4291–4312. [CrossRef]
34. Morasso, P. Spatial Control of Arm Movements. *Exp. Brain Res.* **1981**, *42*, 223–227. [CrossRef]
35. Huber, M.E.; Seitz, A.L.; Leeser, M.; Sternad, D. Validity and Reliability of Kinect Skeleton for Measuring Shoulder Joint Angles: A Feasibility Study Chartered Society of Physiotherapy. *Physiotherapy* **2015**, *101*, 389–393. [CrossRef]
36. Shotton, J.; Sharp, T.; Kipman, A.; Girshick, R.; Fitzgibbon, A.; Cook, M.; Finocchio, M.; Moore, R.; Kohli, P.; Criminisi, A.; et al. Difficult Human Pose Estimation from Single depth Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2821–2840. [CrossRef]

