

Research Article

Real-Time Capture of Snowboarder's Skiing Motion Using a 3D Vision Sensor

Zhipeng Li,¹ Jun Wang,² Tao Zhang,² Dave Balne,³ Bing Li,² Ruizhu Yang,² Wenli Song ,² and Xingfu Zhang⁴

¹Winter Olympic College, Harbin Sports University, Harbin 150008, China

²College of Physical Education and Training, Harbin Sports University, Harbin 150008, China

³Georgia Mountain Club, Ontario, Canada

⁴Heilongjiang Institute of Technology, Harbin 150050, China

Correspondence should be addressed to Wenli Song; songwenli@hrbipe.edu.cn

Received 9 October 2021; Accepted 30 October 2021; Published 16 November 2021

Academic Editor: Alireza Souri

Copyright © 2021 Zhipeng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the influence of environmental interference and too fast speed, there are some problems in ski motion capture, such as inaccurate motion capture, motion delay, and motion loss, resulting in the inconsistency between the actual motion of later athletes and the motion of virtual characters. To solve the above problems, a real-time skiing motion capture method of snowboarders based on a 3D vision sensor is proposed. This method combines the Time of Flight (TOF) camera and high-speed vision sensor to form a motion acquisition system. The collected motion images are fused to form a complete motion image, and the pose is solved. The pose data is bound with the constructed virtual character model to drive the virtual model to synchronously complete the snowboarding motion and realize the real-time capture of skiing motion. The results show that the motion accuracy of the system is as high as 98.6%, which improves the capture effect, and the motion matching proportion is better and more practical. It is also excellent in the investigation of motion delay and motion loss.

1. Introduction

Snowboarding is an intense sport, which is mainly carried out in winter, and it is one of the necessary events of the Winter Olympic Games [1]. The sport is exciting, cool, and dangerous, and the training requirements are rather high.

Physical quality consists of technology, psychology, tactics, and other factors. In recent years, the development of physical fitness training in China has made great progress than before. In theory, we no longer blindly follow the traditional training mode of “three importance and one greatness,” and the training means and methods are constantly innovated. However, China's physical training development is still in the stage of catching up and learning, and there is still a big gap with foreign countries. In foreign countries, physical training pays more attention to the comprehensive and balanced develop-

ment of physical elements and puts forward many powerful concepts of physical training. Functional physical training is a popular training concept in recent years, but due to the lack of theoretical knowledge of functional physical training in China, we still lack a comprehensive and systematic understanding. This study collects a large number of the latest relevant literature and foreign literature, summarizes, arranges, and applies them to cross-country skiing physical training, so as to make a certain contribution to improving the competitive ability of cross-country skiing teams. Motion capture, also known as motion tracking, collects athletes' motion data in real time and then uses these data to drive virtual character simulation to display athletes' motions [2]. In the skiing motion training of professional athletes, motion capture is very important. It can be used to observe athletes' moving posture in real time, which is of great practical significance to correct wrong

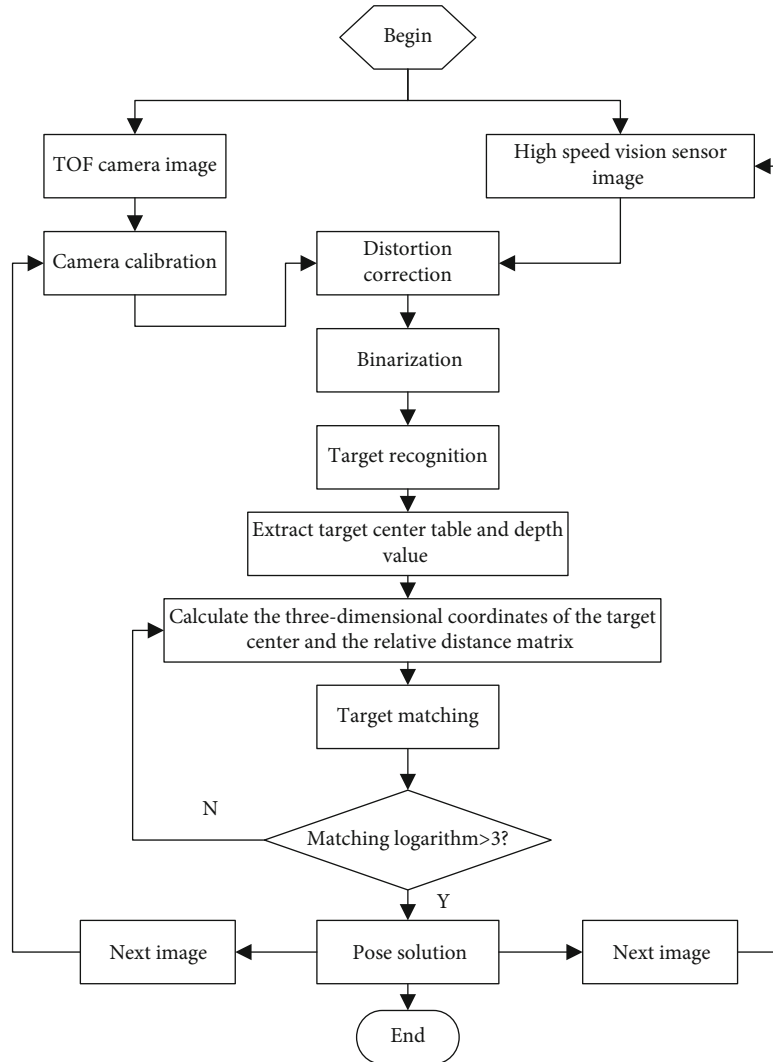


FIGURE 1: Pose solution process.

motions and improve athletes' skill level. At present, the accuracy of the motion capture system used by Chinese athletes in skiing training is not high. Because the skiing motion is too fast and difficult, it is easy to miss capture. If there is no more professional system to capture the motion of skiers for a long time, the coach will not be able to correct their mistakes in the skiing process in time, which will seriously affect the training speed of skiers [3]. To solve the above problems, a more suitable method for skiing motion capture is proposed, that is, snowboarding athletes' skiing motion real-time capture based on a 3D vision sensor. 3D display technology is a new technology with rapid development in recent years. At present, it is widely used in animation production. Compared with the 2D presentation effect, 3D display technology can present the picture more realistically and make people feel immersive. 3D vision sensor is the main device to realize 3D technology. Based on the original system, 3D technology is applied to this study in order to provide effective help for ski training.

The main contributions of this paper are as follows:

- (1) Reduce missed capture and improve the accuracy of the system in capturing motion
- (2) The application of a 3D vision sensor in real-time capture of snowboarding athletes' skiing motion improves the performance of motion capture, which is convenient to capture feature points and better simulate motion
- (3) Different datasets are used to simulate the skiing motion real-time capture system in this paper, and the feasibility of the system is tested

2. Related Work

Based on the above background, motion capture technology has been widely used in various sports professional training. For example, in literature [4], Wang et al. established a multi-view camera system to capture moving human body images from multiple views in real time, then estimated human body

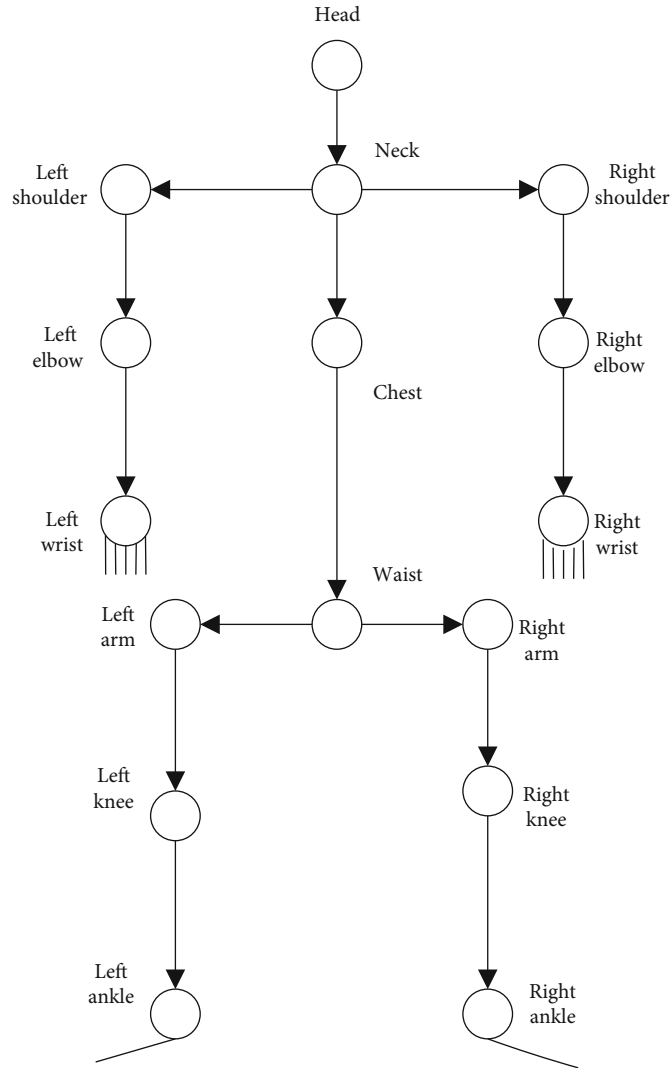


FIGURE 2: Human skeleton model.

TABLE 1: Training parameters of snowboarder's skiing motion.

Parameters	Value (or running state)
Number of input images	20000 sample images
Iterations	800
Optimizer	Random gradient descent
Initial value of weight correction	0.01

3D key points from multiview images using the OpenPose method, reconstructed dense 3D point clouds from the images, and finally fitted the SMPL model to 3D key points and 3D point clouds to represent human motion. However, the motion of this method is too complex. In literature [5], Guo and Song proposed a hand running direction capture method based on soft sensing. This method uses HTC Vive equipment to capture human hand motion and then combines the double-layer LSTM to estimate the hand motion and

clarify the hand motion characteristics. However, the matching proportion of this method is low. In literature [6], Yu et al. proposed a human motion capture method based on a single RGB-D sensor. This method realizes motion tracking refinement at the body part level through semantic tracking loss calculation, which can improve the tracking accuracy in the case of severe occlusion and fast motion. However, the motion accuracy of this method is low. In literature [7], Maruyama et al. proposed a MoCap capture method based on an inertial measurement unit. This method does not need any optical equipment and can measure the motion attitude even in the outdoor environment. However, the motion delay of this method is too long. Literature [1] proposed to use a series of precise physiological and biochemical instruments to detect the physical function of athletes, so as to master the law of the change of athletes' mechanical energy in sports training, scientifically organize and arrange the training and practical activities of snowboarding skills on veneer U-shaped venues, and reasonably guide athletes' diet. This can effectively prevent sports injury, prolong sports life, and improve sports ability. However,

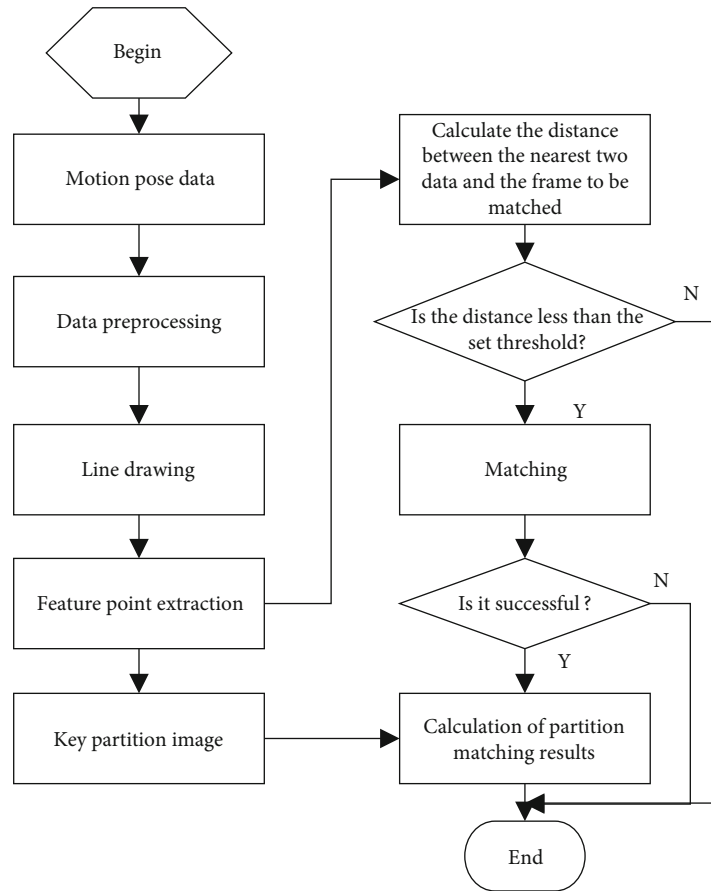


FIGURE 3: Comparison method of skiing motion details.

due to the loss of motion in this method, it is impossible to accurately capture snowboarding athletes' skiing motion.

Compared with the above systems, this paper uses a TOF camera and high-speed vision sensor as a motion acquisition system. This can greatly improve the accuracy of system motion simulation, and it is less affected by environmental interference. It can be applied to motion capture in various environments and has a good application prospect.

3. Real-Time Capture of Snowboarding Motion Using a 3D Vision Sensor

Snowboarding is dangerous, and motion learning is difficult. Ordinary people generally need several months to ensure that they can slide alone. However, for professional athletes with competition pressure, optical gliding is not enough. They also need continuous training to achieve excellent results in the formal competition. The training intensity of professional athletes is much higher than that of ordinary people, especially for the control of some details of motion. In order to assist the training of skiers, a real-time snowboarding capture method based on a 3D vision sensor is proposed in this study. This method mainly solves the following three problems:

(1) Interference by ambient light

(2) The motion speed is too fast to capture

(3) The athlete's actual motion cannot be consistent with the virtual character model motion

The first two problems are the main reasons for the last problem. Therefore, the latter problem can be solved by solving the first two problems and finally improved the accuracy of motion capture. This study is divided into three parts: snowboarding motion data acquisition, pose solution, and virtual character model establishment and driving. Specific analysis is made for these three parts.

3.1. Snowboarding Motion Image Acquisition. In order to solve the influence of ambient light interference and motion speed on motion capture, this design uses two kinds of 3D vision sensors to form an acquisition system to improve the comprehensiveness and integrity of motion data acquisition [8–10]. One of the two sensors is a TOF camera with less interference from ambient light, and the other is a high-speed vision sensor, which can be used to capture fast-moving targets. Through the combination of the two sensors, the problem of data loss in motion capture is reduced.

3.1.1. TOF Depth Camera. TOF depth camera collects reflected light to build depth mapped images. The distance from the camera to the object can be calculated according to time and

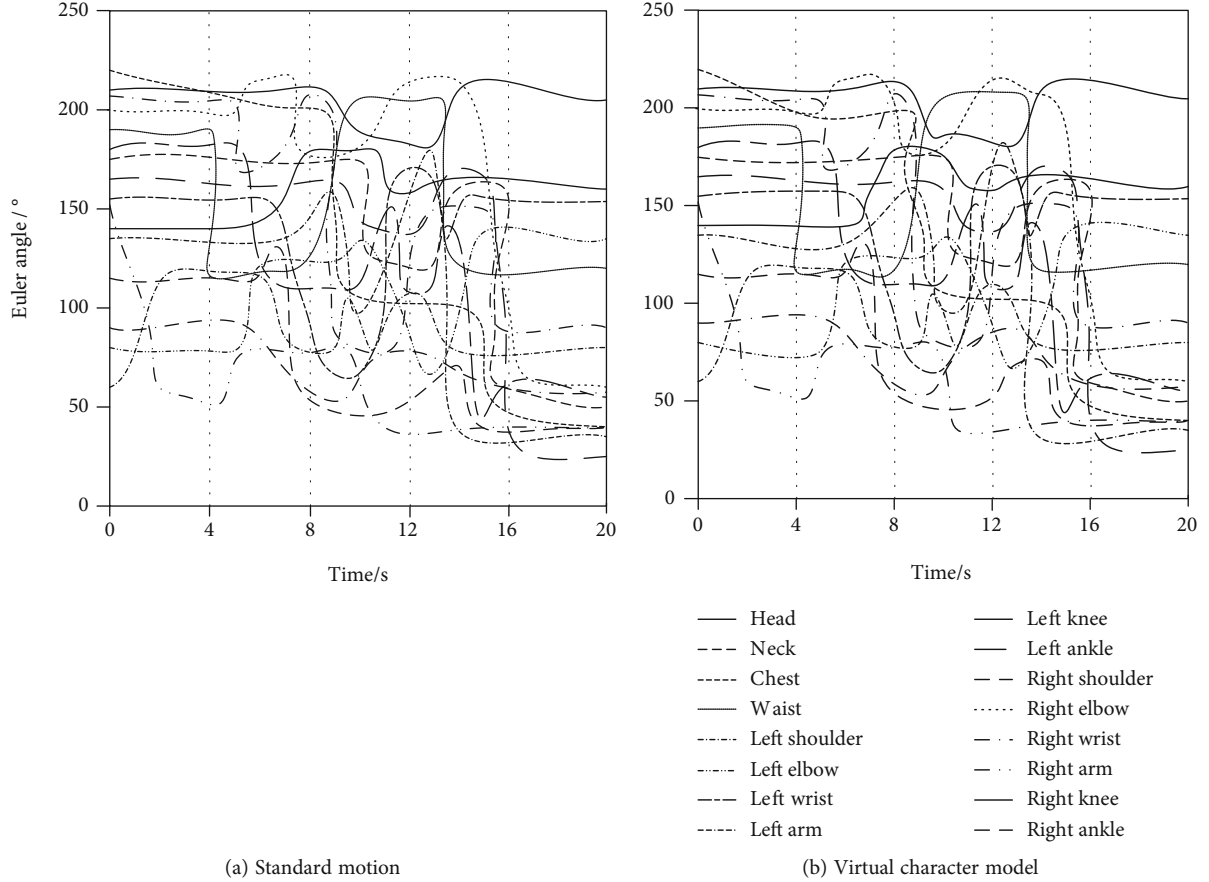


FIGURE 4: Line image segmentation results.

the speed of light. Formula (1) is as follows:

$$D = \frac{c \cdot \Delta t}{2}, \quad (1)$$

where D refers to the distance and Δt refers to the time for the pulse signal to travel from the camera and the target and then return [11].

3.1.2. High-Speed Vision Sensor. Compared with ordinary sensors, high-speed vision sensors adopt stacked structure and back-illuminated pixel array to improve their parallel processing ability [12, 13]. According to relevant research, the speed of detecting and tracking targets by high-speed vision sensor can reach 1000 frames per second. In this paper, a high-speed vision sensor is used to collect the data of dynamic target in real time, so as to reduce the missed acquisition probability of motion.

After the snowboarding athletes' skiing motion images are collected by the two 3D vision sensors, image fusion is required to integrate the two images. The fusion process is as follows:

Step 1. Multiscale transformation.

The motion images of two kinds of source snowboarding are decomposed by wavelet transform, and two kinds of

coefficients are obtained, namely, high-frequency coefficients and low-frequency coefficients.

(1) Edge feature

Step 2. Feature extraction.

The Canny operator is used to analyze the low-frequency coefficients of motion images of two kinds of source snowboarding $Y_{j,A}$ (low-frequency coefficient of TOF depth image) and $Y_{j,B}$ (low-frequency coefficient of high-speed vision sensor image) for the extraction of edge features and then construct binary edge feature $Z_{j,A}$ and $Z_{j,B}$ [14]. The construction formula is as follows:

$$\begin{aligned} Z_{j,A} &= \text{XOR}(Y_{j,A}, Y_{j,B}) \text{AND} Y_{j,A}, \\ Z_{j,B} &= \text{XOR}(Y_{j,A}, Y_{j,B}) \text{AND} Y_{j,B}, \end{aligned} \quad (2)$$

where XOR refers to the XOR operation symbol, $\text{XOR}(Y_{j,A}, Y_{j,B})$ means to calculate the low-frequency coefficients of TOF depth image and high-speed vision sensor image, and AND means the calculation symbol.

(2) Gradient feature

TABLE 2: Comparison results of key points in motion.

Category	Method	Standard motion	The proposed method	Literature [4] method	Literature [5] method	Literature [6] method	Literature [7] method	Literature [1] method
Motion complexity	One segment	58	58	56	55	55	56	54
	Second segment	125	126	120	122	120	124	120
	Third segment	256	250	236	237	233	235	237
	Fourth segment	240	238	230	224	220	227	224
	Fifth segment	105	103	100	101	100	98	96
Matching proportion	One segment	58	0.958	0.856	0.847	0.910	0.877	0.798
	Second segment	125	0.992	0.874	0.872	0.902	0.869	0.850
	Third segment	256	0.925	0.885	0.865	0.878	0.892	0.882
	Fourth segment	240	0.972	0.902	0.877	0.893	0.874	0.865
	Fifth segment	105	0.921	0.888	0.886	0.873	0.882	0.872
Motion accuracy (%)	One segment	100	98	90	89	87	85	86
	Second segment	100	98	95	90	86	87	90
	Third segment	100	99	93	91	85	85	82
	Fourth segment	100	99	96	93	90	97	97
	Fifth segment	10	99	95	96	91	90	85
Motion delay rate (ms)	One segment	0	15	50	60	40	30	20
	Second segment	0	14	45	60	44	35	21
	Third segment	0	12	48	58	43	28	28
	Fourth segment	0	15	49	57	40	29	29
	Fifth segment	0	13	52	55	39	26	25
Motion loss rate	One segment	Unclear	Unclear	Clear	Clear	Clear	Unclear	Clear
	Second segment	Unclear	Unclear	Clear	Clear	Clear	Clear	Unclear
	Third segment	Unclear	Unclear	Unclear	Unclear	Clear	Clear	Clear
	Fourth segment	Unclear	Unclear	Unclear	Clear	Clear	Clear	Unclear
	Fifth segment	Unclear	Unclear	Clear	Unclear	Clear	Unclear	Clear

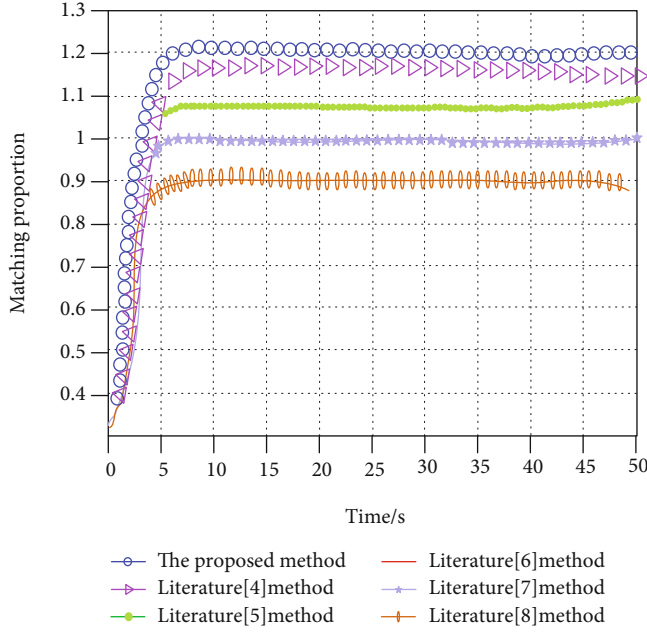


FIGURE 5: Comparison of matching proportion of skiing motion.

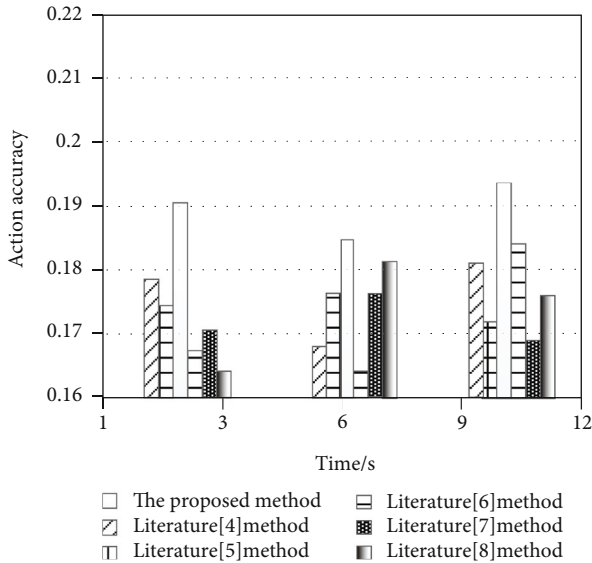


FIGURE 6: Comparison of skiing motion accuracy.

The gradient feature extraction formula [15] is as follows:

$$S(i, j) = \frac{\sum_{i=1}^X \sum_{j=1}^Y \sqrt{(\Delta f_x^2(i, j)^2 + \Delta f_y^2(i, j)^2) / 2}}{X \cdot Y}, \quad (3)$$

where $S(i, j)$ refers to the average gradient of the pixel at the low-frequency coefficient of (i, j) , $\Delta f_x(i, j)$ and $\Delta f_y(i, j)$ are the difference of the image at the regional window's central pixel (i, j) at direction x and y , and X means the number of pixels within the regional window at the direction x and y .

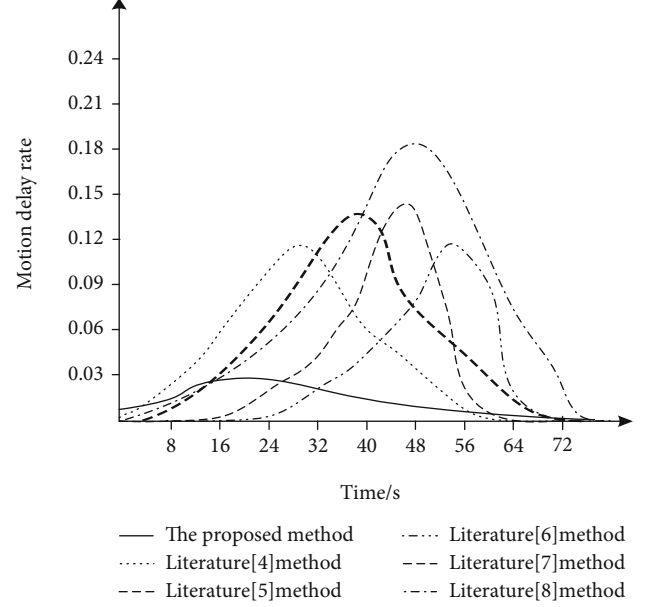


FIGURE 7: Comparison of skiing motion delay rate.

(3) Extraction of signal intensity feature

For the high-frequency coefficients after image decomposition, the signal intensity features are extracted. The extraction formula is as follows:

$$R_{AB} = \frac{G_{j,A}}{G_{j,B}}, \quad (4)$$

where R_{AB} means the intensity ratio of correlation signal, $G_{j,A}$ means signal intensity in TOF depth image area window, and $G_{j,B}$ means signal intensity in image region window of a high-speed vision sensor.

Step 3. Feature integration.

According to the above features, the fusion is carried out, and the fusion formula is as follows:

$$H_{AB}(i, j) = Z_{j,A} \cdot S_{j,A}(i, j) \cdot R_{AB} + Z_{j,B} \cdot S_{j,B}(i, j) \cdot R_{AB}, \quad (5)$$

where $H_{AB}(i, j)$ means the integrated feature of the two kinds of images, $S_{j,A}(i, j)$ means the gradient feature of the TOF depth image, and $S_{j,B}(i, j)$ means the gradient feature of the high-speed visual sensor image [16].

Step 4. Multiscale inverse transform.

The fused high- and low-frequency coefficients are transformed by wavelet multiscale inverse transform to obtain the fused athlete motion image.

3.2. Snowboarding Motion Pose Solution. Based on the fused image, the pose is solved. The solution process is shown in Figure 1. Attitude solution is based on Euler angle.

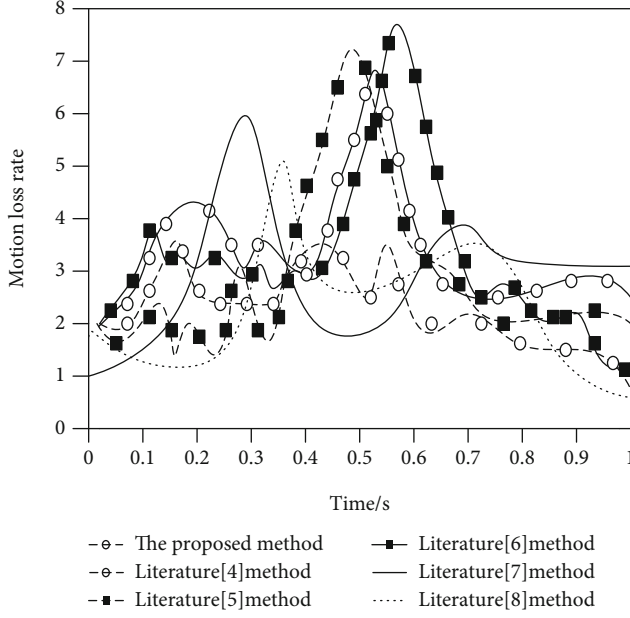


FIGURE 8: Comparison of skiing motion loss rate.

According to Figure 1. The TOF camera image calibrated by the camera and the high-speed vision sensor image are first input into the distortion correction module for image correction, then the image is binarized, the gray value of image pixels is set, and the image with black-and-white effect is obtained. The image data is transmitted to the target recognition module to recognize it, extract the most table and depth value of the target center, then calculate the three-dimensional coordinates and relative distance matrix of the image target center, and finally enter the target matching. When the matching logarithm is greater than 3, the position can be solved. If it is less than 3, the three-dimensional coordinates and relative distance matrix of the image center need to be recalculated. After the image pose is solved, the next image can be solved. Pose solution is the key and core. Pose solution is to calculate the pose data of the target [17, 18]. Suppose the same point f is defined as f in the 3D coordinate of the object coordinate system Oc-XcYcZc, and the 3D coordinate in the camera coordinate system Oo-XoYoZo is defined as g

$$f = Rg + V, \quad (6)$$

where R represents the rotation matrix and V represents the translation matrix.

It is transformed into the following form and then optimized to solve the rotation matrix and translation matrix.

$$f' = \sum_{i=1}^n w_i \|(Rf_i + V) - g_i\|^2, \quad (7)$$

where f' is the optimized solution, f_i is the coordinate value of the target i in Oc-XcYcZc, g_i is the coordinate value of the same target i in Oo-XoYoZo, and w_i is the weight of the target i in calculation [19].

Then calculate

$$\begin{aligned} f &= \frac{\sum_{i=1}^n w_i f_i}{\sum_{i=1}^n w_i}, \\ g &= \frac{\sum_{i=1}^n w_i g_i}{\sum_{i=1}^n w_i}, \end{aligned} \quad (8)$$

where f is the weight center of the target in Oc-XcYcZc, g is the weight center of the target in Oo-XoYoZo, w_i is the weight of target i in calculation, f_i is the coordinate value of the target i in Oc-XcYcZc, g_i is the coordinate value of the same target i in Oo-XoYoZo, i is the initial value of the summation calculation, $i = 1$, and n is the end value of the summation calculation.

Find the center vector of the target in two coordinate systems.

$$\begin{aligned} \alpha_i &= f_i - f, \\ \beta_i &= g_i - g, \end{aligned} \quad (9)$$

where $i = 1, 2, \dots, n$, α_i and β_i are the central vector of the target in Oc-XcYcZc and Oo-XoYoZo, f_i is the coordinate value of the target i in Oc-XcYcZc, f is the weight center of the target in Oc-XcYcZc, g_i is the coordinate value of the same target i in Oo-XoYoZo, and g is the weight center of the target in Oo-XoYoZo.

This forms the center vector matrix $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ and $\beta = (\beta_1, \beta_2, \dots, \beta_n)$. This calculates the scale factor S . The calculation formula is as follows:

$$S = \alpha W \beta^T, \quad (10)$$

where W is the diagonal matrix composed of w_i and T is the transpose symbol. Decompose the S matrix and then get the rotation matrix R and the translation matrix V .

$$\begin{aligned} R &= xy^T \cdot \det(xy^T), \\ V &= \bar{g} - R\bar{f}, \end{aligned} \quad (11)$$

where (x, y) refers to the pixel coordinate, g refers to the weight center of the target in Oo-XoYoZo, f refers to the weight center of the target in Oc-XcYcZc, R is the rotation matrix, V is the translation matrix, xy^T is the rotation matrix of the pixel coordinate (x, y) , and $\det(xy^T)$ is the determinant of the transposed matrix of the pixel coordinate (x, y) .

Finally, the rotation matrix R is transformed into Euler angle form to obtain snowboarding motion data.

3.3. Establishment and Driving of the Virtual Character Model. The establishment of the virtual human model is to establish a moving three-dimensional human model. On this basis, combined with the calculated snowboarding motion data, the synchronous driving of virtual character motion can be realized. The specific process is divided into the following three steps:

Step 1. Establish a human skeleton model. Human movement is mainly driven by bones. The human skeleton structure can be simplified into 16 main bones, and all movement postures can be reflected on these 16 bones, as shown in Figure 2.

Step 2. Skin the bones. In order to ensure the fidelity of the established athlete virtual model, a layer of human-like skin is covered outside the skeleton to improve the three-dimensional of the model.

Step 3. Fit the snowboarding motion data obtained in the previous chapter with the manikin.

Step 4. Judge whether the fitting curve is close to the motion trajectory. If it is close, the output bone changes continuously; otherwise, refit.

Step 5. Drive the virtual model to complete a series of skiing motions.

4. Experimental Analysis and Results

4.1. Experimental Environment and Dataset. PKU MMD is a large long sequence multimodal and multiview dataset released in 2017. The images in the dataset have the characteristics of multiview, including 1076 untrimmed long sequence albums and 20000 trimmed samples; miniImageNet is a subset of ImageNet dataset, which contains more than 100 categories; each category contains 600 pictures. At present, it is widely used in small sample simulation experiments. In this paper, PKU MMD and miniImageNet datasets are selected as the datasets. The training parameters of snowboarder's skiing motion are shown in Table 1.

There are many snowboarding motions, among which the air grabbing is the most typical one. Athletes need to complete technical movements such as turning and somersault with a veneer. This technology has a large range of motion, has high requirements for the mastery of athletes' skill level, and is also one of the necessary items for athletes' training. Based on this, taking the skiing movement as the research object, the motion sequence images are collected by the visual acquisition system, the acquisition frequency is 0.20 frames/s, a total of 100 sequence images are collected, and the time is 20 s.

The TOF camera in the vision acquisition system is the Intel RealSense l515 radar TOF camera depth realistic camera, which can be used under various indoor and outdoor lighting conditions, and the image resolution can reach 1280×720 30 fps, the maximum working range is 0.4~20 m, and the accuracy error is less than 2.7%. At the same time, it is equipped with IMU to realize the automatic monitoring of the moving route of the object. The high-speed vision sensor is imx382, which can detect and track the moving target at the speed of 1000 frames per second, so as to avoid the problem of missing image information due to the moving speed of the target during image capture. The algorithm is used for motion detail comparison, and the process is shown in Figure 3.

The pose solution results are given to the established athlete virtual model to drive the model to move synchronously, and then, the consistency between the athlete's actual motion and the virtual character model motion is compared.

Using the above research method, a series of motion poses in snowboarding are calculated, and the Euler angles of each bone node are obtained, and then, the line diagram is shown in Figure 4.

The standard motion line image in Figure 4 and the motion line image presented by the virtual character model are segmented at equal time intervals and then analyzed according to the experimental indicators.

4.2. Experimental Indicators. There are two main evaluation indexes for skiing motion comparison, and the consistency of motion capture is comprehensively analyzed from these two indexes. These two items are analyzed in detail in the following:

(1) Motion complexity

The more feature points, the more complex the motion is. By comparing the key points between the virtual character model and the standard motion, the execution results of the two groups of skiing motions are judged.

$$X = M_v - C_N(r), \quad (12)$$

where M_v represents the accuracy of motion extraction and $C_N(r)$ represents the accuracy of motion extraction.

(2) Matching proportion

Matching proportion refers to the ratio between the matching results of two groups of motion key points and the number of standard motion features. The greater the ratio, the higher the matching degree.

(3) Motion accuracy: obtain motion accuracy by comparing the number of frames

(4) Motion delay rate: obtain the motion delay of each system by comparing video animation

(5) Motion loss rate: obtain the loss of motion execution by comparing each image

4.3. Results and Discussion. In order to verify the effectiveness of snowboarding athletes' skiing motion real-time capture based on a 3D vision sensor, experimental analysis was carried out. Under the same test conditions, carry out motion capture according to the four methods given in literature [4] to literature [1], and then carry out fine motion comparison, and compare the comparison results with the studied methods. It is shown in Table 2.

According to Table 2, compared with the other five methods, the difference between the key points of 100 frame standard motion sequence images and 100 frame virtual character motion sequence images is the smallest, and the matching proportion with the standard motion sequence image is larger. This shows that the skiing motion displayed by the

virtual character is more consistent with the standard motion, which proves that the motion capture data collected by the research method in this paper is more comprehensive. In addition, from the segmentation results, the error mainly occurs in the third and fourth segments. The motion of these two segments is the most complex. During image acquisition, it is easy to miss the acquisition of image information or collect in advance and delay, resulting in deviation from the standard motion. Finally, the presentation of the virtual model is not completely consistent with the standard motion. Compared with other methods, the accuracy of this method is as high as 98.6%, which is much higher than other systems, and its motion delay and loss are also better than other systems. The system studied in this paper has a better capture effect.

According to Figure 5, the motion matching ratio of literature [4] method is 1.15, the motion matching ratio of literature [5] method is 1.1, the feature extraction stability of literature [6] method is 1.05, the motion matching ratio of literature [7] method is 1, the motion matching ratio of literature [1] method is 0.9, and the motion matching ratio of the method in this paper is 1.2. Therefore, the snowboarding motion matching proportion using the method in this paper is better and more practical.

According to Figure 6, the motion accuracy of literature [4] method is 0.183, the motion accuracy of literature [5] method is 0.178, the motion accuracy of literature [6] method is 0.186, the motion accuracy of literature [7] method is 0.176, the motion accuracy of literature [1] method is 0.186, and the motion accuracy of the method in this paper is 0.195. It can be seen that the snowboarding motion using the method in this paper has high accuracy and improves the capture effect.

According to Figure 7, the motion delay rate of literature [4] method is 0.12, the motion delay rate of literature [5] method is 0.15, the motion delay rate of literature [6] method is 0.10, the motion delay rate of literature [7] method is 0.13, the motion delay rate of literature [1] method is 0.18, and the motion delay rate of the method in this paper is 0.03. Therefore, the snowboarding motion delay rate using the method in this paper is lower.

According to Figure 8, the motion loss rate of literature [4] method is 6.8, the motion loss rate of literature [5] method is 7.2, the motion loss rate of literature [6] method is 7.8, the motion loss rate of literature [7] method is 6, the motion loss rate of literature [1] method is 5.2, and the motion loss rate of the method in this paper is 3.4. Therefore, the snowboarding motion loss rate using the method in this paper is low, and the motion capture is more accurate.

5. Conclusions

Skiing is a standard event in the Winter Olympics. Because its motion is cool and thrilling, skiing attracts many young people to participate. Compared with other sports, skiing is much more difficult and requires higher precision of athletes' motion during training. Therefore, in order to better assist athletes to complete training, this paper studies the real-time capture method of snowboarding athletes' skiing motion based on a 3D vision sensor. By combining two kinds of 3D vision sensors, this study makes up for the prob-

lem of information loss during image acquisition of the existing system and improves the accuracy of motion real-time capture. In the test process, the accuracy of the system studied in this paper can reach more than 98.6%, which is much higher than other systems and is better than other systems in motion accuracy, motion delay, and loss. However, this study still needs to be improved. The motion capture test only takes a skiing motion as the test object, and the results have certain limitations, which need to be further analyzed and discussed.

Data Availability

In this study, we used PKU MMD and miniImageNet datasets for experiments. These two datasets are public which have been deposited in the PKU MMD and miniImageNet dataset website.

Conflicts of Interest

The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work is supported by the key research and development plan of the Ministry of Science and Technology of the Science and Technology Winter Olympics key special targeted projects (2018YFF0300506), General Project of the China University Sports Association (202013508), Heilongjiang Province Higher Education Teaching Reform Key Project (SJGZ20190033), and Harbin Institute of Physical Education Subject Leader Project (XKL04).

References

- [1] Z. Haifeng, Z. Can, L. Meixiao, S. Cuirong, and LP. Yin, "Analysis of motion ability of hip joint in various degrees of freedom based on motion capture technology," *China Tissue Engineering Research*, vol. 25, no. 12, pp. 1815–1819, 2021.
- [2] L. Qingqing and Y. Yuxiu, "Research on the protective performance of basketball knee pads based on three-dimensional motion capture," *Modern Textile Technology*, vol. 27, no. 2, pp. 53–58, 2019.
- [3] Z. Xia, J. Xing, C. Wang, and X. Li, "Gesture recognition algorithm of human motion target based on deep neural network," *Mobile Information Systems*, vol. 2021, Article ID 2621691, 12 pages, 2021.
- [4] Y. Wang, F. Xu, C. M. Pun et al., "New multi-view human motion capture framework," *IET Image Processing*, vol. 14, no. 12, pp. 2668–2674, 2020.
- [5] H. Guo and Y. Sung, "Movement estimation using soft sensors based on bi-LSTM and two-layer LSTM for human motion capture," *Sensors*, vol. 20, no. 6, p. 1801, 2020.
- [6] T. Yu, J. Zhao, Y. Huang, Y. Li, and Y. Liu, "Towards robust and accurate single-view fast human motion capture," *IEEE Access*, vol. 7, no. 99, pp. 85548–85559, 2019.
- [7] T. Maruyama, M. Tada, and H. Toda, "Riding motion capture system using inertial measurement units with contact

- constraints,” *International Journal of Automation Technology*, vol. 13, no. 4, pp. 506–516, 2019.
- [8] Z. Haifeng, Z. Can, L. Meixiao, S. Cuirong, and P. Yin, “Analysis of various degrees of freedom of the hip movement based on motion capture technology,” *Chinese Journal of Tissue Engineering Research*, vol. 25, no. 12, pp. 1815–1819, 2021.
- [9] D. Yan, Z. Deng, and Y. Zhang, “An attitude determination algorithm based on MEMS inertial and magnetic measurement unit,” *Acta Armamentarii*, vol. 40, no. 12, pp. 2447–2456, 2019.
- [10] X. Liu and X. Sun, “Application of Vicon motion capture in the design of linear motion feature tracking algorithm for sequence images,” *Bulletin of Science and Technology*, vol. 36, no. 10, pp. 28–32, 2020.
- [11] W. Xia and S. Zhang, “Design of motion capture system based on inertial sensor,” *Computer Measurement and Control*, vol. 27, no. 9, pp. 283–287, 2019.
- [12] W. Chen, “Research on motion capture technology in dance posture analysis and teaching,” *Microcomputer Application*, vol. 36, no. 3, pp. 102–105, 2020.
- [13] Y. Morimoto, “High-speed measurement of shape and vibration: whole-field systems for motion capture and vibration modal analysis by OPPA method,” *Sensors*, vol. 20, no. 15, p. 4263, 2020.
- [14] D. Huang, C. Yang, Z. Ju, and S. L. Dai, “Disturbance observer enhanced variable gain controller for robot teleoperation with motion capture using wearable armbands,” *Autonomous Robots*, vol. 44, no. 7, pp. 1217–1231, 2020.
- [15] G. Giarmatzis, E. I. Zacharaki, and K. Moustakas, “Real-time prediction of joint forces by motion capture and machine learning,” *Sensors*, vol. 20, no. 23, p. 6933, 2020.
- [16] L. González, J. C. Álvarez, A. M. López, and D. Álvarez, “Met- rological evaluation of human–robot collaborative environments based on optical motion capture systems,” *Sensors*, vol. 21, no. 11, article 3748, 2021.
- [17] J. M. Konrath, A. Karatsidis, H. M. Schepers, G. Bellusci, M. de Zee, and M. Andersen, “Estimation of the knee adduction moment and joint contact force during daily living activities using inertial motion capture,” *Sensors*, vol. 19, no. 7, p. 1681, 2019.
- [18] L. Liu, Y. Liu, and J. Zhang, “Learning-based hand motion capture and understanding in assembly process,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9703–9712, 2019.
- [19] L. Tong, R. Liu, and L. Peng, “LSTM-based lower limbs motion reconstruction using low-dimensional input of inertial motion capture system,” *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3667–3677, 2020.