Research on Human Motion Capture Technology of Depth Map and 3D Model Based on Hybrid Reality

Chen Xiaoman

Xiamen Huaxia University, Fujian, Xiamen, 361024, China

Email: 446917527@qq.com

Keywords: Motion capture, 3-D model, Motion parameters, Data

Abstract: Human motion capture is an important research direction in the field of computer vision. It aims to analyze images or videos to obtain human posture and motion parameters, and to further carry out posture recognition, semantic analysis and behavior understanding. Human motion capture refers to extracting the region containing useful information from the background image from the image sequence to make it a meaningful entity. Three-dimensional human motion capture data is increasingly used in animation, film production, games and other industries due to its strong visual reality and expressive force. Both professional animators and non professionals are inseparable from the browsing, searching and accumulation of materials. Therefore, the retrieval and reuse of large-scale motion capture data need to be supported by efficient, accurate and reliable retrieval methods. Based on the hybrid reality technology, this paper studies the human motion capture technology which combines depth map and 3D model.

1. Introduction

With the development of computer graphics technology and virtual reality technology, the creation of biodynamic virtual human technology in three-dimensional space has been widely used in people's daily life [1]. Visual analysis of human motion is an important research direction in the field of computer vision. It aims to analyze images or videos to obtain human posture and motion parameters, and further carry out posture recognition, semantic analysis and behavior understanding [2]. Computer motion capture is a technology that senses human motion through some sensors and can accurately store and record it. Three-dimensional human motion capture data is increasingly used in animation, film production, games and other industries due to its strong visual reality and expressive force. With the continuous upgrading of computer hardware system and the continuous improvement of artificial intelligence theory, human motion analysis has gained a broader application prospect. With the improvement and continuous development of motion capture equipment, people have gradually accumulated a large number of realistic motion capture data [4]. With the development of motion capture technology, there are many methods to capture motion, and many of them have been applied to practical engineering [5]. In order to query motion accurately and efficiently in the database, people have proposed content-based motion retrieval, which has now developed into an important branch of the field of 3D human animation research.

Human motion capture refers to extracting the area containing useful information from the image sequence in the image sequence, and making it a meaningful entity [6]. Human motion capture can be divided into two broad and narrow senses. The broad sense of motion capture generally includes facial expressions, gestures, and capture of human bones and joints, while the narrow sense of human motion capture refers only to the capture of human bones and joints [7]. The purpose of computer vision is to realize the perception, understanding and interpretation of the scene environment, so as to realize the simulation of human vision by computer [8]. There are many shortcomings in currently widely used motion capture systems. For example, the capture system has various hardware devices that need to be worn by performers. These devices are expensive and complicated, and the environment requirements for performers are also high. Whether it is a professional animator or a

non-professional, the browsing, searching and accumulation of materials are inseparable when creating [9]. Therefore, the retrieval and reuse of large-scale motion capture data need to be supported by efficient, accurate and reliable retrieval methods. Human motion capture technology uses computer vision and digital image processing technology. At the same time, intelligent video related modules are added to the monitoring system based on pattern recognition [10]. Human target tracking is a kind of target state estimation based on the initial state of the captured target and the target model obtained by feature extraction. Based on the hybrid reality technology, this paper studies the human motion capture technology which combines depth map and 3D model.

2. 3D Model Based on Depth Image

At present, 3D models are widely used. In fact, the application of 3D models is much earlier than their popularity on computers. At any given moment, the points on the human target in the image correspond to the points on the real three-dimensional target object one by one. During shooting or transmission, the random interference signal to which the image signal is subjected is image noise. After the spatial domain image is converted into the frequency domain image, if the low frequency part is to be enhanced, it can be realized by smooth filtering [11]. When acquiring a video frame image, the image is inevitably affected by noise and environmental factors, and it is likely that the obtained image is not a qualified registered image. Spatial domain enhancement of video frames performs some processing on each pixel of the image, such as gray scale transformation, sharpening of the image, etc. In a video sequence, if there is motion of an object, the gray value in the motion region will suddenly change. Unqualified original images may lead to excessive error matching. To construct an image motion vector field, each pixel in the image is given a velocity vector. Only by accurately separating the target can the characteristic information of the human target be accurately calculated. The reserved low-frequency components are restored to the spatial domain of the image for subsequent operations.

The target to be tracked is detected in the frame to be tracked, and the image contour is binarized. Finally, the two templates obtained from the first two parts are compared. If the matching is successful, the tracking is successful. The algorithm has the probability of misclassifying the background or part of the background into targets, but the misclassification probability is very small. If the pixel features or pixel areas at corresponding positions have certain differences, it is considered that the pixel points or pixel areas at these positions in the video image of the frame have changed. Fig. 1 is a human body height estimation model using a reference height.

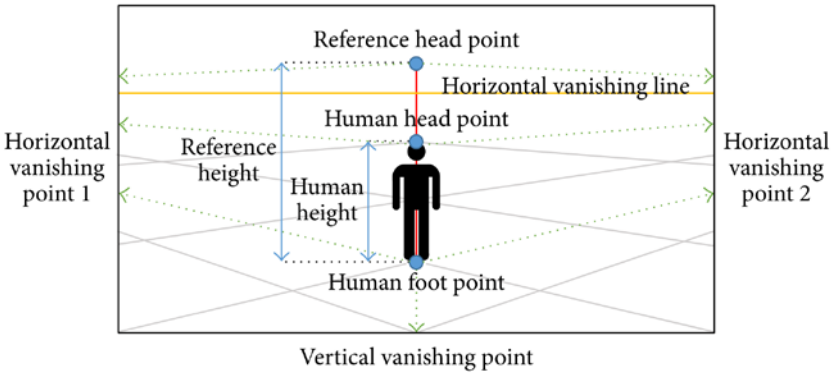

Figure 1 Human height estimation model using reference height

Factor analysis is carried out on the indexes of sports events, and the index system is suitable for factor analysis. The key posture corresponding to the key frame is usually the limit posture in human motion, and it is also the posture with the most abundant content and the most representative meaning. In general, the faster the update speed of the background model, the better the effect order of the obtained motion region. The original monitoring technology used a fixed camera and its viewing angle was limited. The principle of image fidelity is not followed, but information of interest is highlighted and useless information is suppressed [12]. When the speed of the target is too slow, the

frame difference method will get the image of the target over covering and almost completely overlapping, which makes it impossible to detect the human target. We use the context sensitive feature of human body in the video image sequence and the template matching method based on correlation to solve the occlusion problem. There is no general theory of image enhancement technology. The final judgment of the observer determines the quality of an enhancement technology. The difference operation is performed on two or several adjacent images in time in the image sequence. There are different geometric correction methods in different fields. Generally, remote sensing image has large geometric deformation, which can be corrected by using geometric deformation.

3. Human Motion Capture

In order to realize omnidirectional human motion capture, it is necessary to arrange multiple cameras in the human motion capture scene to assist in the work. When the camera shakes, the original stationary target in the background moves or the foreground target becomes part of the background, the original background model is no longer suitable. Therefore, the background model needs to be continuously updated so as to extract better foreground targets. Passive 3D reconstruction method is based on pictures or videos. Usually, a camera is used to capture images or videos, and then the images or videos are processed by algorithms and the 3D coordinates of objects are calculated. In a virtual environment, surface models are usually used to simulate human motion in real time. For a three-dimensional model human motion capture system, there is a high real-time requirement for the human target detection process [13]. If there is no object moving, the gray value of pixels in the adjacent two frames will not change greatly. Frame difference method takes advantage of the continuity between images in video sequence. The real-time representation of human motion is essentially the coordinate transformation of virtual human body model.

In the video sequence, the gray values of the corresponding pixels in the adjacent two frames are compared, and then the difference is found. If there is no human body target in the image, the optical flow vector changes continuously in the whole image area. However, if there is a human target, that is, there is relative motion between the human target and the background image. Then the velocity vector formed by the human target must be different from that of the neighborhood background. As shown in Figure 2, a motion detection and tracking system framework using dense parallax variance technology is proposed.

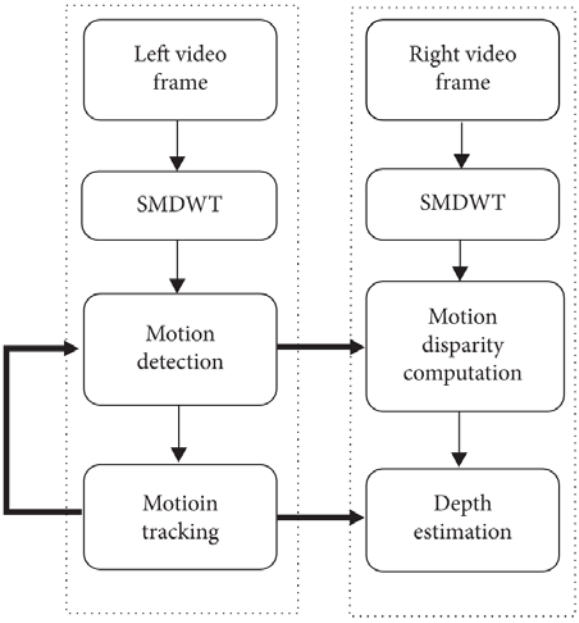

Figure 2 Framework of motion detection and tracking system using dense parallax variance technology

According to the position of the center of mass of the target, it is necessary to control the tripod head to drive the camera to rotate so as to change the field of view and make the target always within the field of view. Table 1 shows the detection rate when the human body is not occluded in human motion capture. Table 2 shows the detection rate under occlusion.

Table 1 Detection rate of unobstructed human body

	Detection rate (%)	Group discrimination rate (%)
Video 1	93.8	81.5
Video 2	92.7	78.7

Table 2 Detection rate of human body being blocked

	Detection rate (%)	Group discrimination rate (%)
Video 1	81.5	76.6
Video 2	77.5	73.63

Because two adjacent frames in a video image have certain continuity, when there is motion of a moving object, the difference of pixel gray values between adjacent frames will increase. The input and output data are randomly selected and submitted to the network. Calculate the output of each neuron in the hidden layer:

$$f(t)=\sum_{j=1}^{N}\sum_{k\in Z}d_k^j\phi_{jk}(t)+\sum_{k\in Z}c_k^N\phi_{Nk}(t) \tag{1}$$

Calculate the response of neurons in the output layer:

$$E_{mi}=\sum_{i=1}^{k}(i\Delta t)\cdot\left|S_{mi}\right|^2 \tag{2}$$

Use the given output data to calculate the error of output layer neurons:

$$\vec{E}=\frac{E_{mi}}{\sqrt{\sum_{i=1}^{k}E_{mi}}} \tag{3}$$

According to the deviation between the target coordinate position and the center of field of view, the PTZ is controlled to drive the camera to rotate to change the monitoring angle of view. Relevant parameters of collinear equation can be obtained through prediction, and the correction number of pixel points can be obtained by using the least square method to achieve the correction purpose. The target template is not fixed. On the contrary, in order to adapt to the changes of targets in different scenes, the template needs to be updated in time. The target is not completely occluded, and some features can be identified to characterize the action of the target, so this algorithm can continue tracking human targets for a long time.

4. Summary

With the development of depth camera, the acquisition of depth images is more convenient. Because depth images contain three-dimensional information and data information is more abundant, motion capture technology based on depth images has been paid attention and studied in recent years. Human motion capture is a key research issue in the field of computer vision, and it is also an active research topic that crosses many disciplines. At present, the difficulty of target tracking lies in the contradiction between real-time and accuracy. If we want to obtain high accuracy at the same time, we must sacrifice real-time performance and reduce the accuracy of applications requiring high real-time performance. Three-dimensional model human motion capture technology has a wide application prospect. It can contain richer and more advanced functions and also bring more and more

complex problems. This paper presents a method of motion capture which combines depth map and 3D model. This method is cheap and convenient, and it paves the way for the further development of motion capture technology. The movement of human body can be varied, and if we want to describe a movement completely, we need to use many images to express it, so the requirement of database is very high. The more motion sequences in the database, the more detailed the description, the more accurate the motion matching recognition.

Acknowledgments

This article is the research result of the young and middle-aged teachers' educational research project "skeleton skin animation based on motion capture data" (project number: JAT60636) of Fujian education department.

References

[1] Tang Ling. Construction of Hierarchical Standing Human Motion Model of Sports Biomechanics [J]. Automation and Instrumentation, 2018 (9): 188-191.

[2] Fu Yanming. Establishment of 3D simulation model of freestyle skiing aerial athletes [J]. Heilongjiang Science and Technology Information, 2014 (23): 116-117.

[3] Zhao Jianjun, Wei Yi, Xia Shihong, et al. Human motion synthesis based on a second-order inverted pendulum [J]. Chinese Journal of Computers, 2014, 37 (10): 2187-2195.

[4] Ke Wende, Peng Zhiping, Wu Chunying. A review of humanoid robot motion planning based on motion similarity [J]. Journal of Guangdong Institute of Petrochemical Technology, 2014 (3): 33-37.

[5] Gu Jie, Guo Zhenxing, Li Enjie. Mechanical analysis of Taijiquan under the action of a three-dimensional rigid body [J]. Journal of Handan University, 2016, 26 (2): 67-87.

[6] Wang Jialei, Liu Qi. Quick construction of human personalized model and bone matching and motion simulation [J]. China Medical Imaging Technology, 2014, 30 (10): 1579-1582.

[7] Duan Shibo. Application of Computer Aided Analysis of 3D Human Motion in Sports [J]. Electronic Test, 2014 (19): 149-150.

[8] Wang Chunjie. Simulation of motion behavior recognition based on 3D virtual incomplete limb images [J]. Computer Simulation, 2014, 31 (7): 399-402.

[9] Ying Rui, Cai Jin, Feng Hui, et al. Human motion recognition based on motion blocks and key frames [J]. Fudan Journal: Natural Science Edition, 2014, 53 (6): 815-822.

[10] Zhang Lei, Shi Xuegang, Lu Linping, et al. Design method of work seat based on reverse human body modeling technology [J]. Mechanical Design, 2014 (7): 113-117.

[11] Wang Xin, Chen Qiudi, Liang Chaokai, et al. Multi-role semantic motion synthesis based on deformable motion model [J]. Computer Science, 2015, 42 (10): 292-296.

[12] Chang Aiduo, Xia Shuhua, Du Xiaozhong. A human joint rotation control model with complementary balance of external force field [J]. Science and Technology Bulletin, 2015 (4): 31-33.

[13] Wang Yan, Yang Gang. Research and Implementation of Collision Interaction between Virtual and Real Objects Using Depth Images [J]. Computer Engineering and Applications, The Only Official Website, 2015, 51 (21): 161-167.