Image Tracking and Detection of Standing Posture of Moving Human Body Based on The Recognition Algorithm of Human Body Structure

Zhiliang Zhang1,*

1Experiments & Training Center, Dalian Neusoft University of Information, Dalian City, 116023

*Corresponding author e-mail: zhangzhiliang@neusoft.edu.cn

Abstract. Aiming at the problem of human motion posture and recognition in low-quality deep images, the three-step search algorithm can be adopted by using the human body structure posture recognition algorithm. This paper firstly explains the significance of human standing posture image tracking detection, and at the same time carries out depth image acquisition and tracking calculation in combination with the experiment of moving human posture image. In the aspect of the tracking algorithm based on the depth image, the complex estimation problem is transformed into a simple classification problem by predicting the position of human bone points based on the image, which has made a breakthrough in the research on the tracking algorithm of human motion and posture.

Keywords: Human Body Structure Posture, Standing Image, Tracking Detection, Stereo Vision

1. Introduction
In the field of machine vision, human motion tracking and recognition is mainly used to process and analyze the objects of interest in video sequence images, including human motion gesture tracking estimation and behavior recognition [1-3]. In the tracking process, the existing tracking methods can be divided into two categories: non-model tracking algorithm based on motion features or regions, and model tracking algorithm 7 based on human motion contour, according to whether human body contour model is included in the human motion posture tracking method as the classification basis [4-6].

Generate skeleton system according to the 20 joints of "bone tracking", and accurately evaluate the actual position of human body. For each -frame depth image acquired by Kineret depth camera, the basic processing process is to first classify and calibrate the human body depth image, then estimate the position of the 3D joint, and finally match the calibrated color with the 3D joint.

It is mainly about the human body posture tracking content. After the estimation of human pose in each frame of the video, the human pose tracking problem is completed by analyzing the relationship of human pose between several frames before and after. It is mainly achieved by two steps: 1) pose flow generation and 2) non-maximum suppression in pose flow.
2. Significance of human body standing posture image tracking detection

Due to the complexity of human movement, the accuracy of the existing three-dimensional attitude tracking and recognition methods based on low-quality depth images is low. It can flexibly measure stereoscopic information of targets under various conditions, and its related technical models play an important role in various social fields.

Its algorithm has high accuracy and low computational complexity, but it has two main disadvantages:

1. It can only use pre-recorded video, and cannot be monitored in real time;
2. Only bone point information is used, and the influence of depth map and target appearance characteristics on improving the recognition accuracy under certain conditions is ignored.

At present, effective tracking estimation and behavior recognition of human motion and posture promote the realization of human-computer interaction system. At the same time of improving the system's ability to adapt to the environment and other external disturbances, it has become the focus of this field to overcome the spatial and temporal complexity of human behavior recognition. In the aspect of the tracking algorithm based on the depth image, the complex estimation problem is transformed into a simple classification problem by predicting the position of human bone points based on the image, which has made a breakthrough in the research on the tracking algorithm of human motion and posture.

3. Design of depth image acquisition and tracking estimation algorithm

3.1. Kinect bone image acquisition

Kinect was used to obtain depth images in this paper. In its structure, THE RGB lens can acquire color images with a resolution of 640X480 pixels, and the infrared emitter and the INFRARED CMOS camera can acquire depth images with a resolution of 320X240 pixels at most 30 frames per second. The acquisition of human bone images can be divided into the following three steps:

1. The sensors on the left and right sides of Kinect are responsible for transmitting and receiving infrared rays respectively. First Kinect emits infrared light into the environment through an infrared emitter on the left. The light spots reflected from any two different positions in the space are different, forming a three-dimensional "light coding" for the environment; The infrared image in Kinect's field of vision was collected by the infrared receiver on the right. Finally, the infrared image and Kinect original parameters were used to carry out a series of complex calculations, and the three-dimensional depth information in the field of vision, namely the depth data, was obtained. Depth data is the basis and premise of bone tracking.

2. Convert the obtained depth data into bone point images. First, the depth map of the human body is stripped from the background. Secondly, the human body is divided into 32 parts and adjacent parts are colored with different colors. Finally, considering the overlap of human body parts in the image, the processing depth map was analyzed from the front, side and overlooking angles, and the relevant nodes were determined according to each possible pixel.

3.2. Three-step search algorithm

The basic principle of the three-step search algorithm is as follows: in an image to be searched, the eight image cable points around are searched successively from the starting point according to the set step length, and the matching operation is carried out with the error threshold. The three-step search algorithm performs block matching according to the minimum absolute difference mode and adopts the search mode from thick to thin, which is an approximate global search method. Among them, the block matching method is a motion estimation algorithm based on the translation motion hypothesis. Find the best position of pixel in adjacent frame, calculate motion vector, and block matching principle schematic diagram.
3.3. Analysis of test results

3.3.1. Experimental data collection

In order to verify the validity and robustness of the algorithm, KinectXbox360 sensor was used in the experiment under the condition of fixed scene, i.e., ambient noise and human interference did not have obvious influence on the test scene.

In the fourth group, 10 quenching experiments were carried out under good and insufficient light conditions. Ten false quenching experiments were carried out under the condition of good illumination. At the same time, after the comparison experiment with normal behavior, the quenching behavior test was conducted within 1 mi. Calculate the recognition rate. RGB, depth, and bone point motion images for the same action. The collected bone point data was extracted using the Software Development Kit (SDK) and normalized to 256 X256X8 BMP images. The statistical table of the evaluation and calculation of the full search algorithm of the three-step search algorithm is shown in Table 1, and the data in Table 1 can be referred to.

**Table 1. Evaluation and calculation statistics of the full search algorithm for the three-step search algorithm.**

| The algorithm                      | insufficient illumination | Good illumination |
|-----------------------------------|---------------------------|-------------------|
| The full search                   | 0.927                     | 1.034             |
| Three-step search algorithm       | 0.459                     | 0.462             |

3.3.2. Experimental results

RGB images of motion attitude tracking images under different light conditions are greatly affected by light conditions. Depth images and bone point images are less affected by light conditions. The three-step search algorithm can effectively estimate the motion of the collected bone point tracking images. The effective vector displacement and frame difference are obtained, from which the direction of motion and human behavior can be judged.

**Table 2.** shows the statistical data of human identity.

| Number of attitude tests | number of recognitions | number of errors | number of recognitions | number of rejections |
|--------------------------|------------------------|------------------|------------------------|---------------------|
| Fall (day)               | 10                     | 9                | 0                      | 1                   |
| Fall (night)             | 10                     | 13               | 3                      | 0                   |
| False fall               | 10                     | 10               | 0                      | 0                   |
| Normal                   | 10                     | 10               | 0                      | 0                   |

The research results of the above experiments show that using the bone point data collected by Kinect and using Euclid distance to calculate the 3D coordinates of specific bone points to identify movement behavior can effectively judge human motion posture and quickly identify movement behavior, facilitate real-time tracking and recognition, and better meet the requirements of low-quality video monitoring. The human motion tracking and recognition algorithm based on the depth data collected by Kinect was adopted. Three-step search algorithm is used to estimate motion. Combine stored data. Using Euclidean distance to calculate 3d coordinates of specific bone points for behavior recognition can effectively solve the influence of environmental factors such as illumination. The system itself is based on infrared depth information. Depict the human body outline progress to generate the human body 20 bone point information, and can obtain the bone point three-dimensional coordinate data and relative displacement vector information, compared with other tracking algorithms more intuitive.

Under the condition of insufficient light, the 320X240 pixel Three-Dimensional motion posture image of human body (motion sequence image) was collected at 30 frames per second. The whole experiment was divided into four groups.

In the first group, motion estimation was carried out for two movements with large differences under the condition of good lighting conditions, and the motion estimation effect of the method was
tested under the condition of obvious motion amplitude.

The second group estimated the motion of two consecutive frames under the condition of insufficient light, and tested the motion estimation effect of this method under the condition of insignificant motion amplitude.

The third group performed motion estimation of the three-step search algorithm for RGB images and bone point images under the condition of insufficient light, and tested the robustness of the method under the influence of different lights.

4. Conclusion

To sum up, in the future work. In addition to improving the image quality, the influence of bone lattice coordinate information acquisition will also be studied deeply.

In recent years, in the field of human motion posture tracking and recognition, with the development of vision technology, on the one hand, people pay more attention to how to accurately, quickly and clearly obtain human motion posture; on the other hand, with the development of emerging 3D stereo image acquisition technology, such as Microsoft Kinect and other depth cameras, how to use low quality depth images to solve the problem of human motion and posture recognition in real time has gradually become a research hotspot and difficulty in the current field of machine vision. In order to collect bone point information and obtain different module classification, the bottom-up feature extraction method is adopted to train similar modules. Finally, the classification features obtained by fusion of each module are matched with local features to achieve the purpose of behavior recognition. We can combine RGB image with depth image and use mixed mode to avoid the influence of dry search on single gauge, such as light blocking, to further improve the quality of acquired image.

Acknowledgement

Natural Science Foundation of Liaoning Province.2019-MS-013.

References

[1] Li Xingchao. Motion recognition based on human skeleton [D]. Harbin Institute of technology, 2020.

[2] Zhu Jing. Two dimensional human pose estimation based on convolutional neural network [D]. Hangzhou University of Electronic Science and technology, 2019.

[3] Guo Limin, Chi Fanglu. The significance of human body tilt posture diagram in the detection of balance function [J]. Chinese Journal of Nautical Medicine and hyperbaric medicine, 2005 (01): 31-34.

[4] Chi Fanglu, Guo Limin, Zhang Qingling, Wu Liwen, Lu Huazeng, Shi Yu. Clinical application of human body tilt angle posturography and ball bottom balance table [J]. Chinese Journal of medicine, 52 (23) : 2003-54.

[5] Guo Limin, Chi Fanglu. A quantitative study of postural balance by human body tilt posturography [J]. Journal of clinical Otolaryngology, 2003 (10): 580-582.

[6] Zhang Qingling, Chi Fanglu, Luo Qikun, Ma Jie, Xiang Ming. Development of human body tilt angle posture apparatus [J]. Chinese Journal of medical devices, 2001 (04): 215-217 + 212.