2019 2nd International Conference on Intelligent Systems Research and Mechatronics Engineering (ISRME 2019)

# The Construction of Security Warning Zone Based on Kinect

**Yangwei Cheng[1, a,*], Ying Sun[1, 2, b], and Yao Liu**

[1]Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China.

[2]Hubei Key Laboratory of Metallurgical Transmission and Manufacturing Engineer, Wuhan University of Science and Technology, Wuhan 430081, China.

[a]970619543@qq.com, [b]sunying65@wust.edu.cn

**Keywords:** Kinect Sensor; Human Skeleton; Safety Warning Zone

**Abstract:** With the development of the robot technology, industrial robots have been widely used in various factories. In view of the fact that robots can not guarantee human safety while working, a safety alert zone based on Kinect sensor is proposed. This method uses the RGB and 3D images collected by Kinect sensor to measure the distance and set up a security alert area. Among them, it is very difficult to identify the whole human body and the recognition rate is very low. By recognizing the human skeleton, people who will enter this safe area will be positioned safely in three-dimensional space. Finally, the method is validated by experiments. The experimental results show that the method can be used for effective safety monitoring.

## 1. Introduction

With the continuous development of modern science and technology, robotics technology is becoming more and more mature, and robots can be widely used in life. Robots have also become the most effective tool for human beings to explore space and the ocean. In the enterprise, in order to obtain higher profits, the application of robotics technology to industry can not only ensure the quality of products, but also double the production rate. In recent years, many scholars have done in-depth research on the recognition of the working environment of robots, for example, Zheng Weicai, Yu Zhenzhong and Hui Jing[2] uses Kinect's built-in CMOS infrared camera and optical coding technology to obtain information about the surrounding environment, so as to plan the path of mobile robots. Zhang Bo[3] uses Kinect sensor to process the captured image and video information and depth data, so as to identify known and unknown targets in the environment. But these methods have low recognition rate for human beings. To overcome this shortcoming, Kinect sensor is mainly used to extract the dynamic environment around industrial robots, as a visual sensor of robotic hand to set up a security alert area, to identify the skeleton of people entering this area, and finally to issue an alarm. This method guarantees the safety of the workers around the industrial robot when it works.

## 2. Kinect's Ranging Method

Microsoft released a sensory peripheral called Kinect in June 2010. It was designed for XBOX360. The device contains Microsoft's human skeleton recognition library, which can recognize human body and capture human movements. Kinect can also be used as a 3D camera. It can provide us with visual images and depth images at low cost. Because of this, many scholars have used Kinect in the field of robotics and achieved good results.

Light coding is a distance measurement technology, which uses light source to digitally encode the space to be measured. Kinect ranging uses its own laser speckle as a light source. When laser irradiates an object, it will form a diffraction spot. The advantage of these diffraction spots is that they are highly random, and they produce different patterns with distance. After irradiating the space to be measured by laser, the pattern of any two places in the space is different, and the whole

  DOI: 10.25236/isrme.2019.035

space is marked when the structured light is spread over the whole space to be measured. Placing an object in this space can identify its position accurately through the speckle pattern. Of course, in order to record the speckle pattern of the whole space, it is necessary to calibrate the light source first. The light source is calibrated as follows: we set up a reference plane at intervals, and then record the speckle pattern on the reference plane. For example, within the range of 1-3 m, 30 speckle images can be obtained by selecting a reference plane every 10 cm. In the measurements, the 30 images were cross-correlated to obtain 30 correlation images. If an object exists at a certain position in space, the peak values will be displayed in the relevant image, and then the three-dimensional shape of the object can be obtained by stacking these peaks layer by layer and interpolating them.

## 3. Kinect Sensor Monitoring Environment

### 3.1. Deep Camera Calibration.

The so-called camera calibration is to correlate the pixels in the image with the physical coordinate system in real life, so as to get the mapping from two-dimensional image to three-dimensional space and then use this mapping relationship to measure and locate the detected objects. Since the imaging model of depth camera is similar to that of pinhole camera, assuming that the image pixel coordinate system is $O_P X_P Y_P$ and the camera coordinate system is B, then using the sensor imaging model, we can calculate the three-dimensional coordinate of the image center position $O_V X_V Y_V Z_V$ by deviating the pixel point $(u,v)$ from the image center position $(u_0,v_0)$.

$$X_V = |u-u_0| \frac{depth(u,v)}{f_x} \tag{1}$$

$$Y_V = |v-v_0| \frac{depth(u,v)}{f_y} \tag{2}$$

$$Z_V = depth(u,v) \tag{3}$$

Where $f_x$, $f_y$ are internal parameters of the camera in the X and Y directions. $X_V$, $Y_V$ and $Z_V$ are the offset distances of pixels from the image center in different directions. $depth(u,v)$ is $(u,v)$ spatial depth distance value

### 3.2. Structure of Skeletal Tracking Data.

Skeletal tracking technology is mainly used to calibrate 20 special points of human body, and to track these 20 special points in real time in the monitoring environment. In the Kinect for Windows SDK, there is an API for recognizing bones, through which at most two people can be identified in their location information, and their gestures and three-dimensional coordinates of bones are also included. In addition, the Kinect for Windows SDK can provide up to 20 skeletal points of the human body. What Kinect observes is that people are represented as skeletal frames, with 20 skeletal points on each frame, as shown in Fig.1.

In SDK, each skeleton point is defined by Joint type, and 20 skeleton points contained in each frame are combined to form a set based on Joint type. Joint type includes three types: Joint Type, Position and Trajectory State.

(1) Joint Type: An enumeration type, it mainly expresses the type of 20 skeletal points with specific names. For example, "WRIST_RIGHT" represents the right wrist of the human body.

(2) Position: The Skeleton Point type is used to define the location information of each skeleton point. SkeletonPointly uses three data information (X, Y and Z) to store the location information of skeletal points.

(3) Tracking State: It is also an enumeration type, this type mainly captures skeletal points and

tracks them in real time. Tracked represents the points that have been captured and tracked in real time, Inferred represents the unknown state, and Not Tracked indicates that no information about the points has been obtained.

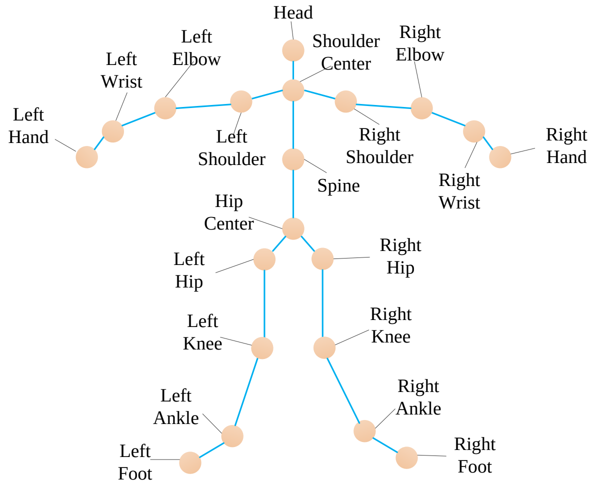


Figure 1.Human Skeleton Frame Form

### 3.3. Method of Obtaining Bone Tracking Data.

The way to get the skeleton data of the next frame is mainly to call the callback function Open Skeleton Frame () and pass the cached skeleton frame below, which is the same way to get the color image and the depth image. If the latest skeletal data is ready, it will be copied into the cache. However, if the new skeletal data is not ready when the application makes the request, you can choose to wait for the next skeletal data until it is ready, or return immediately to send the request later. For the NUI Skeletal API, the same skeletal data will be provided only once. Skeletal API can provide two modes: Polling Mode and Time Mode.

(1) Polling Mode is a very simple way to get the next skeleton data, mainly by calling OpenNextFrame () function to set the waiting time. When the new data is ready or exceeds this waiting time, the function will return the next skeleton data.

(2) Time Model is a more convenient, flexible and accurate method, which mainly obtains skeletal data by event-driven method. The principle of this model is to use the application program to pass the processing function to Skeleton Frame Ready in Kinect Sensor class. When the next skeleton data is ready, the application immediately calls back function to get the skeleton data in real time.

The results are displayed by running the program as shown in Fig.2.

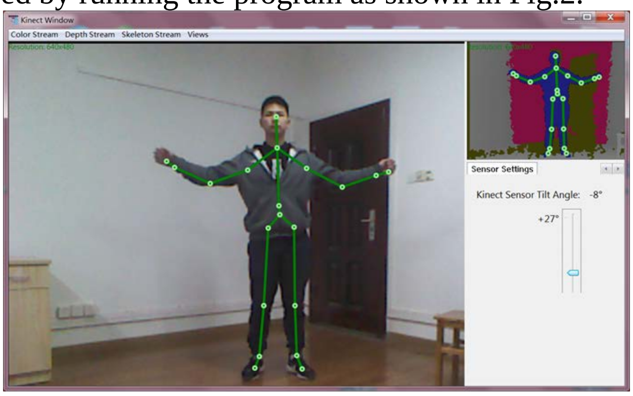


Figure 2. Kinect Recognition Skeleton

## 4. Experiment

In order to verify the feasibility of this method, relevant experiments are carried out to verify it. The color image and depth image collected by Kinect sensor are used to express the depth information with the depth of color. When someone enters the safety zone, the collected images are shown in Fig.3.

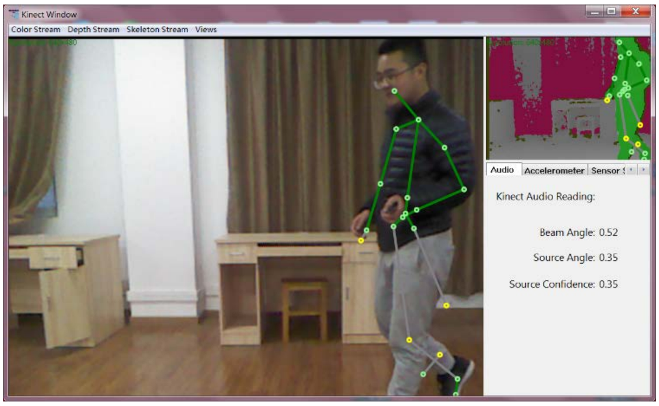


Figure 3.Identify People Entering The Security Zone

From Fig.3, it can be found that the recognition effect of human skeleton is very good. Compared with Fig.2, the farther the distance is, the bigger the pixel value is, the darker the color is, on the contrary, the lighter the color is. When human skeleton is reflected, the signal of this characteristic is transmitted to the signal processing circuit, and the alarm reacts immediately. The experiment shows that the safety monitoring method can achieve good results.

## 5. Conclusions

In view of the fact that industrial robots in factories can not guarantee human safety while working, a safety alert zone based on Kinect sensor is proposed. The image and video information and depth data collected by Kinect sensor will enter the environment, and the skeleton will be reflected by skeleton tracking technology. Finally, the experimental verification of this method is carried out. When someone enters the security alert area, the features of the bone image collected by Kinect sensor are transmitted to the signal processing circuit, and the alarm can react immediately. The validation results show that this method can effectively monitor security.

## References

[1] Y. Sun, C.Q. Li, G.Z. Jiang, D. Jiang, H.H. Liu, Z.G. Zheng, W.N and Shu: Mobile Networks & Applications, Vol. 23 (2018) No.4, p.797.

[2] B. Zhang: Electronic Science and Technology, Vol. 31, (2018) No.7, p.89.

[3] G.G. Zhu and L. Cao: Computer Simulation, Vol. 31, (2014) No.12, p.329.

[4] Y. Duan, D.L. Sheng, X. Yu: Journal of Shenyang University of Technology, Vol. 39, (2017) No.2, p.170.

[5] W.C. Zheng, Z.Z. Yu and J. Hui: Journal of Jiangnan University (Natural Science Edition), Vol. 14, (2015) No.2, p.156.

[6] S.L. Wang and W.S. Xu: Computer Engineering and Applications, (2016), No.53, p.169.

[7] H.B. Lin, W.L. Mei, Y. Zhang and Y. Luo: Computer Applications and Software, Vol. 30, (2013), No.2, p.157.

[8] R. Gu, Y.M. He and L.Y. Jin: Computer Applications and Software, Vol. 30, (2013), No.2, p.276.

[9] Q. Wu, Z. Chen, Y.X. Zhou and J.Z. Liang: Journal of ChangZhou University (Natural Science Edition), Vol. 26, (2014), No.3, p.6.

[10] Z.D. Chen, J.M. Yao and T.L. Guo: Cable TV Technology, (2018), No.1, p.28.