

Research Article

Human Position Detection Based on Depth Camera Image Information in Mechanical Safety

Cheng Zhou, Dacong Ren, Xiangyan Zhang, Cungui Yu, and Likai Ju 

School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Correspondence should be addressed to Likai Ju; jwcjlk@njust.edu.cn

Received 11 November 2021; Accepted 14 December 2021; Published 13 January 2022

Academic Editor: Miaocho Chen

Copyright © 2022 Cheng Zhou 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.

The devices used for human position detection in mechanical safety mainly include safety light curtain, safety laser scanner, safety pad, and vision system. However, these devices may be bypassed when used, and human or equipment cannot be distinguished. To solve this problem, a depth camera is proposed as a human position detection device in mechanical safety. The process of human position detection based on depth camera image information is given; it mainly includes image information acquisition, human presence detection, and distance measurement. Meanwhile, a human position detection method based on Intel RealSense depth camera and MobileNet-SSD algorithm is proposed and applied to robot safety protection. The result shows that the image information collected by the depth camera can detect the human position in real time, which can replace the existing mechanical safety human position detection device. At the same time, the depth camera can detect only human but not mobile devices and realize the separation and early warning of people and mobile devices.

1. Introduction

Mechanical safety is the state and condition that the human is protected from external factors under various states of using machinery from human needs. Aiming at achieving mechanical safety in the mechanical design stage and use stage, three steps are mainly adopted: intrinsic safety measures, safety protection and supplementary protection measures, and risk reduction of use information [1]. The purpose of using safety protection and supplementary protection devices is to prevent moving parts from causing danger to human. When using safety protection devices related to the approach speed of human parts, first, we need to accurately detect the presence or position of human. At present, the detection devices of human existence in mechanical safety mainly include safety light curtain, safety laser scanner, safety pad, and vision system. When using these devices, the orientation, angle, height, and other factors of the detection area and the possibility of bypassing need to be considered. For example, when using a safety light curtain, a human may climb in from the lowest beam, cross over from the highest beam, cross between two beams,

etc. [2]. In addition, most of these devices belong to hardware devices and have no function of object recognition. The detection of human existence is artificially specified, that is, only a human appears in the detection area. However, vision-based object detection can make up for the above shortcomings, using the camera to collect images and combined with the object detection algorithm for accurate human detection.

Intelligent plant pays more attention to the automation and flexibility of product manufacturing and use a large number of mobile devices such as industrial robots and AGV. In order not to affect the passage of AGV and other mobile devices, it is necessary to remove the safety fence and set a certain dangerous area through the safety laser scanner. When a human enters the dangerous area, the robot slows down or stops to protect human safety. For AGV and other mobile devices, the moving path is planned in advance, which generally will not affect the work of the robot. However, in addition to detecting human, the laser scanner also detects AGV and other mobile devices, resulting in triggering the deceleration or stop of the robot and affecting the work efficiency.

Scholars have carried out research on human position detection based on depth camera, because in many detection scenarios, not only accurate human detection but also human position measurement is required. Compared with ordinary cameras, the biggest feature of depth camera is that it can collect depth information for position measurement. At the same time, it also has functions such as voice recognition, gesture recognition, and facial expression recognition. Moreover, compared with some hardware devices that can realize position measurement, it can realize accurate detection of human based on image information. At present, the widely used depth cameras mainly include Intel RealSense and Microsoft Kinect. These cameras have the characteristics of small overall size, strong environmental perception, image acquisition accuracy, and low price. Mathé et al. [3] determined the position between the surgeon and the robot through Kinect to prevent the robot from interfering with the doctor's work. Tupper and Green [4] used RealSense combined with the Mask R-CNN [5] algorithm to determine the relative position relationship between pedestrians and cameras to realize pedestrian proximity detection. Jian et al. [6] realized human recognition and position measurement through depth camera combined with cascade classifier AdaBoost and RGB-D image. Based on the 3D skeleton information obtained by Kinect, Li et al. [7] registered with the SMPL model to obtain the real posture of a human. Yu et al. [8] used deep camera Kinect to detect human position in the study of robot automatic avoidance of pedestrians. Li et al. [9] proposed a human posture tracking system based on dual Kinect, which can determine the human position by obtaining accurate and stable joint position trajectory. To sum up, the depth camera can accurately detect the position of human within a certain distance and range and can be applied to the detection of human existence or position in mechanical safety. This paper presents the human position detection process based on depth camera and takes Intel RealSense depth camera combined with MobileNet-SSD [10] algorithm as an example to give the human position detection method, which is applied to robot safety protection.

2. Human Position Detection Method Based on Depth Camera Image Information

2.1. Human Position Detection Process. The human position detection process based on depth camera image information is shown in Figure 1, which mainly includes image information acquisition, human presence detection, distance measurement, and other links. Firstly, human image information is collected by depth camera. Secondly, the human presence detection is realized according to the collected image information and the human detection algorithm. Finally, the distance is measured through the detection information and the depth information provided by the camera to determine the human position.

2.1.1. Image Information Acquisition. The depth camera generally includes RGB camera and infrared laser emission module. Before image acquisition, it is necessary to establish

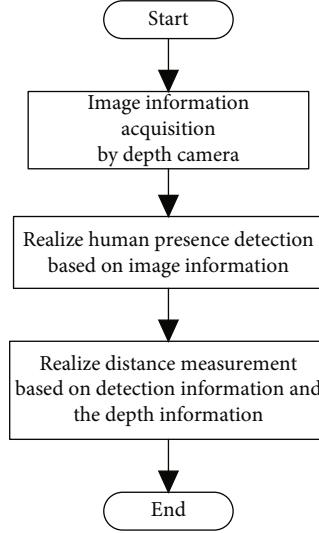


FIGURE 1: Human position detection process based on depth camera image information.

the communication between the vision software library and the camera, and then, the vision software library drives the camera through the program to realize the acquisition of color image, depth image and infrared data. And the collected images are converted into corresponding image data for follow-up work.

2.1.2. Human Presence Detection. Human presence detection belongs to the category of object detection. Object detection mainly includes machine learning and deep learning algorithms. The machine learning detection algorithm firstly selects the detection region based on the sliding window traversal and then extracts the characteristics of the image in the sliding window, such as HOG (histogram of oriented gradient), Haar, and LBP (local binary patterns); finally, SVM (support vector machine), AdaBoost, and other classifiers are used to classify the extracted features to realize human presence detection [11–12]. The deep learning detection algorithm realizes object detection through the self-learning characteristics of multilayer convolutional neural network. The detection accuracy and speed are significantly improved compared with the machine learning detection algorithm [13–15]. At present, human presence detection algorithms based on deep learning mainly include Faster R-CNN [16], SSD [17], and YOLO [18]. In addition, based on the consideration of industrial scenes related to mechanical safety, the human detection algorithm used in this paper must meet the requirements of real-time detection and ensure a certain detection accuracy.

2.1.3. Distance Measurement. Depth camera distance measurement mainly includes binocular stereo vision, structured-light, and time of flight (TOF) method. The camera depth using binocular stereo vision does not need to project external active light source. It obtains images of the object measured from different positions by binocular camera, calculates the position deviation (i.e., parallax) between the

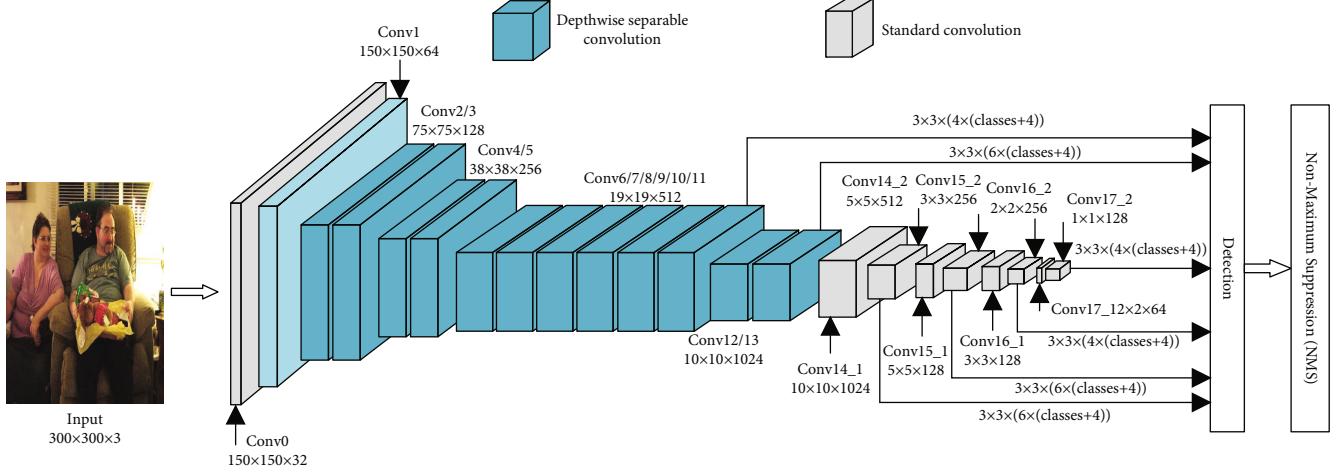


FIGURE 2: Network structure of MobileNet-SSD.

corresponding pixels of the two images and obtains three-dimensional information of the object. Reference [19] combines object detection and binocular distance measurement to detect and measure the distance of the front object of the engineering vehicle, so that the engineering vehicle can work safely and independently. Structural-light method uses a specific wavelength of invisible infrared light as the light source to illuminate the object and then obtains the position and depth information of the object according to the returned optical distortion image. Reference [20] bases on the structured-light; a high-performance, small volume, and modular structural-light 3D camera is designed, which can directly obtain 3D data. TOF is to continuously transmit light pulses to the observed object, then receive the light pulses reflected from the object, and calculate the distance between the measured object and the camera by detecting the flight (round trip) time of the light pulses. In reference [21], TOF camera is used to provide the distance information and three-dimensional coordinates of the object in real time, and the geometric structure of the three-dimensional object is reconstructed based on the distance information and camera parameters.

3. Human Position Detection Method Based on RealSense Image Information

Mechanical safety requires high real-time detection of human position. It is necessary to quickly detect the human position under the condition of ensuring a certain accuracy. Based on the measurement of accuracy and speed, this paper selected RealSense depth camera and MobileNet-SSD algorithm to realize human position detection. Firstly, the real-time color image and depth image of human are obtained by using RealSense depth camera. Then, the MobileNet-SSD algorithm is used to detect the human according to the color image. Finally, the pixel value corresponding to the depth image is obtained according to the detected pixel position information, and the distance between the human and the camera is calculated to determine the human position.

3.1. RealSense Image Information Acquisition. The steps of acquiring color image and depth image with RealSense depth camera are as follows:

- (1) Declare RealSense pipe stream object
- (2) Create a configuration object, define the image pixel size, and specify the number of frames read by the camera and the type of image collected
- (3) Use the pipe stream object to open the configuration and start cyclic reading of video frames
- (4) Create an alignment object so that the depth aligns the colors
- (5) Read video frames, align processing video frames, and get aligned depth images and color images
- (6) Get camera built-in parameters from color images
- (7) Data format conversion of color images for human existence detection

3.2. Human Existence Detection Based on MobileNet-SSD. MobileNet SSD algorithm changes the backbone network of the original SSD algorithm from VGG to MobileNet network. The network structure is shown in Figure 2. Based on streamlined architecture, MobileNet uses depthwise separable convolutions instead of standard convolutions to build lightweight deep neural networks. Depthwise separable convolutions decomposes the standard convolution into depthwise convolution, and 1×1 pointwise convolution plays the role of filtering and linear combination, respectively, while reducing the amount of parameters and calculation. The network detection speed has been greatly improved and is suitable for mobile terminals.

Table 1 shows the detection performance of common target detection algorithms on the general dataset VOC2012, FPS (frame per second) in the table is used to measure the detection speed of the algorithm, and mAP (mean average precision) is used to measure the detection

TABLE 1: Performance of common target detection algorithms on VOC2012 dataset.

Object detector	Backbone	FPS	mAP (%)
Faster R-CNN	VGG16	5	70.4
YOLO	GoogleNet	45	57.9
SSD	VGG16	46	75.8
MobileNet-SSD	MobileNet	>46	72.7
YOLOv2	DarkNet19	40	73.4

accuracy of the algorithm; the larger the FPS, the faster the detection speed, and the larger the mAP, the more accurate the detection. From the table, we can find that the algorithms of MobileNet-SSD not only has the absolute advantage of detection speed but also has good detection accuracy, and the mAP is as high as 72.7%, which fully meets the needs of human presence detection.

In this paper, OpenCV [22] (an open source computer vision and machine learning software library) is used to load MobileNet-SSD model to realize human detection. OpenCV integrates a module called DNN, which is specially used to realize the related functions of deep neural network. When OpenCV loads the relevant detection model, the DNN module will rewrite the detection model to make the operation efficiency of the model higher.

The MobileNet-SSD model loaded by OpenCV requires two model files; one is the binary description file and the other is the model text file. The general process of OpenCV loading MobileNet-SSD model to realize target detection is shown in Figure 3, and the specific steps are as follows.

- (1) Load the MobileNet-SSD model. Using the “dnn.readNetFromCaffe()” method to load the model. The two parameters of the method are the text file and description file of the model, respectively
- (2) Read the image and format the image data. Using the “dnn.blobFromImage()” method to achieve image format conversion, the converted data can be used for network broadcast, that is, it can be used by the loaded model
- (3) Use the image as the input of the model. Using the “dnn.setInput()” method to load the image
- (4) Forward broadcast. Using the “dnn.forward()” method to complete forward broadcast, that is, model prediction. The predicted result is a four-dimensional matrix, focusing on the data of the third and fourth dimensions. The third dimension is the detected target, and the fourth dimension is the detected information of each target. The information mainly includes the target category number, confidence level, and object location
- (5) Traverse all the predicted results to judge whether the confidence of each target is greater than the given confidence threshold. If it is greater than the given threshold, it is considered that the prediction is

correct, and the target position and target category are drawn on the original image. If it is less than the given threshold, it is considered that the prediction is wrong, and continue to traverse

The human body detection effect of MobileNet-SSD is shown in Figure 4. It can be seen from the figure, MobileNet-SSD has a general effect on small target detection, but for the scene proposed in the paper, the human body belongs to large targets, so it does not affect the overall detection effect. In addition, the test video is used to measure the MobileNet-SSD, and the measured detection speed is about 25 FPS; the reason for the large difference from the data in Table 1 is that the performance of the computer and the output frame rate of the video will have a great impact on the detection speed, but it is clear that MobileNet-SSD has good timeliness; because for video processing, it is generally considered that 12.5 FPS is real time. The measured results are shown in Figure 5.

3.3. Human Distance Measurement. RealSense depth camera is mainly composed of left camera, right camera, infrared projector, and RGB camera, as shown in Figure 6. RealSense uses binocular stereo vision to measure distance [23–25], which is shown in Figure 7. Camera L and Camera R are left and right cameras, respectively. Image planes are the imaging planes of two cameras, which are located in front of the camera plane and parallel to the camera plane. Baseline represents the camera baseline; P' and P'' represent the two projections of the space point P to be measured on the imaging plane. Based on the principle of similar triangles, distance measurement is calculated as follows:

$$Z = \frac{fb}{x_L - x_R}, \quad (1)$$

where Z is the measured distance, mm; f is the camera focus, mm; b is the center distance of the left and right cameras, mm; x_L is the x coordinate of point P' on the image plane, mm; and x_R is the x coordinate of point P'' on the image plane, mm.

4. Application Case of Robot Safety Protection

Figure 8 is a schematic diagram of the mechanical safety protection system of a robot in a production line, Figure 9 is a physical image, and Figure 10 is a mechanical safety protection monitoring system of a robot for real-time monitoring of the system status.

In order not to affect the passage of AGV, the traditional safety fence is removed. According to the distance between a human and machine, a hierarchical early warning system is constructed by using laser scanner. The range of human activities is divided into four areas: early warning area I, early warning area II, early warning area III, and dangerous area. Protect human safety by projecting light subtitles, broadcasting warning voice, robot deceleration or stop, etc. When people intrude into the early warning area I, the light projection subtitle “Early warning area I” appears on the

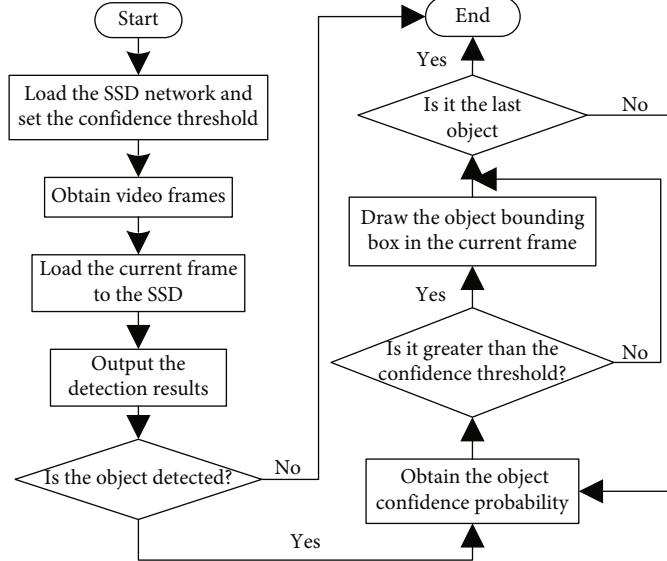


FIGURE 3: Human detection process based on MobileNet-SSD.

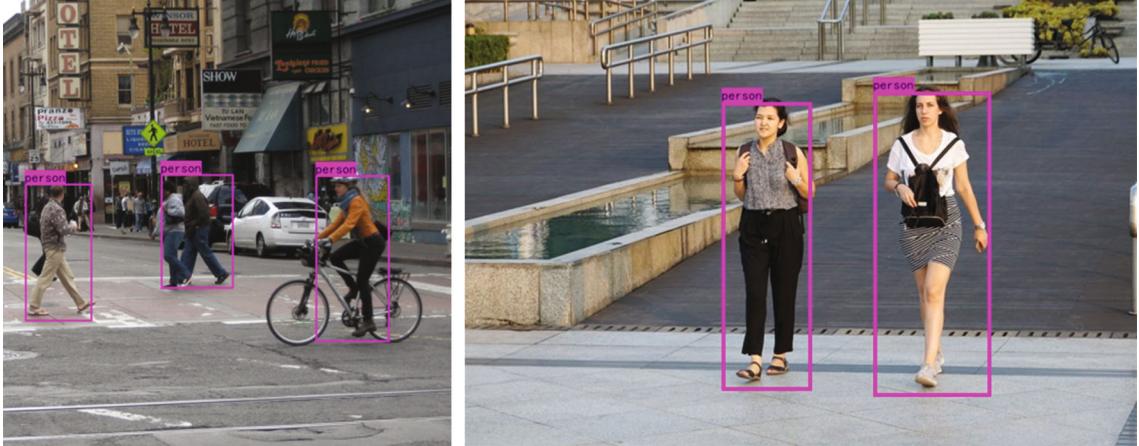


FIGURE 4: Human body detection effect display.

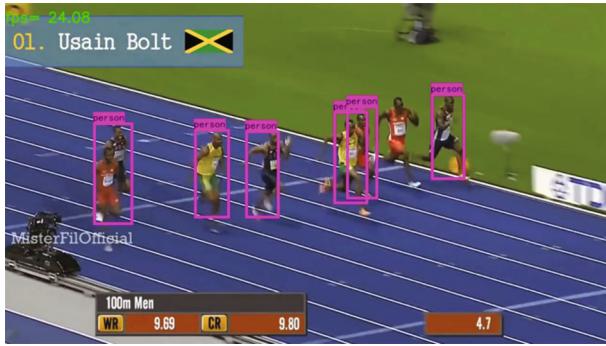


FIGURE 5: Video measurement effect display.

ground, and the voice broadcast “You have entered the early warning area I”, as shown in Figure 11(a). When people invade the early warning area II, the light projection subtitle “Early warning area II” appears on the ground, the voice broadcast “You have entered the early warning area II”,

and the speed of 1 axis of the robot automatically decreases to 70% of the working speed. The monitoring system is shown in Figure 11(b). When people invade the early warning area III, the light projection caption “Early warning area III” appears on the ground, the voice broadcast “You have entered the early warning area III”, and the speed of 1 axis of the robot automatically decreases to 20% of the working speed. The monitoring system is shown in Figure 11(c). When people invade the dangerous area, the robot stops, and the monitoring system is shown in Figure 11(d). Therefore, the hierarchical mechanical safety protection system can realize early risk early warning under the condition of ensuring human safety, significantly reduce the number of downtime, and improve the working efficiency of the machine.

However, the AGV frequently enters the early warning area III and early warning area II in the working process. Because the laser scanner cannot recognize that the object entering the early warning area III and early warning area II is human or AGV, the robot is frequently in a deceleration

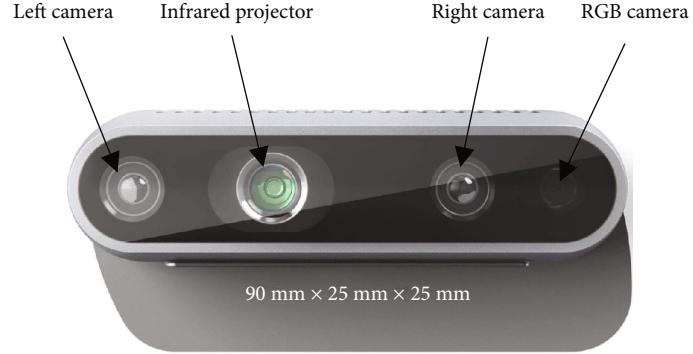


FIGURE 6: RealSense depth camera.

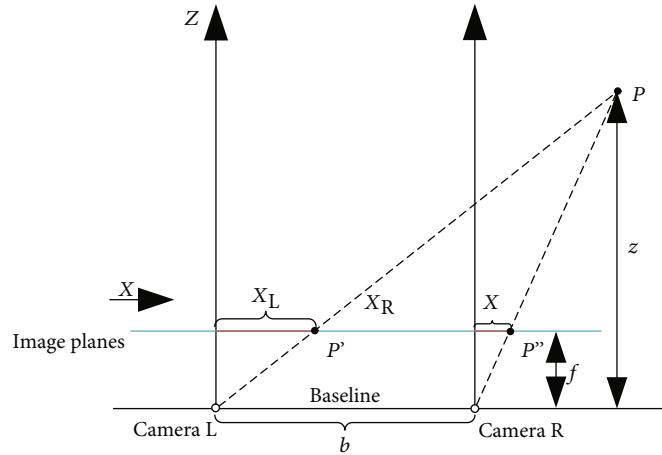


FIGURE 7: Principle of binocular stereo vision ranging.

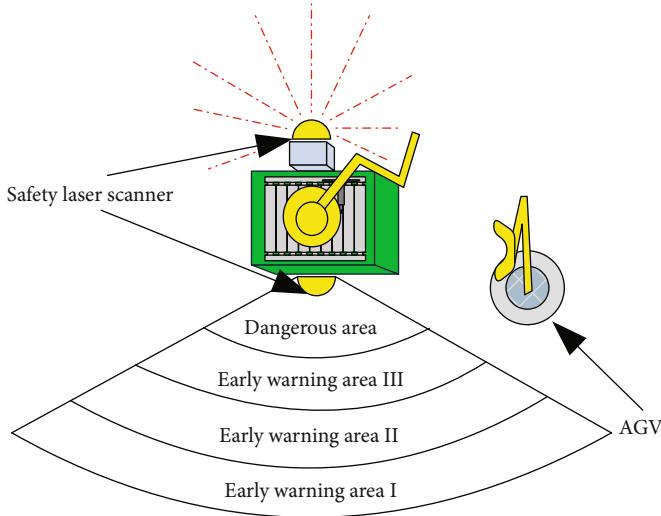


FIGURE 8: Diagram of mechanical safety protection system.

state, affecting the working efficiency. However, if the RealSense camera is installed at the robot and combined with the MobileNet-SSD algorithm to measure the distance between the human and the robot in real time, the corresponding signal is given to control the motion state of robot according to the distance, and the RealSense camera

does not monitor the AGV position information, so it can realize the early warning when the human enters the early warning area, but AGV does not alarm when entering the early warning area. The real-time status of the monitoring system when AGV enters the early warning area is shown in Figure 12.



FIGURE 9: Mechanical safety protection system.

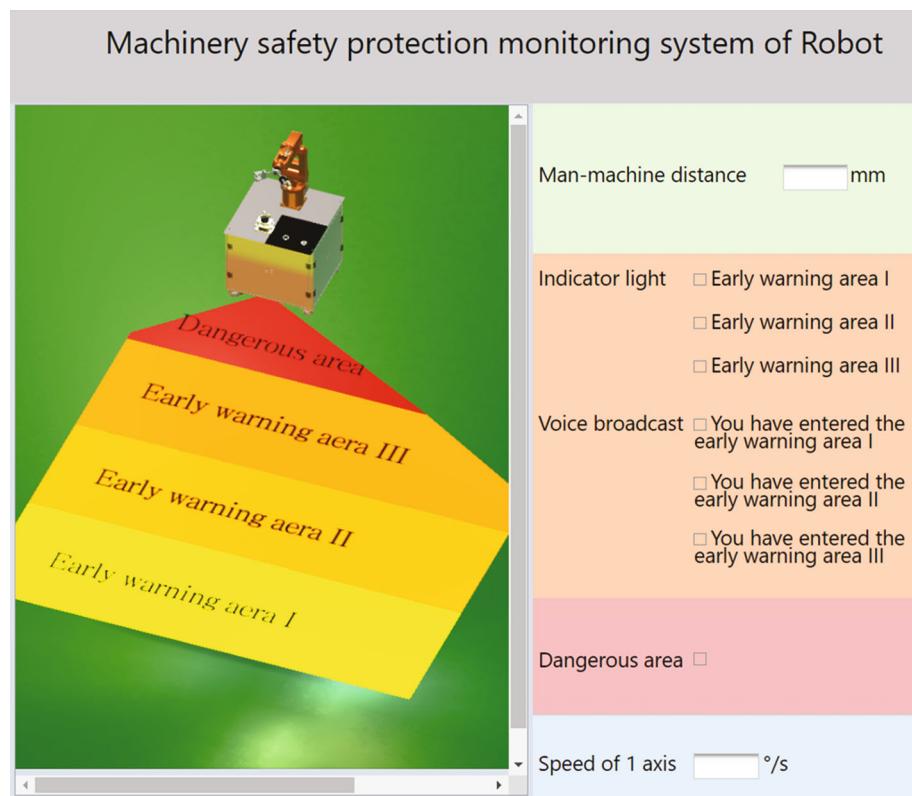


FIGURE 10: Machinery safety protection monitoring system of robot.



FIGURE 11: Real-time status of monitoring system.

5. Discussion

5.1. Advantages of Using Depth Camera to Realize Human Position Detection

- (1) Using depth camera to detect human position has cost advantages over safety carpet, safety light curtain, and safety laser scanner
- (2) Compared with human position detection devices such as safety carpet and safety light curtain, the depth camera can measure the distance between human and hazard sources in real time and carry out hierarchical early warning as described in Section 2
- (3) When the depth camera is used as the human position detection device, it can detect the human position without detecting the position of the movable equipment, so as to achieve the early warn-

ing of separation between the human and the movable equipment

- (4) When using the safety light curtain or safety pad as the human position detection device, it is necessary to prevent people from crossing the detection area of the detection device. If people are already in the dangerous area, they cannot be detected. If multiple depth cameras are fused to increase the field of view, detection blind area can be eliminated

5.2. Detection Range of Depth Camera. The detection range of depth camera is limited. For example, the field of view (FOV) of RealSense depth camera is $85 \times 58^\circ$, namely, 85° in horizontal direction and 58° in vertical direction. The detection distance is $0.1 \sim 10$ m. If we want to build a detection area larger than the FOV of camera, only using a single depth camera cannot meet the detection requirements.

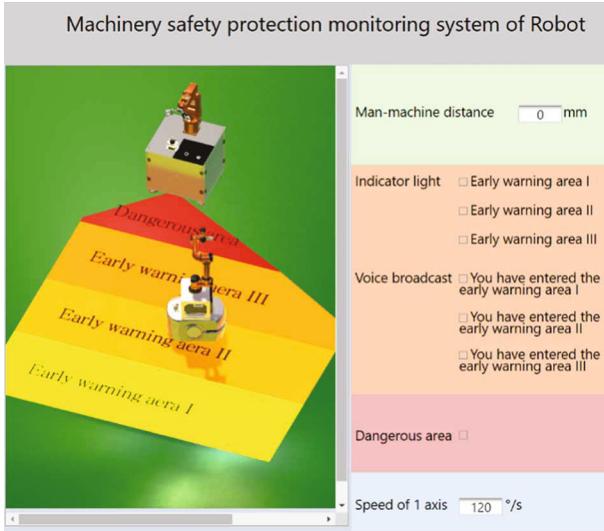


FIGURE 12: Monitoring system status when AGV invades early warning area.

However, the fusion of multiple depth cameras can increase the field of view and expand the range of the detection area. At present, scholars have carried out research on fusion system of multiple depth cameras. For example, Yan et al. [26] used four depth cameras to build a tracking system and used the FOV method to track and match the motion characteristics of objects, so as to solve the problem of object switching in overlapping areas between multiple cameras; In order to solve the problem of blurred image when detecting parts with a single camera, Wan [27] studied the methods of multicamera calibration and image mosaic and used the good panoramic mosaic image for detection. Hayat et al. [28] proposed a cost-effective 360° panorama generation system, which could process single view and three-dimensional panoramas, and eliminated the splicing gap of overlapping areas between adjacent cameras.

5.3. Minimum Distance between Man and Machine. There are some limitations in the measurement of man-machine distance. For example, if a part of the human (such as hands and legs) enters the dangerous area and the body trunk is in the safe area, the measured man-machine distance cannot judge whether it enters the dangerous area. Therefore, it is necessary to further measure the man-machine minimum distance. The man-machine minimum distance can be obtained by calculating the distance between each part of the human and the robot body. Taking the man-machine minimum distance as the judgment basis of whether to enter the dangerous area is one of the important directions of man-machine cooperative safety research. At present, scholars have carried out such research, such as Chen and Song [29] separated the human depth image from the collaborative space background, generated the point cloud, and clustered the point cloud using the k -nearest neighbor algorithm to find the minimum distance between man and machine. Wang et al. [30] established a man-machine distance model in a cooperative environment based on the

structural features of the robot and the human bone features extracted by three-dimensional vision sensors and iteratively calculated the minimum man-machine distance based on this model.

6. Conclusion

- (1) The Intel RealSense depth camera combined with MobileNet-SSD algorithm can detect human position in real time and replace the mechanical safety human position detection devices such as safety carpet, safety light screen, and safety laser scanner in specific application scenarios
- (2) When the depth camera is used as the position detection device of human, it can only detect human but not movable devices and realize the separation and early warning of human and movable devices

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was financially supported by the National Key R&D Program (No. 2017YFF020720).

References

- [1] ISO 12100, 2010, Safety of machinery-general principles for design-risk assessment and risk reductioner.
- [2] ISO 13855-2010Safety of machinery- positioning of safeguards with respect to the approach speeds of parts of the human body.
- [3] L. Mathe, A. Caverzasi, F. Saravia, G. Gomez, and J. Pedroni, “Detection of human-robot collision using kinetic,” *IEEE Latin America Transactions*, vol. 11, no. 1, pp. 143–148, 2013.
- [4] A. TUPPER and R. GREEN, “Pedestrian proximity detection using RGB-D data,” in *2019 Inter-National Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, Dunedin, New Zealand, 2019.
- [5] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2961–2969, Venice, Italy, 2017.
- [6] Z. Jian, F. Zhu, and L. Tang, “Research on human body recognition and position measurement based on AdaBoost and RGB-D,” in *2020 39th Chinese Control Conference (CCC)*, pp. 5184–5189, Shenyang, China, 2020.
- [7] L. Jian, Y. Biaobiao, and Z. Haoruo, “Human pose estimation based on SMPL model,” *Computer simulation*, vol. 38, no. 3, 2021.
- [8] Y. Jiayuan, Z. Lei, and Z. Kaibo, “Autonomous avoidance pedestrian control method for indoor mobile robot,” *Journal*

- of Chinese Computer Systems*, vol. 41, no. 8, pp. 1776–1782, 2020.
- [9] L. Qi, W. Xiangdong, and L. Hua, “3D human pose tracking approach based on double Kinect sensors,” *Journal of System Simulation*, vol. 32, no. 8, pp. 1446–1454, 2020.
- [10] D. Biswas, H. Su, C. Wang, A. Stevanovic, and W. Wang, “An automatic traffic density estimation using single shot detection (SSD) and MobileNet-SSD,” *Physics and Chemistry of the Earth*, vol. 110, pp. 176–184, 2019.
- [11] W. Li, X. S. Feng, K. Zha, S. Li, and H. S. Zhu, “Summary of target detection algorithms,” *Journal of Physics: Conference Series*, vol. 1757, no. 1, 2021.
- [12] L. Mingbo, “A survey of target detection algorithm based on machine learning,” *Computer products and circulation*, vol. 6, pp. 154–155, 2019.
- [13] L. I. Yinan, “A survey of research on deep learning target detection methods,” *China New Telecommunications*, vol. 23, no. 9, pp. 159–160, 2021.
- [14] F. Lu and Z. Yi, “A new pedestrian multi-target tracking algorithm,” *Computer Applications and Software*, vol. 38, no. 4, 2021.
- [15] Z. Bo, S. Yuanbin, X. Ruoxin, and Z. Shichao, “Helmet-wearing detection considering human joint,” *China Safety Science Journal*, vol. 30, no. 2, pp. 177–182, 2020.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [17] W. Liu, D. Anguelov, D. Erhan et al., “SSD: single shot multi-box detector,” in *European Conference on Computer Vision*, pp. 13–17, Springer, 2016.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, USA, 2016.
- [19] G. Yufei, *Research on target detection and ranging based on binocular vision*, Taiyuan University of Science and Technology, Taiyuan, China, 2020.
- [20] I. Anwar and S. Lee, “High performance stand-alone structured light 3D camera for smart manipulators,” in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp. 192–195, Jeju, Korea (South), 2017.
- [21] S. L. X. Francis, S. G. Anavatti, and M. Garratt, “Reconstructing the geometry of an object using 3D TOF camera,” in *2011 IEEE Workshop On Merging Fields Of Computational Intelligence And Sensor Technology*, pp. 13–17, Paris, France, 2011.
- [22] J. Wang, “Study on the performance identification of OpenCV in cashew nut shell-based activated carbon,” *IOP Conference Series: Earth and Environmental Science*, vol. 769, no. 3, 2021.
- [23] P. Huang, J. Jiang, and B. Yang, “Research status and progress of binocular stereo vision,” *Optical Instruments*, vol. 40, no. 4, pp. 81–86, 2018.
- [24] M. Li, B. Lu, and C. Jin, “Binocular ranging based on binocular vision stereo matching technology,” *Journal of Langfang Normal University*, vol. 19, no. 2, pp. 18–20, 2019.
- [25] P. Guo and H. Du, “Study on the robotic binocular distance measurement,” *Wireless Internet Technology*, vol. 15, no. 5, pp. 99–101, 2018.
- [26] M. Yan, Y. Zhao, M. Liu, L. Kong, and L. Dong, “High-speed moving target tracking of multi-camera system with over-lapped field of view,” *Signal, Image and Video Processing*, vol. 15, no. 7, pp. 1369–1377, 2021.
- [27] W. A. N. Yuandong, *Multi-Camera calibration and image mosaic method for detecting parts*, Huazhong University of Science and Technology, Wuhan, China, 2017.
- [28] H. Ullah, O. Zia, J. H. Kim, K. Han, and J. W. Lee, “Automatic 360°mono-stereo panorama generation using a cost-effective multi-camera system,” *Sensors*, vol. 20, no. 11, p. 3097, 2020.
- [29] J. H. Chen and K. T. Song, “Collision-free motion planning for human-robot collaborative safety under Cartesian constraint,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4348–4354, Brisbane, QLD, Australia, 2018.
- [30] Z. Wang, Y. Gan, and X. DaiRobot, “Human-robot distance modeling and minimum distance iterative algorithm in collaboration environment,” *Robot*, vol. 40, no. 4, pp. 413–422, 2018.