Mechanical arm teleoperation control system by dynamic hand gesture recognition based on kinect device

Pu Tu1✉, Chen Huang2
1Sichuan University of Arts and Science, Dazhou, People’s Republic of China
2Dazhou Public Security Bureau, Dazhou, People’s Republic of China
✉E-mail: tupu0803@126.com

Abstract: The research achieved to control the mechanical arm by using real-time dynamic gesture recognition based on Kinect. It uses the unmarked gesture segmentation algorithm based on the palm neighbourhood and the threshold detection algorithm based on the palmar contour to identify the operator’s gestures and moving trajectories, and to convert it into the specific action of the mechanical arm. Through wireless network, this system sends control instructions to the mechanical arm to realise teleoperation control. The system also provides video feedback of mechanical arm operation site. It improves the teleoperation telepresence and interactivity and avoids operation error such as grasp nothing or fall. The results of the experiment indicate that the operation of the gesture control system is simple and easy, the response of mechanical arm is quick and accurate and the human–machine interaction is intuitive and friendly.

1 Introduction
With the rapid development of machine vision technology, mechanical arm interaction control system based on gesture recognition has become a research hotspot in the field of human–computer interaction. In the research of human–machine interaction for gesture recognition at home and abroad, it is mainly divided into gesture recognition technology based on data glove and gesture recognition technology based on visual information.

Wu Jiang qin and Li Zhen improved the gesture recognition rate of data glove by combining the algorithm of ANN and HMM, but it has not solved the problems of high cost and inconvenience of wearing [1, 2]. Lu Xiao min et al used the Kinect sensor to obtain the depth data, through the HMM, SURF, and other algorithms for hand gesture recognition, used to control autonomous mobile robots, mechanical arm, and intelligent wheelchair and other equipment [3–7]. Wang Yi et al used the application of Kinect sensor in augmented reality, the mechanical arm's motion path planning and teaching learning are completed [8, 9]. Xiong You jun and Liu Jun et al proposed that the machine, people, and target object can be operated in different space by the teleoperation technology of connecting the mechanical arm to the network [10–12]. The execution ability and security of the machine in the special environment of deep sea exploration, explosion, and radiation can be improved to a great extent. Tang Wei cai et al proposed a scheme that increased the video feedback information for the teleoperation of the mechanical arm, it improved the performance of the mechanical arm operation [13]. Here, we design a teleoperation control system for mechanical arm based on Kinect sensor. It uses the unmarked gesture segmentation algorithm based on the palm neighbourhood and the threshold detection algorithm based on the palmar contour to identify the operator's gestures and moving trajectories. At the same time, it uses embedded system and wireless network to build a field monitoring auxiliary system. This system can provide video feedback for mechanical arm operation to increase the accuracy. The system not only has the advantage of natural body sense interaction but also realises the integration of virtual operation and real scene, making the control process more natural and real (Fig. 1).

2 System structure
This system mainly includes hand gesture segmentation and feature recognition, mechanical arm, operation instruction wireless transmission, video capture, and transmission system and so on. First, we divide the depth image of the operator’s gesture and map the behaviour characteristics and movement trajectories of the gesture to the coordinate space of the mechanical arm. Then, through the wireless network, we send the upper computer operation instructions to the mechanical arm to carry out the specific operation. Finally, we use video acquisition and transmission system to collect video images from the visual angle of the outside scene and the target plane, and display and store the video image on the operator's terminal device based on the streaming media technology.

3 Gesture segmentation and feature recognition
3.1 Gesture segmentation
At present, the OpenNI and NITE middleware provided by the third party has provided the Kinect developer with a more accurate location of the palm of the hand. Mai Jian hua et al have proposed a depth threshold gesture segmentation algorithm [14]. Based on this algorithm, we have designed an unmarked gesture segmentation algorithm based on the neighbourhood of the palm. The implementation process of the algorithm is as follows: using NITE middleware to obtain the location of the current palm, and
caching all points within the distance from the palm of the hand, traversing all points in the depth image pixel by pixel. We make the grey value of the point with a distance greater than the threshold value of the palm is 0, and the grey value of the rest points is 255.

As shown in formula (1):

\[
\text{handpoint}(\text{index}) = \begin{cases} 
0, & \text{realsegment}(\text{index}) \leq \text{Dis} - T \\
0, & \text{realpoint}(\text{index}) \geq \text{Dis} + T \\
255, & \text{otherwise}
\end{cases}
\]  

(1)

handpoint (index) is the palm heart depth image after the segmentation, realpoint (index) is the full depth pixels of the cache. \(D\) is represents the current position of the palm, and \(T\) represents the neighbour threshold of the palm. Through a number of tests, a relatively stable threshold value of 95–105 mm was determined. The test contrast is shown in Fig. 2.

Figs. 2a–d are hand-segmented images with thresholds of 50, 100, 150 and 200 mm, respectively. Experiments show that when the threshold value is 100 mm, the segmentation effect is the best, which meets the system requirements.

### 3.2 Gesture recognition

This paper designs an algorithm based on palm-contour threshold detection. The realisation of the algorithm is that when the palm is fully opened, the total width and total height of the external contour of the hand will increase – the horizontal interval between the most right and the leftmost point (the tip of the thumb and the tip of the finger) in the hand point cloud will increase, and the vertical distance between the highest and lowest hand points in the hand point cloud (the middle finger tip and the wrist root) will also increase. Conversely, when the palm is closed in a ‘grasp’ state, the total width and total height of the external contour will be reduced. Therefore, when the palm contour is greater than or less than the threshold value, the ‘release’ or ‘grab’ gestures can be determined. As shown in formula (2) and (3):

\[
\text{handVec} = \begin{cases} 
\text{handRight} = \max (\text{handPoint}(\text{index}), x) \\
\text{handLeft} = \min (\text{handPoint}(\text{index}), x) \\
\text{handTop} = \max (\text{handPoint}(\text{index}), y) \\
\text{handBottom} = \min (\text{handPoint}(\text{index}), y) \\
\end{cases}
\]  

(2)

\[
\text{boolean}() = \begin{cases} 
\text{open}, & \text{handRight} - \text{handLeft} \geq D1 \\
\text{open}, & \text{handTop} - \text{handBottom} \geq D3 \\
\text{close}, & \text{handTop} - \text{handBottom} \leq D4 \\
0, & \text{otherwise}
\end{cases}
\]  

(3)

HandVec represents the external contour of the hand. handRight, handLeft, handTop, and handBottom represent the right, left, highest, and lowest positions of the hand obtained from the hand depth image after the segmentation.

Boolean () represents the decision function of the contours of the opponent. When the total width of the contour exceeds \(D1\) and the total height exceeds \(D3\), the ‘release’ gesture is determined. In the same way, the ‘grab’ gesture is determined when the total width of the external contour of the hand is less than \(D2\), and the total height is less than \(D4\). Through many experiments, the width threshold of the ‘release’ gesture is about 165–200 mm, and the height threshold is about 155–180 mm. The width of the grabbing gesture is about 55–70 mm, and the height threshold range is about 80–95 mm.

### 3.3 Gesture tracking and coordinate transformation

The mechanical arm has 3 degrees of freedom. The base moves in the X-axis plane, the big arm moves in the Y-axis plane, and the clamping moves with open and close in the X-axis plane, and the clamping range is 0–100 mm. We adopt a trajectory tracking algorithm based on the recognition centre point. The algorithm takes the position of palm as the origin when the system first recognises the palm of the operator. By the relative displacement of the operator's current hand point and the position of the origin, we calculate the motion of the gesture. The line drawn from the origin to the palm of the palm is displayed on the terminal. As shown in Fig. 3.

In Fig. 3, point \(O\) represents the position of the palm when the system first recognises the operator's palm – the origin point, point \(A\) represents the position of the current palm, the line \(OA\) represents the line of the origin to the hand point, \(D\) represents the straight line length.

Here, the Kinect cone is drawn on the operator's screen through the third party library to provide the operator with accurate gesture tracking view. On this basis, the relative motion vector of the forearm is calculated and mapped to the relative displacement in the direction of \(X\) and \(Y\) axes of the mechanical arm, and sent to the mechanical arm. It completes the transformation between the Kinect sensor coordinate system, the screen coordinate system, and the mechanical arm coordinate system, as shown in Fig. 4:

### 4 Wireless network transmission

#### 4.1 Operation instruction transmission

After the upper computer completes the computation from gesture recognition and tracking to the movement command of the specific mechanical arm, the transmission and reception of operation instructions are realised through wireless network system. The transmitter and receiver adopt the connection of point to point and the pairing of one to one. It sends and receives data according to the transparent serial interface protocol. It is shown in Fig. 5.
In this system, the transmitter of the wireless network is connected with the host computer, it receives the motion instruction by the serial port and sends the motion instruction to the wireless receiver based on ZigBee. The wireless receiver is connected to the mechanical arm, and the hardware composition is the same as the sending end. The difference is that the receiving end parses the received motion instructions and sends them to the mechanical arm for specific execution.

4.2 Video acquisition and transmission

In the position of the shoulder joint of the mechanical arm, we set a camera which provide the video information concerned with the view of the outside scene. At the end of the holder of the mechanical arm, we set a camera which provide the video information of the position and posture of the target plane and the grasping object directly. At the same time, we also build an embedded video image processing and transmission platform. Finally, we transmit video images to the operator's terminal equipment for display and storage based on streaming media technology and WiFi. It is shown in Fig. 6.

5 Experimental test results

First, we use gesture to guide the mechanical arm in four directions: upper, lower, left, and right movements to observe the movement. Then, the gripper grabs the object through the ‘grab’ and ‘release’ gestures. Fig. 7 shows the experimental results of the motion of the mechanical arm when tracking the operator's gestures. Fig. 8 shows the gesture tracking and field operation video displayed on the operator's terminal.

In Figs. 7 and 8, the system accurately identifies the gestures and moving trajectories of the operator and converts them to the moving trajectory of the mechanical arm. For gesture control commands, the mechanical arm responds in sequence, the movement process is smooth, the action is accurate, and there is no malfunction and other control anomalies. It is proved that the host computer conveys the operation instruction to the mechanical arm through the wireless network and realises teleoperation control. This shows that the system has a good interactivity.

6 Conclusion

Here, the dynamic gesture information collected by the Kinect sensor is used to control the mechanical arm in real time. It uses the unmarked gesture segmentation algorithm based on the palm neighbourhood and the threshold detection algorithm based on the palmar contour to accurately identify the operator's gestures and moving trajectories, and convert it to the specific action of the mechanical arm. The mechanical arm has the characteristics of quick response, smooth movement, and accurate action. It is no malfunction and other control abnormality. At the same time, this proves that the host computer sends instructions to the mechanical arm in real-time and accurately through wireless network and realises teleoperation control. The system also provides the video feedback of the mechanical arm operation, which helps the operator to make the accurate judgment and decision on the remote operation scene in the process of teleoperation. It improves the
teleoperation sense of field and interactivity and avoids the operation errors such as grab nothing and fall.

7 Acknowledgments
This research was supported in part by Sichuan Province Education Department Project (18ZB0508)

8 References
[1] Wu, J.Q., Gao, W., Chen, X.L.: 'A system recognizing Chinese finger-spelling alphabets based on data-glove input', PR&AI, 1999, 12, (1), pp. 74–78
[2] Li, Z.: 'The design of tele-peration manipulator controlled by hand gesture'. Master thesis, Wuhan University of Technology, 2014
[3] Cicirelli, G., Attolico, C., Guaragnella, C., et al.: 'A kinect-based gesture recognition approach for a natural human robot interface', Int. J. Adv. Robot. Syst., 2015, 12, (3), pp. 12–22
[4] Lu, X.M.: 'The research of human-robot interaction technology based on hybrid sensors'. Master thesis, South China University of Technology, 2014
[5] Tang, M., Ponce, C., Selman, B.: 'Unstructured human activity detection from RGBD images'. Proc. Int. Conf. Robotics and Automation. Minnesota, USA, 2012, pp. 842–849
[6] Lin, H.B., Wang, H., Zhang, Y.: 'Markerless human-machine interaction control system using kinect sensor', Comput. Eng. Des., 2015, 36, (2), pp. 551–555
[7] Liao, Y., Xie, Y., Zhang, Y.: 'Design and implementation of a gesture-driven system for intelligent wheelchairs based on the kinect sensor', Robot, 2012, 34, (1), pp. 110–113
[8] Fang, H.C., Ong, S.K., Nee, A.Y.C.: 'Novel AR-based interface for human-robot interaction and visualization', Adv. Manuf., 2014, 2, (4), pp. 275–288
[9] Wang, Y.: ‘Augmented reality system based on range image with collision interaction between virtual and actual objects’. Master thesis, Beijing Forestry University, 2015
[10] Xiong, Y.J.: ‘A dissertation submitted in partial fulfillment of the requirements’. PhD thesis, Huazhong University of Science and Technology, 2005
[11] Liu, J.: ‘The research of 3D reconstruction of remote control robot based on stereo vision method’. Master thesis, Beijing University of Posts and Telecommunications, 2009
[12] John, J.C.: ‘Introduction to robotics mechanics and control’ (China Machine Press, Beijing, 2007)
[13] Tang, W.C., Chen, S.G., Xiao, Y.: ‘Visual information presentation mode for teleoperation of manipulator’, Space Med. Med. Eng., 2017, 30, (1), pp. 73–78
[14] Mai, J.H.: ‘Natural gesture-based robot control’. Master thesis, South China University of Technology, 2013