

Human's Gesture Recognition and Imitation Based on Robot NAO

Chen Wenbai, Wu Xibao, Wang Sai and Gao Hui

*School of Automation, Beijing Information Science & Technology University,
Beijing 100192, P. R. China*

*chenwb@bistu.edu.cn, wuxibao@bistu.edu.cn, wangsai@bistu.edu.cn,
gaoh@bistu.edu.cn*

Abstract

Based on Kinect platform, human's gesture recognition and imitation were realized by humanoid robot NAO in this paper. The hardware system structure of Kinect platform and the principle of human skeleton extraction were mainly introduced firstly. From the image, Kinect camera gets the skeletal point information which is used to compared with preset posture of the skeletal information. Through calculate the angle between the different bones, the computer obtains the human's posture and sends the instructions to the robot. In the physical experiments, NAO robot receives the pre-programmed corresponding action instructions and imitates the corresponding motions of humans.

Keywords: *Humanoid Robot; NAO; Human's Gesture Recognition; Human's Gesture Imitation; Skeleton Extraction*

1. Introduction

Human's gesture recognition is one of the important research area of Machine Vision in the modern computer vision research areas, Where more and more people pay attention to human-computer interaction. In a wide variety of human gesture recognition algorithm[1], the technical knowledge of human posture based on knowledge of skeletal information is an excellent way. Kinect [2] appeared in this context born, which makes machines to interact with the world expanded from two-dimensional to three-dimensional space, and enters into a non-contact interactions. Skeletal tracking, face recognition, microphone input, voice recognition based on Kinect & NAO robot are studied in this paper. From the image, Kinect camera gets the skeletal point information which is used to compared with preset posture of the skeletal information. Through calculate the angle between the different bones, the computer obtains the human's posture and sends the instructions to the robot. In the physical experiments, NAO robot receives the pre-programmed corresponding action instructions and imitates the corresponding motions of humans.

2. Platform Introduction

2.1. Kinect

Kinect sensor is shown in Figure 1. It is a horizontal bar connected to a small base with a motorized pivot and is designed to be positioned lengthwise above or below the video display. The device features in "RGB camera, depth sensor and multi-array microphone running proprietary software", which provide full-body 3D motion capture, facial recognition and voice recognition capabilities[1].

The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions [2].

The sensing range of the depth sensor is adjustable, and Kinect software is capable of automatically calibrating the sensor based on gameplay and the player's physical environment, accommodating for the presence of furniture or other obstacles [3].

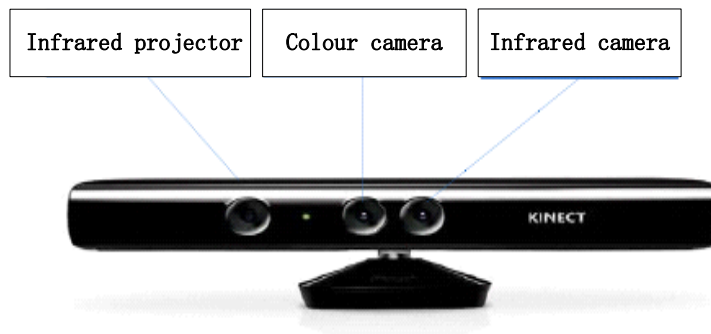


Figure 1. Kinect Appearance and Structure Diagram

2.2. Humanoid Robot NAO

Humanoid robot Nao is shown in Figure 2. It is an autonomous, programmable robot developed by Aldebaran Robotics company. Nao had replaced Sony's robot dog Aibo as the robot used in the RoboCup Standard Platform League (SPL) [4] With a flattering shape, and a certain emotional intelligence, NAO robot can interact with human kind. As shown in Table 1 [5], the technical parameters of NAO robot have reached the advanced level, not only the production process but also the internal system components, the NAO humanoid robot sufficiently meets the current needs of the experiment.



Figure 2. Robot NAO

Table 1. Technical Parameters of Nao Robot

Function	paramepers
Degree of freedom	25DOF
Hight	58 cm
Weight	4.3 kg
Charger	AC 90-230 V/DC 24 V
Autonomy	90 min
Voice	Two speakers and voice synthesis
Vision	Two CMOS camera
Connectivity	Ethernet, Wi-Fi
Programming languages	Python, C++, java

3. System Implementation

3.1. Human Gesture Recognition Method

As shown in Figure 3, visual posture actions recognition, commonly can be divided into three methods, that is, statistics- based, template matching-based and semantics-based method.

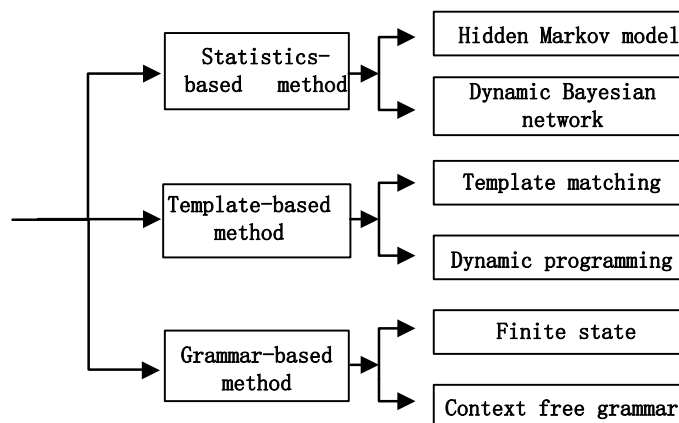


Figure 3. Human Gesture Recognition Method

The basic idea of template-matching method can be introduced as follows. Computer system uses the obtained image to match the image in database, and the most similar matching image will be its posture. So, it is necessary to establish a database on body posture image, and enter a variety of body posture image as many as possible. When computer recognizing the human's gesture, the accuracy depends on the quality and quantity of the database largely. Therefore, the database must have a lot of postures to ensure a high reserve of image data recognition accuracy.

3.2. Data Extraction Based on Kinect

Data extraction process based on kinect is shown in Figure 4. In the human skeleton image captured by Kinect, (X, Y, Z) denotes the position of one joint. The infrared projector of Kinect emits a beam of infrared light, which will be divided into many light

beam after through scattering sheet. The beams form a scattering spot when they converge together at the target plane.

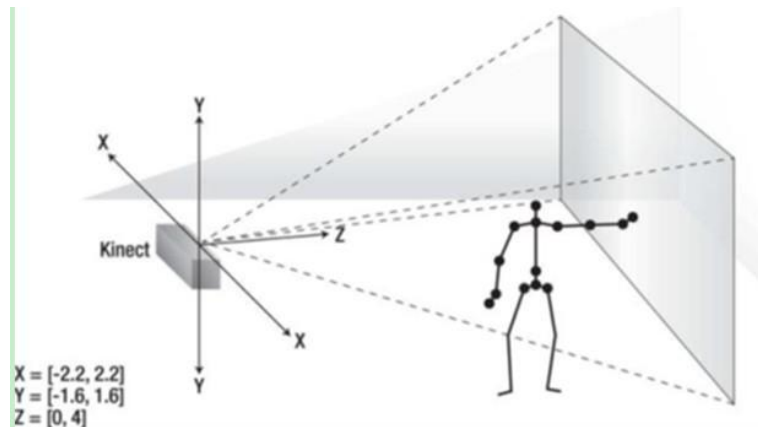


Figure 4. The Image Coordinates of Kinect

The infrared camera will capture scattering spot which is compared with the primary internal stored reference model. When it is different between the distance of captured spot and the primary internal stored reference model, the spot at the position of infrared image will move along the reference line. Range of all spot movement can measured through the correlation of image, and this system will form a parallax image. Through the corresponding displacement of parallax image, the depth distance of each pixel can be calculated, the specific process is shown in Figure 5[6].

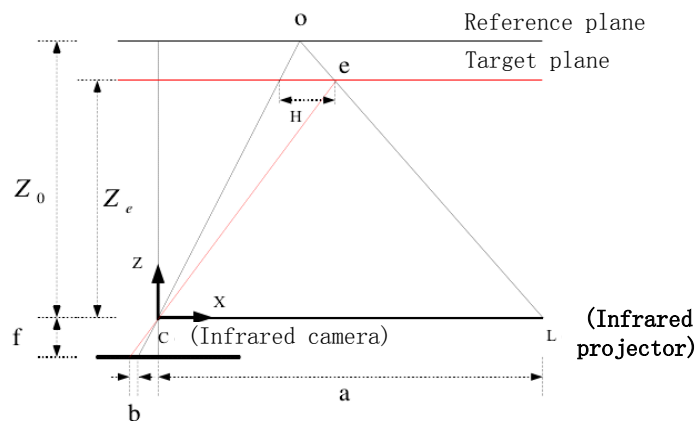


Figure 5. Obtaining the Depth Image Schematic Diagram

Assuming point e as a reference point, b as the displacement distance measured, the origin of coordinates for infrared camera center location, the Z axis to the camera point at objects direction as positive direction, the X axis is perpendicular to the Z axis to the right direction, the direction of Y axis is perpendicular to vertical of the X axis and Z axis. It hypothesizes that objects on the reference plane, and considers the distance Z_0 , infrared spot reflected by Infrared from Kinect can be obtained through infrared camera.

When the object moves back and forth, the spot will also changes at the same time. According to the triangle similarity theorem, the object change of displacement b can be written as [7]:

$$\frac{H}{a} = \frac{Z_o - Z_e}{Z_o} \quad (1)$$

$$\frac{b}{f} = \frac{H}{Z_e} \quad (2)$$

Where, Z_e denotes the point e in object space depth, a is the baseline length, f is the infrared camera focal length. H denotes the displacement of reference point e in real space, b is the displacement of observed object. From formulas (1) and (2), it can get:

$$Z_e = \frac{Z_o}{1 + \frac{Z_o}{fa}b} \quad (3)$$

Formula (3) is used to calculate the depth of objects by the displacement in the image. Assuming the image coordinate system parallel to the baseline, parallel to the depth coordinate system, the coordinate of each point in the image can be obtained by Z_e :

$$X_e = -\frac{Z_e}{f}(x_e - x_o + \delta_x) \quad (4)$$

$$Y_e = -\frac{Z_e}{f}(y_e - y_o + \delta_y) \quad (5)$$

Here, (x_o, y_o) represents the coordinate origin, (x_e, y_e) represents the image coordinates of the points, (δ_x, δ_y) represents the distortion of Kinect camera.

3.3. Gesture Recognition

In the process of calculating the skeletal angle, we can identify with the two adjacent joint angle information by the position information of the joint, and describe the gesture of the robot, Posture of the type and complexity determines the complexity of the gesture recognition process.

Through the joint point cross posture recognition is the joint point hit test. We can determine a position of the joint is at the interface of the effective range of a certain visual elements or not. We can do the same test on the joint point. But it needs less work because all the joint point are in the same coordinate space, which makes the calculation easier. For example, akimbo action (hand-on-hip), you can position from skeletal tracking data acquisition left hip and left and right hand. Then calculate the left hand and left hip position. If the distance is less than a threshold, we can say the two points are intersect.

There are some simple gestures can have very high recognition rate. For example, a user with open arms and shoulders in a line in this position, called the T posture. He can use node intersection, hand, elbow, and shoulder are in approximately the same position on the Y axis. Another method is to calculate the joint point between certain angle. Skeletal tracking engine can recognize up to 20 joint point data. Any of the three joint points can form a triangle. Use a triangular geometry can calculate the angle between them. We only need to two joints can draw a triangle, the third point location can be decided to two joints have been determined.

As shown in Figure 6, the coordinates of an arbitrary two nodes can calculate the value of the corresponding side.

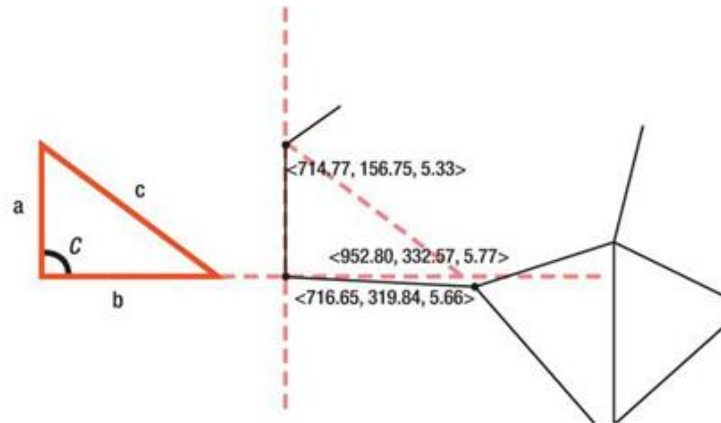
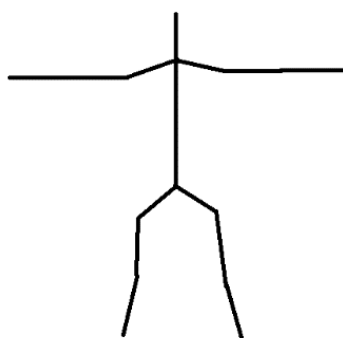


Figure 6. Schematic Diagram of Calculating Angle

In Figure 6, the graph is the elbow and wrist graphics. The graph can be calculated from three angle joint according to coordinate these three joints. There are two methods in calculating the angle. The first method is the most obvious as in the example above, we use the coordinates of three nodes to construct a triangle. Another method is to use two nodes, the remaining node set freely. In the above diagram, we use the first method, method of using three nodes are composed of the wrist, elbow and shoulder. No matter how changes in the other parts, the three triangular shape relatively unchanged.

In this experiment, we use the second methods. First of all, we need to customize a gesture library to store the gesture we defined.

Figure 7(a) shows the gesture of "open the arm", in which the shoulder, the wrist, the palm are in the horizontal direction. In the program, the threshold angle of joint is limited in 8 degree. Figure 7(b) shows the gesture of "Close the arm". The left and right side of the shoulder are in horizontal direction, and the angle between wrist and vertical direction is 25 degree. As shown in Figure 7(c), the left shoulder, wrist, palm are located in downwards vertically, and the threshold angle is limited in 10 degree. The right shoulder, wrist, palm stretch upward, the threshold angle is set at 30 degree. As shown in Figure 7(d), the gesture of "Raise the left hand" is similar to the gesture of "Raise the right hand" in Figure 7(c).



(a) Open the Arms



(b) Close the Arms



Figure 7. Gesture Library of Humans: (a) Open the Arms, (b) Close the Arms, (c) Raise the Left Hand, (d) Raise the Right Hand

4. Experiment Results

It is the key to establish robot posture library and record the relevant data in the experiment. At present, posture library contains the following basic motion data:

- Sit down: Instruct the robot to sit down;
- Open the arms: Instruct the robot to open the arms;
- Close the arms: Instruct the robot to close the arms;
- Raise your right hand: Instruct the robot to raise right hand;
- Raise your left hand: Instruct the robot to raise left hand.

After Kinect camera obtains human's skeletal point information from depth images, gesture recognition system will judge the bone body posture by comparing the skeletal point information. The recognized attitude will be sent to the robot NAO by means of voice command. And NAO can imitate the corresponding motions. Human's gesture recognition and imitation program flow shown in Figure 8.

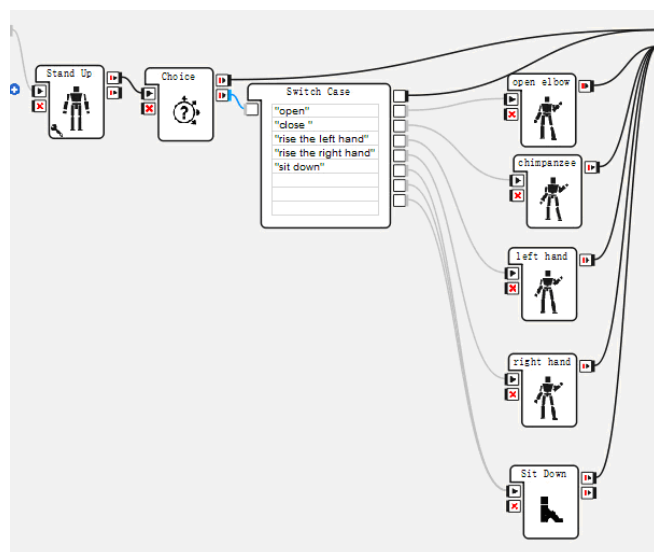


Figure 8. Human's Gesture Recognition and Imitation Program Flow Diagram

As shown in Figure 9-12, the experiment procedure can be divided into the following three parts. The first one is the person's gesture, the second is the skeletal frame which is processed by Kinect sensor, and the third is that the robot obey the voice instructions and make the same action with the person.

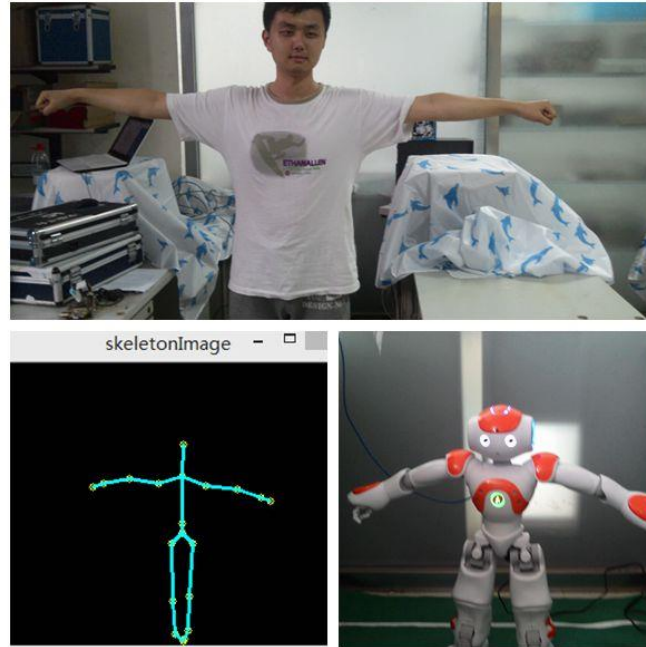


Figure 9. Open the Arms

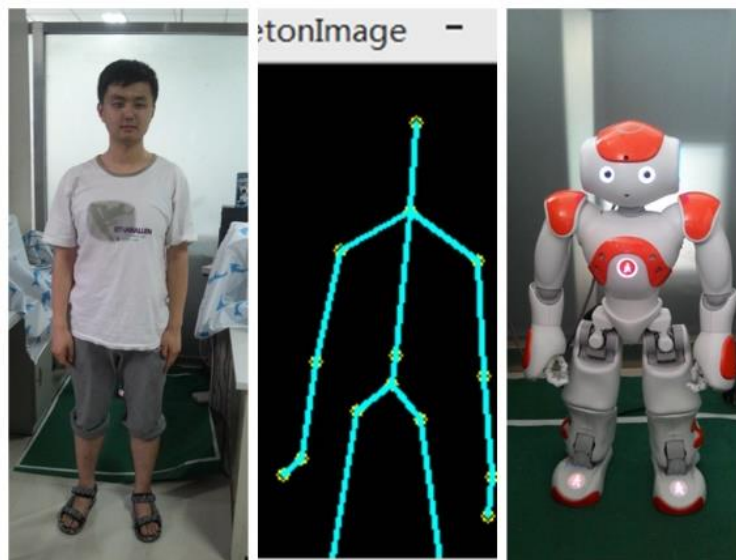


Figure 10. Close the Arms

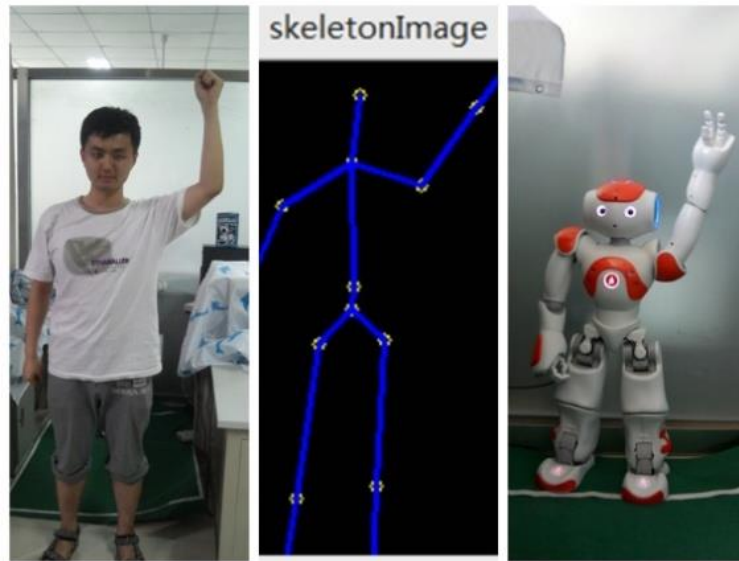


Figure 11. Raise the Left Hand

Figure 9 shows the recognition and imitation action of opening the arms by robot NAO. Figure 10 shows the recognition and imitation action of closing the arms. Figure 11 and Figure 12 show the recognition and imitation of raising the left and right hand action of robot NAO.

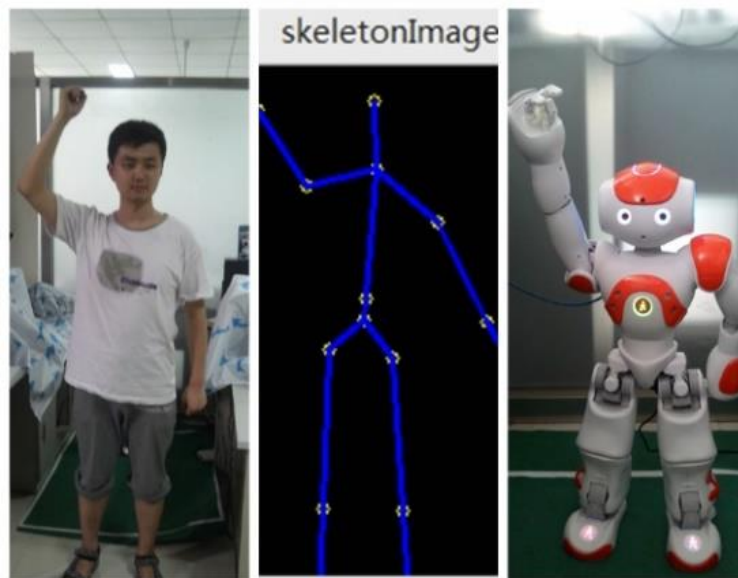


Figure 12. Raise the Right Hand

5. Conclusion

Based on Kinect platform, human's gesture recognition and imitation were realized by humanoid robot NAO in this paper. The hardware system structure of Kinect platform and the principle of human skeleton extraction were mainly introduced firstly. From the image, Kinect camera gets the skeletal point information which is used to compared with preset posture of the skeletal information. Through calculate the angle between the different bones, the computer obtains the human's posture and sends the instructions to

the robot. In the physical experiments, NAO robot receives the pre-programmed corresponding action instructions and imitates the corresponding motions of humans.

Acknowledgments

This work was partly supported by The Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (NO. CIT&TCD201404125), Youth Innovation Fund program of AGRS (2013YFL06) and Comprehensive Reform of Personnel Training (PXM2014_014224_000091).

References

- [1] T. Stephen, (2010) January 7, "Natal Recognizes 31 Body Parts, Uses Tenth of Xbox 360 "Computing Resources"", Kotaku, Gawker Media, Retrieved, (2010) November 25.
- [2] T. Stephen, (2009) June 5, "Microsoft: Project Natal Can Support Multiple Players, See Fingers", Kotaku, Gawker Media, Retrieved, (2009) June 6.
- [3] W. Mark and B. Matt, (2009) June 3, "Testing Project Natal: We Touched the Intangible", Gizmodo, Gawker Media, Retrieved, (2009) June 6.
- [4] "Nao robot replaces AIBO in RoboCup Standard Platform League", Engadget, (2007) August 16, Retrieved (2012) October 4.
- [5] L. Ismail, S. Shamsuddin, H. Yussof, H. Hashim, S. Bahari, A. Jaafar and I. Zahari, "Face Detection Technique of Humanoid Robot NAO for Application in Robotic Assistive Therapy", Control System, Computing and Engineering (ICCSCE), 2011, International Conference on IEEE, (2011), pp. 517-521.
- [6] J. Guan and Meng and M. Q.-H, "Study on distance measurement for NAO humanoid robot", Robotics and Biomimetics (ROBIO), 2012, International Conference on IEEE, (2012), pp. 283 – 286.
- [7] L. Ming, "Research on Skeleton Localization Based on Kinect Sensor[D]", Wuhan University of Science and Technology.

Authors



Wenbai Chen, he was born in Sichuan province, China in 1975. He received his B.E. degree in Control Technology and Instruments from northeast university at Qinghuangdao, China, in 1997 and his M.E. and Ph.D. degree in Electronic Science and Technology from Yanshan University and Beijing University of Posts and Telecommunications, China, in 2004 and 2011 respectively. Now he is an associate professor in the School of Automation, Beijing Information Science & Technology University, Beijing, China. He has more than 17 years of experience in teaching. His current research interests include intelligent robot and its applications, digital image and video processing. He is member of a council in China Association of Artificial Intelligence (CAAI).



Xibao Wu, he was born in Jiangxi province, China in 1979. He received his B. E. degree in Electronic Information Technology from Beijing Information Science & Technoogy University, Beijing, China in 2000 and his M.E. degree in Control Technology from Beijing Institute of Technology, China, in 2007. Now he is an Lecturer in the School of Automation, Beijing Information Science & Technology University, Beijing, China. His current research interests include intelligent robot, sensor technology and its application.



Sai Wang, he was born in Shandong province, China in 1992. She received her B.S. degree in The intelligent science and technology from Beijing Information Science and Technology University, Beijing, China, in 2014. His interest is to research vision processing and intelligent system.



Hui Gao, she was born in Jilin province, China in 1989. She received her B.S. degree in Automation from North China Institute of Science and Technology, Hebei province, China, in 2012. Now she is pursuing his M.E. degree in Control Engineering in Beijing Information Science and Technology University, Beijing, China. His interest is to research sensor technology and intelligent system.

