Research and Implementation of a Humanoid Robot Control Method Based on Action Mapping

Xuezhen Li*, Yuanwen Zhang, Rui Yang and Jiayun Zhang
University Of Science & Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing, China

*Corresponding author email: 41804344@xs.ustb.edu.cn

Abstract. In order to reduce programming difficulty and let humanoid robot act more human-like, we designed a robotics control method for humanoid robots. Firstly, based on human body movements collected by Kinect camera, we calculated 3D data of joint point with pinhole model, and established space model of robot and skeleton model of human. Then, we transformed 3D coordinates of human joints into joint values by space vector method, mapping to the joints of the robot and guides the robot to move. Finally, with semi-autonomous control experiment, it shows that humanoid robot action mapping strategy is feasible and effective. Compared with the existing similar research, this paper implements the action guidance control of humanoid robot, enhances the intelligent degree of robot and expands the application scene and scope of humanoid robot. Depending on the human-like appearance, humanoid robot with action mapping could provide a much more intelligent interaction method.

Keywords: Machine vision; Space vector method; Humanoid robot; Action mapping.

1. Introduction
In recent years, with the continuous development of artificial intelligence and robot technology, humanoid robot with its humanoid appearance characteristics, occupies an unparalleled advantage in human-computer interaction and plays an increasingly important role in teaching, scientific research, industry and other fields. However, there are a large number of steering gears in a humanoid robot. The traditional control of humanoid robot based on pre-programming is complex, costly and unintelligent, which brings huge obstacles to the application of humanoid robot. Therefore, it is imperative to explore a more intelligent and convenient humanoid robot control method.

The accuracy and stability of biped walking are the basic problems in the motion control of humanoid robot. Japan's Ichiro Kato first firstly developed the WL-9DR biped robot, which is controlled by the program with the actuator value of each joint designed in advance, which implements the quasi dynamic walking with the step length of 45 cm and each step of 10 s[1]. Japan's Honda company launched a fully self-supporting two legged walking machine P3, and the application of decentralized control technology makes it possible to miniaturize and lightweight humanoid robots[2]. Pratt of Massachusetts Institute of technology put forward virtual model control strategy in the control of spring Turkey and spring Flamingo biped robots, which can avoid tedious inverse kinematics effectively, improving the immediacy of humanoid robot motion[1,3].

Action mapping is a kind of robot control strategy to improve the cognitive ability and behavior ability of robots. L. Zhang used the skeleton data captured by Kinect camera to represent the whole body movement of human, with which he constructed the eight-chain human body model of the humanoid robot and real human to simulate human motion[4]. Yang used a lower-body DRM to find valid leg
configurations on uneven floors, presenting a framework using a paired forward-inverse dynamic reachability map to accomplish manipulation tasks on uneven terrains while avoiding obstacles[5].

Wang used the EWMA in the action mapping strategy to smooth the acquired skeleton data based on the local and entire process without prior knowledge of noise, improving the feasibility of collected data[6].

Most of the existing research only discuss the mapping between manipulator and human arm. Aiming at the motion mapping problem of humanoid robot, this paper proposes an action mapping strategy based on Kinect depth camera. The depth information of the coordinate value is obtained by tf-pose-estimation method, with which the three-dimensional coordinates of the human body can be obtained. Then we established the skeleton diagram of the human body and the space model of the robot. The joint values of the human body are obtained according to the space vector method. After the amplitude limiting filtering, the motion of the humanoid robot is guided and controlled. Finally, the feasibility of the whole idea is verified by semi-autonomous experiment, proving that it is practicable to control humanoid robot by real-time action mapping.

2. Human Motion Data Model

The human motion data model mainly includes three parts: human motion data acquisition, human motion modeling, and human motion model modification.

Firstly, RGB-D camera is used to obtain the action, color camera obtains color image, and infrared camera obtains depth image. If the resolution of the depth map obtained by RGB-D is different from that of the color image and the one-to-one mapping of pixels cannot be realized, it is necessary to registration the color image with the depth image and adjust the resolution to be consistent with the depth map. After that, the pixel positions and 2D coordinates of human body's key points in RGB image are obtained based on tf-pose-estimation framework. Then the recognized 2D joint points are mapped to the depth map to obtain the accurate 3D joint point coordinates, so as to realize the 3D pose recognition of multiple people.

2.1. Three Dimensional Conversion Model

We established the pinhole model of camera imaging. In this model, there is a linear relationship between the spatial coordinates of the object and the image coordinates, so the solution of camera parameters is reduced to solving linear equations. The model involves four coordinate systems: world coordinate system, camera coordinate system, image coordinate system and pixel coordinate system. Through the relationship of these four coordinate systems, the coordinates of the related nodes in any coordinate system can be calculated. The relationship between coordinate systems is shown in Figure 1.

![Figure 1. The relationship between the four coordinate systems](image)

In pinhole model, the pixel system represents the coordinates of a point in the image in terms of pixels. Taken pixel value as the element, the image is stored in the form of matrix. Coordinates of a point in pixel coordinate system are shown as row and column labels of the corresponding elements in the matrix. The image coordinate system represents the coordinates of a point in the unit of physical length, and the position relationship with the pixel coordinate system is shown in figure 3, where \((u, v)\) and \((x, y)\) represent the coordinate in pixel coordinate system and in image coordinate system sequentially. \(O_i\) is the intersection point between the camera optical axis and the image plane, which is defined as the origin of image coordinate system. If the physical size of each pixel on the \(X, Y\) axis is represented as \(dx\) and \(dy\), the relationship between the two coordinate systems is shown in equation (1).
CCME 2020
Journal of Physics: Conference Series 1815 (2021) 012044 doi:10.1088/1742-6596/1815/1/012044

\[
\begin{bmatrix}
u \\
v \\
1 \\
\end{bmatrix} = \begin{bmatrix}
\frac{1}{dx} & 0 & u_0 \\
0 & \frac{1}{dy} & v_0 \\
0 & 0 & 1
\end{bmatrix}\begin{bmatrix} x \\
y \\
1 \\
\end{bmatrix} \quad (1)
\]

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c \\
1 \\
\end{bmatrix} = \begin{bmatrix}
f & 0 & 0 \\
0 & f & 0 \\
1 & 0 & 1
\end{bmatrix}\begin{bmatrix} X_c \\
Y_c \\
Z_c \\
1 \\
\end{bmatrix} \quad (2)
\]

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c \\
1 \\
\end{bmatrix} = \begin{bmatrix}
[0 & 1] \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}\begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1 \\
\end{bmatrix} \quad (3)
\]

\[
\begin{bmatrix}
u \\
v \\
1 \\
\end{bmatrix} = \begin{bmatrix}
\frac{1}{dx} & 0 & u_0 \\
0 & \frac{1}{dy} & v_0 \\
0 & 0 & 1
\end{bmatrix}\begin{bmatrix} f & 0 & 0 \\
0 & f & 0 \\
0 & 0 & 1
\end{bmatrix}\begin{bmatrix}
\frac{R}{1} & \frac{1}{1} & \frac{1}{0} \\
0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}\begin{bmatrix} X_w \\
Y_w \\
Z_w \\
1 \\
\end{bmatrix} \quad (4)
\]

In figure 3, camera coordinate system takes camera optical center \(O\) as the origin, axes \(X_c, Y_c\) are parallel to the axes \(X, Y\) in the image coordinate system, and axis \(Z_c\) is the optical axis of the camera, which is perpendicular to image plane. The intersection of axis \(Z_c\) and image plane is \(O1\). Marked focal length of the camera as \(f\), the matrix operation relationship between camera coordinate system and image coordinate system is shown in equation (2).

![Figure 2. Correspondence between image and pixel coordinate system](image)

**Figure 2.** Correspondence between image and pixel coordinate system

The world coordinate system is a rectangular coordinate system established with a point in space as the origin. The relationship between camera coordinate system and world coordinate system can be described by rotation matrix \(R\) and shift vector \(L\) in equation (3).

![Figure 3. Correspondence between image and camera coordinate system](image)

**Figure 3.** Correspondence between image and camera coordinate system

According to the principle of kinematics, we can calculate coordinate transformation matrix, and the mapping relationship between position on image and space can be established.

### 2.2. Camera Calibration

In order to solve the camera internal parameters, camera calibration is required.

Our Kinect camera is equipped with two lens, one is for depth and the other is for RGB value. \((C_x, C_y)\) are the aperture centers of the camera, and \((f_x, f_y)\) are focal lengths of lens\(^7\).

In the equation (5), \(M\) is called the internal parameter matrix, the values in which are only related to the internal parameters of a camera and do not change with the position of the object. \(S\) is the scaling factor of the depth map. In this case, point \((u, v, d)\) and its corresponding spatial coordinate relationship \((x, y, z)\). The transformation relationship between a point \((U, V, d)\) in camera coordinate system and three-dimensional coordinate point \((x, y, z)\) is shown in equation (6)(7)(8).

\[
M = \begin{bmatrix}
\frac{1}{dx} & 0 & u_0 \\
0 & \frac{1}{dy} & v_0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix} f & 0 & 0 \\
0 & f & 0 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
\frac{f}{dx} & 0 & u_0 \\
0 & \frac{f}{dy} & v_0 \\
0 & 0 & 1
\end{bmatrix} \quad (5)
\]
The human body can be abstracted as the line between two adjacent points. In this way, the human movement can be represented by the spatial coordinates of these joint points, the coordinates of which are the computer model of human movement. Because there is certain distance between the depth camera and RGB camera, the depth image obtained should be registered, and the depth image coordinate system should be converted to the RGB image coordinate system.

\[ u = \frac{f_x x}{z} + C_x \]  
(6)  
\[ v = \frac{f_y y}{z} + C_y \]  
(7)  
\[ d = z \times s \]  
(8)

The pixels of a color image are represented as \((u_R, v_R, z_R)^T\), \(u_R, v_R, z_R\) represent the abscissa, ordinate and depth value of color image in pixel coordinate system and camera coordinate system respectively. The pixels of the depth image are \((u_L, v_L, z_L)^T\), following the similar meaning of color image. Then matrix \(W'\) relates these two images in equation (9).

\[
\begin{bmatrix}
  u_R \\
  v_R \\
  1
\end{bmatrix}
= W' 
\begin{bmatrix}
  u_L \\
  v_L \\
  1
\end{bmatrix}
\]  
(9)  
\[
Z_L 
\begin{bmatrix}
  u_L \\
  v_L \\
  1
\end{bmatrix}
= 
\begin{bmatrix}
  f_x & 0 & u_0 & 0 \\
  0 & f_y & v_0 & 0 \\
  0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
  X_L \\
  Y_L \\
  Z_L \\
  1
\end{bmatrix}
\]  
(10)

The transformation between pixel coordinate system and camera coordinate system is known which is shown in equation (10). Making use of relationships above, an equivalent transformation relation is analysed. Then, internal parameter matrix of camera can be solved, shown in equation (11).

\[
Z_L 
\begin{bmatrix}
  u_L \\
  v_L \\
  1
\end{bmatrix}
= 
\begin{bmatrix}
  f_x & 0 & u_0 & 0 \\
  0 & f_y & v_0 & 0 \\
  0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
  X_L \\
  Y_L \\
  Z_L \\
  1
\end{bmatrix}
\]  
(11)
3. Action Mapping Method

3.1. Transformation of Visual Data and Control Data

Generally, the boundary conditions of spatial analytic geometry are so complicated that it is difficult to test and verify acquired results. With the help of space vector, the constraint of boundary conditions can be avoided, and the spatial relationship between vectors can be directly used to obtain the pose relationship, which greatly simplifies the calculation process.

There are seven degrees of freedom on human arms, but different types of robots and manipulators have different degrees of freedom. Therefore, it is necessary to allocate the human body degrees of freedom and joint angle to the robot in order to ensure the normal realization of functions[10].

In the process of semi-autonomous control, the joint vector is directly calculated according to the position coordinates of each part on the object based on image recognition, and then the joint angle is solved by space vector method.

![Figure 5. Calculation of yaw/pitch/rolling degree of freedom](image)

**Figure 5.** Calculation of yaw/pitch/rolling degree of freedom

The yaw degree of freedom corresponds to the motion of the robot elbow joint. In the plane of elbow, wrist and shoulder, vector $\overrightarrow{ES}$ and $\overrightarrow{EW}$ are formed. The yaw angle of the left arm is calculated with equation (14)(15)(16).

$$\overrightarrow{ES} = (SX - EX, SY - EY, SZ - EZ)$$

$$\overrightarrow{EW} = (WX - EX, WY - EY, WZ - EZ)$$

$$\cos\theta_1 = \frac{\overrightarrow{ES} \cdot \overrightarrow{EW}}{|\overrightarrow{ES}| |\overrightarrow{EW}|}$$

The pitch degree of freedom determines the joint angle of the humanoid robot's boos. With space vector method, the component of $\overrightarrow{ES}$ in $XOY$ plane is taken, and the angle between $\overrightarrow{ES}$ projection and $Y$ coordinate axis in positive direction is calculated, which is yaw joint angle.

$$\overrightarrow{ES} = (SX - EX, SY - EY, SZ - EZ)$$

$$\overrightarrow{ES}_{XOY} = (SX - EX, SY - EY, 0)$$

$$\text{n1} = (0,100,0)$$
The rolling degree of freedom determines the state of whole arm rotating along the shoulder-to-shoulder straight line. With space vector method, the angle with the positive direction of Y axis can be calculated by the outer product of $\mathbf{E}S$ and $\mathbf{E}W$.

$$\mathbf{E}S = (SX - EX, SY - EY, SZ - EZ)$$  \hspace{1cm} (21)$$

$$\mathbf{E}W = (WX - EX, WY - EY, WZ - EZ)$$ \hspace{1cm} (22)$$

$$\mathbf{n1} = \mathbf{E}S \cdot \mathbf{E}W$$ \hspace{1cm} (23)$$

$$\mathbf{n2} = (0, 100, 0)$$ \hspace{1cm} (24)$$

$$\cos \theta_3 = \frac{\mathbf{n1} \cdot \mathbf{n2}}{|\mathbf{n1}| |\mathbf{n2}|}$$ \hspace{1cm} (25)$$

When $\mathbf{n1}$ is on the either side of the body at the same angle, it is impossible to distinguish these two actions. Therefore, it is necessary to determine the relationship between the y-coordinate of the shoulder and of the elbow on depth direction.

The calculation method of the right arm is similar, but the direction of the outer product of the two vectors is opposite to that of left arm, so the rotation degree of freedom in Right arm should be in the opposite direction.

In the actual operation process, the data of calculating always fluctuates. Based on the precision limitation of camera and steering gear, it is necessary to select an appropriate filtering method with low memory requirement and high operation efficiency to process the fluctuating data[9]. In this paper, the limiting filtering method is used for the experiment.

According to above ideas, this paper defines the zero position of six angles--the position of each joint when each angle calculated is zero. Therefore, in the calibrating of joint value and actuator value, it is necessary to check the two kinds of data. After that, the information of joints is sent to the Arbotix-M control board to control the robot to complete the motion.

### 3.2. Control of Humanoid Robot

Because the angle change is used to control the motion of the robot's arms, a kind of private server click-the actuator, is selected as the representative of the actuator to elaborate the motion control principle of the robot's arms[11]. The process is shown in the figure 6.

**Figure 6.** Robot motion control flow

After mapping from the human arm to robots' arms, the relevant data of the arm is obtained through the sensor, calculated and processed by the upper control computer, which is transmitted to the control unit and the control signal.

**Figure 7.** Working principle of steering gear
When the steering gear receives the signal from the controller, the pulse width is compared with the pulse width generated by the internal circuit to get two pulses, one is output to the drive circuit, and the other is used to control the driving direction. The rotation of the motor drives the position of the potentiometer to change the pulse width from the future until it is equal to the pulse width of the external signal output. When the motor stops rotating, the rudder angle is fixed. Its schematic diagram is roughly as shown in Figure 7. The steering gear is controlled to a fixed angle, so as to implement the control of the robot arm.

4. Experiments and Results
To test the feasibility of the control method of humanoid robot, the experimenter made a series of arm raising and arm swinging movements. Kinect camera recorded and obtained the depth video and color video. Then, the video was processed into action images according to the frame. After that, we used tf-pose-estimation to get the 3D joint coordinates of arms. The joint values of joints in each image were obtained with space vector method. According to the mapping relationship between humanoid robot and real arms, arm action is mapped to humanoid robot continuously in real time. By comparing the action consistency between humanoid robot and human arm, whether the method is feasible or not can be verified.

In order to verify the feasibility of the method, the design experiment is as follows: experimenter changes his arms’ movement, and the robot arm uses the method above to complete the action change. By analyzing the similarity of the actions of the experimenter and the robot, the feasibility of the humanoid robot control scheme in this paper can be determined.

![Figure 8. Following images of Skeleton model and humanoid robot](image)

The experimenter made six groups of postures, and the humanoid robot followed him. The experimental results of semi-autonomous control are shown in Figure 8, where the humanoid robot is at the bottom left of each image and the experimenter is on the right.

Comparing the action between experimenter and the humanoid robot, we found that the postures of the experimenter were consistent with the robot's, and the error was small. In this semi-autonomous control experiment, it is found that the humanoid robot can follow the action of humans, which proves that the motion control method of the humanoid robot guided by the human’s actions is...
feasible.

5. Conclusions
In this paper, a control method of humanoid robot mapping human arm based on machine vision is proposed. Firstly, we used the RDB-D camera to obtain arm action, so skeleton model is extracted by tf-pose-estimation and the 3D coordinates of the arm are obtained to obtain the action data. Then, the humanoid robot model and the arm model are established, with which the arm joint values are calculated by the space vector method. By establishing a mapping relationship between real arms and humanoid robot arm, the joint values of the robot arm are obtained. Finally, raspberry system and Arbotix-M are used to transfer the control command to the servo actuator, so as to implement the control of the robot arm. The semi-autonomous experimental results show that the mapping method proposed in this paper is feasible. The future research work includes two aspects: one is to further improve the process of human joint recognition and operation capture, and improve the accuracy and instantaneity of robot motion control; the other is to extend the robot motion guided by human motion from the two arms of humanoid robot to the whole robot, so as to improve the intellectualization of humanoid robot.

Acknowledgments
We would like to thank Associate Professor Zhehan Chen for providing some ideas for our research as well as encouraging and helping us when encountering difficulties. And thanks to Beijing University of science and technology for providing the experimental equipment.

References
[1] Jianyu Chen. Research on gait planning of biped robot Research on parameter identification and gait planning of quadruped robot. 2020. Yanshan University, MA thesis.
[2] Qiuyan Yang. Research on humanoid robot servo Grasping object Based on Kinect. 2017. Xihua University, MA thesis.
[3] Xiuli Yu, Shiming Wei and Qizheng Liao. “Development and Technology Research of Humanoid Robot.” Journal of Mechanical Engineering 45.03(2009):71-75. doi:.
[4] Liang Zhang, Zhihao Cheng, Yixin Gan, Guangming Zhu, Peiyi Shen and Juan Song, "Fast human whole body motion imitation algorithm for humanoid robots," 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), Qingdao, 2016, pp. 1430-1435, doi: 10.1109/ROBIO.2016.7866528.
[5] Yiming Yang; Wolfgang Merkt; Henrique Ferrolho; Vladimir Ivan; Sethu Vijayakumar, "Efficient Humanoid Motion Planning on Uneven Terrain Using Paired Forward-Inverse Dynamic Reachability Maps," in IEEE Robotics and Automation Letters, vol. 2, no. 4, pp. 2279-2286, Oct. 2017, doi: 10.1109/LRA.2017.2727538.
[6] Zhifu Wang, Xianfeng Yuan,and Chengjin Zhang. "Design and Implementation of Humanoid Robot Behavior Imitation System Based on Skeleton Tracking". Proceedings of the 31st China Conference on control and decision making(3).Ed., , 2019, 874-879.
[7] Hailiang Zhang. The application research on target recognition and tracking based on image and motion-control robot. 2016. Southwest Jiaotong University. MA thesis.
[8] Kim, Ildoo. "Ildoonet/Tf-Pose-Estimation". Github, 2019, https://github.com/ildoonet/tf-pose-estimation. Accessed 4 Dec 2019.
[9] Yufei Zang. The research on semi-autonomous control system of manipulator based on machine vision. 2018. Southeast University. MA thesis.
[10] Wenbo Li. Intelligent motion training system based on body sense recognition. 2018. Hangzhou Dianzi University. MA thesis.
[11] He Zhu. Research on key problems of action imitation of humanoid robot based on Kinect. 2019. Yanshan University. MA thesis.