Conference contribution:
http://dx.doi.org/10.1109/ICARM.2016.7606908

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.
http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/
Abstract—In this paper, an online tracking system has been developed to control the arm and head of a Nao robot using Kinect sensor. The main goal of this work is to achieve that the robot is able to follow the motion of a human user in real time to track. This objective has been achieved using a RGB-D camera (Kinect v2) and a Nao robot, which is a humanoid robot with 5 degree of freedom (DOF) for each arm. The joint motions of the operator’s head and arm in the real world captured by a Kinect camera can be transferred into the workspace mathematically via forward and inverse kinematics, realistically through data based UDP connection between the robot and Kinect sensor. The satisfactory performance of the proposed approaches have been achieved, which is shown in experimental results.

I. INTRODUCTION

With the rapid development of robot technology, the application of teleoperation with robots for social activities has been widely used. The field of research with respect to motion capture of human based on tracking system attracted great attention during the past decades [1]. In [2], the author used a Microsoft Kinecct sensor intelligently to schedule a platform of multiple RGBD cameras especially in the Bayesian object tracking system. The kinematic validity of a Kinect based skeletal tracking system has been evaluated, which is designed to use with an upper limb virtual reality rehabilitation system [3]. In [4], a work on tracking the fingertips and palm center were developed based on Kinect sensor. However, researches focusing on head following is also necessary.

Head following technology has a wide range of application, especially in the video face recognition authentication, security, surveillance and human-robot interaction [5]. Due to various factors such as, the low resolution of human face in video, illumination change and motion blur factors, the head following involves a number of challenges [6]. According to [7], a multilayer neural-net (NN) controller for limb following is developed. In [8], a research team presents a model-based method for analyzing a human body motion in order to follow a robot arm model which represents a human body motion.

In recent years, the Kinect sensor with depth information has increasing number of applications in some special areas, such as, the game entertainment, health and fitness [9]. In early 2014, Microsoft company developed a new generation of sensor, which is Kinect v2, while providing Kinect development kit SDK, which makes Kinect v2 can collect more accurate color information with depth, become more suitable for the head following project [10].

NAO robot has been widely investigated owing to its opening programming framework [11]. In the field of scientific research, researchers use the Nao robot in computer science, mathematics, physics and artificial intelligence [12]. At the same time, people develop the programming of robots into the education with Nao robot because of its convenience, for example, one Swedish university encourages students to understand the knowledge of voice recognition and visual feedback through the establishment of laboratory with Nao robots [13]. In commercial field, Nao robots have been used in some large production launches to highlight the high technology of the products [14].

II. PRELIMINARIES

The control system consists of two main parts: Kinect v2 camera and the Nao robot. Fig.1 shows the control block of the system, which has been implemented to verify the performance of our approach.

A. Kinect v2

Kinect v2 produced by Microsoft, is an RGB-D device, which can capture depth, colour, and IR images (also sound) [15]. The depth, IR and colour image resolutions are significantly improved from that of the first generation Kinect v1. Raw colour, depth and IR images can be captured from the device using the Kinect 2.0 SDK (Software Development Kit) from Microsoft. Using the SDK, captured colour and depth information can be consolidated (transformed) into real-world co-ordinates, called Camera Space; these co-ordinates are referenced to the centre of the depth sensor [16]. Skeletal tracking can also be achieved with the use of the Kinect 2.0 SDK; although this feature is not used in this
project, it could be used by other researchers to track the position of an operator in front of device. Since the point specied by the operator on the Kinect colour image is 2 dimensional, the Kinect depth space image is required to determine the depth of this point. Both the colour image, and depth image transformed into a common frame, the origin of which is located at the centre of the depth camera, called Camera space; the co-ordinate system of this space follows a right-handed convention [17](as shown in Fig.2).

Fig. 2. Illustration of the origin of the Kinect v2s Camera space; the same as the depth sensor origin, modified from [18]

B. Nao Robot

In this project, we choose the Nao robot (shown in Fig.3), which is produced by the Aldebaran-Robotics in France and it is a humanoid robot with 5 DOF joints [19]. The Nao robot supports multiple sensors and controllers, such as, the head and jaw cameras, the chest sonar sensors, movement motors on the neck, hands and feet, three colour LEDs (in red, green and blue) on the eyes, and the head and feet tactile sensors [20]. NAO has a body mass index (BMI) of about 13.5 kg/m², hence, compared to other robots with the same height, Nao is quite light [20]. According to [21], Nao has a total of 25 degrees of freedom (DOF), wherein 11 joints are for the legs and pelvis and the rest are for the trunks, arms and head. In addition, each arm is supported by a 2 DOF shoulder, a 2 DOF elbow, a 1 DOF wrist and a 1 DOF hands gripper. The head is able to rotate on both yaw and pitch axes [21].

Fig. 3. Illustration of Nao Robot and coordinate system

III. CALIBRATION

The two angles of head for the Nao robot can not be used directly, which needs to be placed into the same coordinate system as the one read from Kinect, where we use the coordinate transformation. We use the following transformation as shown in (1) and then we can obtain Nao’s coordinate transformation from the robot coordinates to the Kinect coordinates.

\[ X_i = TX'_i \]  \hspace{1cm} (1)

where T represents the transformation matrix, we will give it concept as following \( X_i = [x_i, y_i, z_i, 1]^T \), \( X'_i = [x'_i, y'_i, z'_i, 1]^T \), which are the coordinates under both the robot and Kinect systems.

We can retain the transformation matrix T by calculating four points under both the robot and Kinect coordinates systems. Giving the assumption there are four points \( p_1, p_2, p_3 \), whose coordinates are under the robot coordinate system and the Kinect coordinate system respectively, Fig.4 shows an example for one of the four points in relationship between the Nao robot and the Kinect coordinate system. (2) is the equation to calculate the transfer matrix T from the Kinect coordinates to the robot coordinates.

Fig. 4. Illustration of relationship between Nao Robot and Kinect coordinate system

\[
T = \begin{bmatrix}
    x_1 & x_2 & x_3 & x_4 \\
y_1 & y_2 & y_3 & y_4 \\
z_1 & z_2 & z_3 & z_4 \\
1 & 1 & 1 & 1
\end{bmatrix}^{-1}
\]  \hspace{1cm} (2)

In this project, we choose four positions under the same point on the central point of right hand for Nao robot which are placed on both Nao coordinate system and Kinect system as shown in Fig.5. To obtain the positions under Kinect coordinate system, we arrange a person stand in front of a Kinect sensor, which is used to obtain the coordinate data of human’s right hand, and keep his right hand place on the same position as the right hand of Nao robot. The four positions under Nao robot coordinate system are \( (1, 1, 1), (2, 0.5, 0.5), \), \( (1, 0.5, 1.5), (1, -2, -3) \), under Kinect coordinate system are \( (-0.0790, 0.3053, 0.3908), (0.0600, 0.2180, 0.5385), (-0.0012, 0.4000, -0.0703), (0.2364, -0.1443, 0.3090) \).

Substituting all the values of positions both on Nao and Kinect system into (2) we can calculate the transformation matrix of Kinect under ground based coordinate system \( T_{0g} \), where we assumed the coordinate system of Nao robot is ground based because the Nao robot keeps stationary in this project.
\[ T^k_0 = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & -0.5 & 0.5 & -2 \\ 1 & 0.5 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -0.079 & 0.06 & -0.0012 & 0.2364 \\ 0.3053 & 0.218 & 0.4 & -0.1443 \\ 0.3908 & 0.5385 & -0.0703 & 0.309 \end{bmatrix}^{-1} \]

Then we would obtain:

\[ T^k_0 = \begin{bmatrix} 7.87 & 5.09 & 2.37 & 6.58 \\ -10.27 & -0.4 & -0.73 & -1.26 \\ 1.25 & 9.58 & 1.09 & -0.39 \\ 0 & 0 & 0.0001 & 4.72 \end{bmatrix} \]

After that the rotation matrix of Kinect under ground based coordinate system \( R^k_0 \) can be obtained through its transformation matrix, which is given as following:

\[ R^k_0 = \begin{bmatrix} 7.87 & 5.09 & 2.37 \\ -10.27 & -0.4 & -0.73 \\ 1.25 & 9.58 & 1.09 \end{bmatrix} \]

IV. KINEMATICS METHODOLOGY

The image stream is captured from the Kinect v2 colour camera, such as the operators view of the Nao workspace. In order to select a point in the colour image which the operator desires to send the data of orientation angles for user’s head, the angles need to be transformed into those under the basic ground coordinate, which is shown commonly in Fig.6. In this experiment, we keep the flat of Kinect camera be always parallel of the user’s head, where the value of \( \beta_f^k \) equals zero. Then we have the matrix as follow, which represents rotation matrix:

\[ R^f_k (\alpha^k_f, \beta^k_f, \gamma^k_f) = \begin{bmatrix} \alpha^k_f c^k_f c^k_f + \beta^k_f c^k_f \gamma^k_f - \alpha^k_f c^k_f \gamma^k_f \\ \gamma^k_f c^k_f c^k_f + \beta^k_f s^k_f \gamma^k_f - s^k_f c^k_f \gamma^k_f \\ s^k_f c^k_f c^k_f + \beta^k_f s^k_f \gamma^k_f - c^k_f s^k_f \gamma^k_f \end{bmatrix} \]

According to \( R^f_0 = R^k_0 \times R^f_K \), wherein the \( R^k_0 \) has already been obtained through calibration process, as shown in (5), the rotation matrix \( R^f_0 \) for user’s head under the ground base coordinate system could be found by (7), where we denote the abbreviations \( s\alpha^k_f \) for \( s_a \), \( s\beta^k_f \) for \( s_b \), \( s\gamma^k_f \) for \( s_g \), \( c\alpha^k_f \) for \( c_a \), \( c\beta^k_f \) for \( c_b \), \( c\gamma^k_f \) for \( c_g \).

Finally, the orientation angles of user’s head under the basic ground coordinate can be calculated, which are shown in (9)-(11), where we give the abbreviation:

\[ a^0_f = \arctan \frac{r_{21}}{r_{11}} \]

\[ \beta^0_f = \arctan \pm \sqrt{\frac{r_{11}^2 + r_{21}^2}{r_{31}}} \]

\[ \gamma^0_f = \arctan \frac{r_{32}}{r_{33}} \]
\[ R_0^T = \begin{bmatrix}
7.87c_a c_b - 10.27(c_a s_b s_g + s_a c_g) + 1.25(c_a s_b c_g + s_a s_g) & 5.09c_a c_b - 0.4(c_a s_b s_g + s_a c_g) + 9.58(c_a s_b c_g + s_a s_g) \\
7.87s_a c_b - 10.27(s_a s_b s_g - c_a c_g) + 1.25(s_a s_b c_g - c_a s_g) & 5.09s_a c_b - 0.4(s_a s_b s_g - c_a c_g) + 9.58(s_a s_b c_g - c_a s_g) \\
2.37c_a c_b - 0.73(c_a s_b s_g - s_a c_g) + 1.09(c_a s_b c_g + s_a s_g) & -5.09s_b - 0.4c_b s_g + 9.58c_b c_g \\
2.37s_a c_b - 0.73(s_a s_b s_g + c_a c_g) + 1.09(s_a s_b c_g - c_a s_g) & -2.37s_b - 0.73c_b s_g + 1.09c_b c_g \\
\end{bmatrix}
\]

\[ \theta_1 = \pi - \langle \overrightarrow{CO} \times \overrightarrow{CD}, \overrightarrow{CB} \times \overrightarrow{CO} \rangle \]  

\[ \theta_2 = \langle \overrightarrow{CD}, \overrightarrow{CD} \rangle \]  

\[ \theta_3 = \langle \overrightarrow{CB} \times \overrightarrow{CD}, \overrightarrow{CD} \times \overrightarrow{CE} \rangle \]  

\[ \theta_5 = \langle \overrightarrow{EE} \times \overrightarrow{ED}, \overrightarrow{ED} \times \overrightarrow{EH} \rangle \]  

\[ \theta_6 = \overrightarrow{ED} \times \overrightarrow{EF} \] 

\[ \theta_7 = \frac{\pi}{2} + \langle \overrightarrow{EH} \times \overrightarrow{EG}, \overrightarrow{ED} \rangle \]  

These seven angles are all in the joint space, and their initial positions are displayed in Fig.7. The geometry vector method we proposed is based on the principle of cosine value of two vectors shown in (17), in addition, the angle between two planes can be got by using their normal vector [22].

It is simple to find that angle \( \theta_1 \) is the supplement of the angle between plane COD and xOy.

\[ \theta_1 = \pi - \langle \overrightarrow{CO} \times \overrightarrow{CD}, \overrightarrow{CB} \times \overrightarrow{CO} \rangle \]  

Angle \( \theta_2 \) is the angle between \( y-axis \) and vector \( \overrightarrow{CD} \), which is given below:

\[ \theta_2 = \langle \overrightarrow{CD}, \overrightarrow{CD} \rangle \]  

In the same way, we can also get other angle: \( \theta_3 \) is the angle of plane BCD and CDE.

\[ \theta_3 = \langle \overrightarrow{CB} \times \overrightarrow{CD}, \overrightarrow{CD} \times \overrightarrow{CE} \rangle \]  

Angle \( \theta_5 \) is the angle of plane CDE and DEH.

\[ \theta_5 = \langle \overrightarrow{EE} \times \overrightarrow{ED}, \overrightarrow{ED} \times \overrightarrow{EH} \rangle \]  

Angle \( \theta_6 \) is the angle between vector \( \overrightarrow{ED} \) and plane EHF.

\[ \theta_6 = \overrightarrow{ED} \times \overrightarrow{EF} \] 

However, \( \theta_7 \) IS the yaw angle of wrist and it is difficult to calculate using the above method. \( \theta_7 \) can be seen as the angle between \( \overrightarrow{X_5} \) and \( \overrightarrow{Y_7} \). While,

\[ \overrightarrow{X_7} = \frac{\overrightarrow{EF}}{|\overrightarrow{EF}|}; \overrightarrow{Z_7} = \frac{\overrightarrow{EF} \times \overrightarrow{EH}}{|\overrightarrow{EF} \times \overrightarrow{EH}|}; \overrightarrow{Y_7} = \overrightarrow{Z_7} \times \overrightarrow{X_7} \]
So now the problem became to solve $\vec{X}_5$. We know that $\vec{X}_5$ must be in the plane of EFH, in addition, $\vec{X}_5$ is Perpendicular to $\vec{DE}$. Thus when suppose that:

$$\vec{X}_5 = k_1 \vec{EF} + k_2 \vec{EH}$$  \hspace{1cm} (20)

There are:

$$(k_1 \vec{EF} + k_2 \vec{EH}) \times \vec{DE} = 0; |k_1 \vec{EF} + k_2 \vec{EH}| = 1$$  \hspace{1cm} (21)

Therefore,

$$\theta_7 = \langle \vec{x}_5, \vec{x}_7 \rangle$$  \hspace{1cm} (22)

Up to now, all the joint angles can be calculated.

V. EXPERIMENTAL STUDIES

In this section, some experiments have been designed to test the performance of following system. Experimental platform: computer system is windows 10, Kinect SDK for windows, Visual studio 2013 and OpenCV library. The experiment environment is an indoor and adequate illumination environment. Only one person stands forward to Kinect in the distance about 2 meters. Because the degree of freedom for Nao robot’s arm is only 5, thus we choose the ShoulderPitch, ShoulderRoll, ElbowRoll, ElbowYaw and WristYaw joints from human arm to take the experiment.

A. Head Following Experiment

In this part, we run this experiment to test the stability of the head following system where we place a person in front of the Kinect camera. Then we run the program developed by the Kinect SDK 2.0 and send the command to an additional computer, which controls the Nao robot, to imitate the movement of human head. It has been shown in Fig.9, we run 4 times experiments with different postures in order to get the total performance, which are rotating 30° to the left direction, rotating 30° to the right direction, pitching up with 30° and pitching down with 30° respectively. The Fig.10 below shows the process of the first experiment. As a result, a graph can be drawn (seen in Fig.11), wherein the data has been collected from head angles for both operator and Nao robot. From the result of head following experiment, the movement of user’s head can be followed accurately.

![Fig. 9. Illustration of operator’s head for different postures](image)

From the first test above, we are able to conclude that the user’s head with different postures can be captured quite smoothly. The Nao robot imitates the movement of human head well. The result illustrates that there is a fast and accurate response to follow the user’s head and imitate the familiar movement with human in different postures.

B. Limb Following Experiment

In the second experiment, the operator keeps his body standstill and only puts his limb forwards at 90° and raises up at 180° and stretches his limb sidewards at 90°, with a low speed from the origin position to final position, for both of his right and left arms. Then we use the same method in the first experiment, after obtaining the data, we send it to the additional computer, which controls the Nao robot, to imitate the movement of human limb. The Fig.12 below shows the process of the second experiment. As a result, a graph can be drawn (seen in Fig.13), wherein the data has been collected from arms for both operator and Nao robot. From the result above, we can conclude that all the movements of those 5 selected joints from user’s arm can be followed completely.

VI. CONCLUSION

A system for tracking user’s head and arm to interact with Nao robot via a Kinect v2 sensor has been developed in this paper. Kinect v2 is controlled to track the head’s and arm’s motion of the human operator. The operator can move their head and arm in a natural way to manipulate the position of the Kinect device. Additionally, the kinematics equations
carried with this research have been applied to transfer the Euler angles of user’s head from the Kinect coordinate to the ground coordinate and calculate all the joints angles of human arm. We also run experiments and draw the curve according to the test data. The experimental result shows that the Nao robot imitates the human movement accurately. In the near future, we would focus on teleoperation of Nao robot with whole body and leg following would be added to the project.

VII. ACKNOWLEDGEMENT

We would like to thank Junshen Chen for his technical assistance.

REFERENCES


