

Interfacing of Kinect Motion Sensor and NAO Humanoid Robot for Imitation Learning

Melissa Guo, Shuvajit Das, Jacob Bumpus, Esubalew Bekele, and Nilanjan Sarkar

KEYWORDS. Robotics, autism spectrum disorder, robot imitation learning

BRIEF. This research focuses on using the Xbox Kinect sensor for robot imitation.

ABSTRACT. Autism spectrum disorder (ASD) is considered a public health emergency in the United States and affects millions of children. Children affected by ASD generally have difficulty interacting in social environments. Traditional intervention, which involves sessions with therapists, is costly and time-consuming. However, it has been shown that some children with ASD interact better with robots than with humans; compared to humans, robots express no emotion, which ensures that children with ASD are not overwhelmed by their interactions. In order to advance human-robot interactions specifically for autism treatment by creating a relatively new method in which humans and robots can interact, a robot imitation learning platform was developed using the Xbox 360 Kinect and the Aldebaran NAO humanoid robot. The robot imitated movements made by a human based on full body tracking data from the Kinect sensor in real-time. The angle measurements tracked by the Kinect were also compared to angle measurements from an inclinometer and it was shown that the NAO robot can successfully imitate human actions and gestures within its joint and workspace limits. This development will eventually lead to a viable treatment method for children with autism.

INTRODUCTION.

In the U.S., autism spectrum disorder (ASD) is considered a public health emergency with a current estimate of 1 in 88 [1] prevalence rate. ASD is generally characterized by impairments in social interaction, social communication, and repetitive behavioral patterns [2]. Children affected by ASD often have difficulty communicating both verbally and non-verbally, such as in facial expressions and body language [2]. They also have trouble in some social interaction scenarios (e.g. sharing emotions, understanding how others think and feel, and holding a conversation).

Unfortunately, traditional intervention is costly with the average lifetime cost of autism estimated to be around \$3.2 million. The average medical expenditures for individuals with ASD are estimated to be about 4-6 times greater than for those without ASD [2]. Therapy (which could involve behavioral and speech-language therapy) often involves one-on-one, 40-hours-a-week private sessions with a specialist [2]. There are no approved medications specifically for the treatment of autism, but some can treat various associated symptoms. Currently, there are no medications that can alleviate or treat all symptoms of autism [3].

It is well documented that children with ASD, under some circumstances, have been shown to respond and interact better with robotic systems than with humans [1,3]. This may be due to the robots' more simplified and predictable nature, which may allow them to be less intimidating and confusing than humans— compared to humans, robots show no emotions. This implies leveraging these preferences towards robots appropriately might result in improved intervention of ASD. By using robots to help and treat children with ASD rather than using a human, rehabilitation with robots will ideally be faster and more effective than that with humans.

Previous work with human-robot interaction (HRI) has shown that robotic systems are capable of interacting with children with ASD [1-6]. In one such study, robots were programmed with the ability to recognize affective states using peripheral physiological body signals. In the same study, children with ASD played a game of basketball using a robot-controlled basket. Based on the child's

affective states (such as engagement anxiety and liking), the robot would adjust game difficulty accordingly [4].

Previous studies have shown the feasibility of using the Microsoft Kinect to track body position and orientation and to visually represent the data with a skeletal model of the person being tracked [7-9]. A set of lines are used to show the position of the head, torso, and limbs on a screen. In addition, studies involving robot imitation learning have shown that robots are capable of imitation and mimicking when a learning algorithm is applied [10-11]. For example, Lopes *et al.* taught a robotic arm to swing a ball-in-cup by repetitive imitation learning. After the robot completed "learning" the motion, it had more accuracy and precision in swinging the ball into the cup than a human, producing a higher, more efficient success rate [5]. Using such approaches of connecting a robot with an outside system, the Kinect sensor will be interfaced with a humanoid robot to develop a robot imitation system that functions in real-time. These developments will not only advance and improve interaction between robots and humans and provide a method of interfacing readily-available technologies, but will also pave the way for future research to more effectively use robotic systems to aid in technological advancements and to further research in the medical field.

MATERIALS AND METHODS.

The system is a robot-mediated imitation learning platform. The Xbox Kinect (Figure S1) uses a PrimeSensor that "enables the xBox to perceive the gamer's environment in three dimensions and to translate these perceptions into a synchronized depth image" and consists of a PrimeSense PS1080 SoC chip, 3D depth sensors (IR light source and CMOS image sensor), and a RGB camera (color image sensor) [6]. Sensory information, including depth image, color image, and audio, is then transferred back to a console [6]. The robot used is the Aldebaran Robotics NAO robot (Figure S2). NAO is fully programmable and is equipped with different sensors and actuators, rendering it capable of whole-body movement, face and object recognition, and automatic speech recognition [7].

There are software modules for the robot side as well as for the sensor (Kinect) side. Modules on the Kinect side are written in C# and modules on the robot side are written in python. Modules in both sides communicate in real-time via a network interface. The Kinect tracks body position and orientation and stores the data as three-dimensional Cartesian coordinates relative to the console (Figure S3). A software module on the Kinect side then converts the data to joint angles, the angle of a joint between two limb effectors, by calculating the angle between the two vectors in space and determining the orientation of the effector.

Joint angles are the format NAO uses to read and interpret data. A set of joint angles form a "key frame", which is all of NAO's joint angles for any given position. Key frames are then combined to form "motion frames", which is essentially the movement. This is the basic data structure that was used when coding for both the Kinect side and the robot side of the system (Figure 1). Using the client-server network interface, the data is sent from the Kinect side to the NAO side, where a python module will read and interpret the data file. NAO will then execute the movements in real-time.

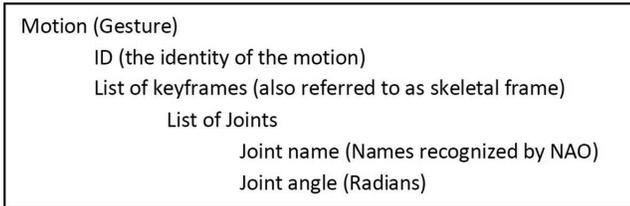


Figure 1. This is the basic data structure used to code for the system. Starting at the bottom, it is seen that each joint consists of a name (“Joint name”) and an angle (“Joint angle”). Several joints (recall that each joint has its own name and an angle) make a “keyframe” (“List of joints”), which represents the robot’s body position at a certain point in time. A set or “list” of keyframes comprises a motion, which also has an “ID”, or name. Therefore, to code for a motion or gesture, it is given a name and a set of keyframes. (Time or duration is not a parameter: the robot is given a certain speed at which to transition from keyframe to keyframe and, essentially, move. This speed is represented as some floating, or decimal, number value between 0 and 1, where 1 is the robot’s maximum possible speed.

Precautions were taken when developing the software modules to avoid self-destructive movements and behaviors by the NAO. For example, collision detection was incorporated to prevent the NAO from hitting itself or its own limbs, which could result in damage. The joint angle data was also smoothed, which includes removing anomalies and in data, to avoid large and dramatic changes in direction, which could damage NAO’s joints and could cause it to function beyond its parameters. This was done at each key frame by evaluating whether the proposed joint angle was within a radius of one epsilon of the previous point. If the point were not within the radius, for example if there was a spike or anomaly in the data, then that point would be discarded and not executed to avoid erratic and potentially dangerous motions.

The accuracy of the Kinect tracking system when tracking certain joints was based on a study by Kar [8]. A trained subject performed ten trials each of six different rotations, which were either 45° or 90° rotations around the shoulder or elbow joints. These accounted for 3 different joint rotations with two different angles of rotation for each joint. The actual joint angles were measured by attaching a digital inclinometer, to the limb that is changing position and measuring the angle difference relative to a “zero position,” which is any initial position, chosen at the start of the motion. The angle change measured by the Kinect was found by referencing the output of the Kinect console on the computer. The differences in angles between the start positions, or “zero positions,” and the final positions collected by the Kinect were compared to the actual angle differences measured by the inclinometer.

RESULTS.

The average difference between the Kinect angle measurements and the inclinometer measurements for each rotation are shown in Figure 2. For the 90-degree rotations, the left shoulder pitch (LShoulderPitch) and right elbow yaw (RElbowYaw) both had relatively low differences (less than 5 degrees) between the Kinect measurements and the digital inclinometer’s measurements, meaning a visual representation of the two measured angles would look the same to the human eye. The right elbow roll (RElbowRoll) had the highest difference of any rotation measured.

Results were not similar for the 45-degree rotations (Figure 2). All three 45-degree rotations had large differences in measurements, with the left shoulder pitch having the highest difference and the right elbow roll having the lowest. There did not seem to be any correlation between the different angle measures or the different joint rotations.

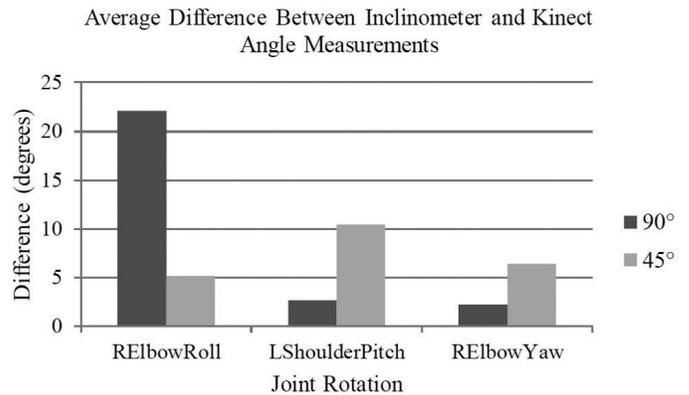


Figure 2. The average difference between the Kinect angle measurements and the inclinometer measurements are represented for the joint rotations.

While there were errors in the joint angle measurements by the Kinect, it did not perceptibly affect the NAO’s execution of the movements and gestures to the point where the movements were unrecognizable as the movements originally performed. Due to the NAO’s joint and workspace limitations, the robot cannot always execute the exact motions of a human in front of the Kinect. Calculations must be made by the Kinect to modify the joint angles to accommodate for the NAO’s capabilities. Therefore, slight variations in joint angle calculations are not expected to significantly affect the NAO’s behavior.

More importantly, the NAO successfully mirrored movements that were performed by a human and tracked by a Kinect. The NAO imitated the positions and orientations of wrist, elbow, shoulder, and head joints in real-time. There was a very slight lag from the time the subject performed the motions to the instant the NAO executed the movements. For example, in an arm-raise motion, the NAO generally completed the motion less than a second after the human subject completed the motion. Although the lag was noticeable, it did not affect the overall accuracy of the movements over time.

In Figure 3, the NAO is standing in a similar position to the Kinect skeletal tracker. On the left is the Kinect visual skeletal tracker showing the body in the same position as the NAO robot, which is executing the movement on the right.

Figure 3. The Kinect visual representation of the skeletal tracker is shown on the left with green lines and dots representing the limbs and joints. The NAO is shown on the right. The Kinect skeletal tracker and the NAO are shown in similar body positions. The NAO is executing the original movement, shown by the Kinect, which was performed by the human. Note that the bottom half of the legs on the Kinect skeletal tracker are not highlighted in green because they were not being tracked at the time. Instead, the Kinect extrapolates where the legs/feet might be.

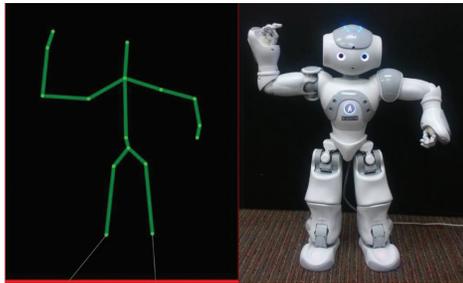
DISCUSSION.

The differences between the angles measured by the inclinometer and the angles sensed and calculated by the Kinect can be expected because there are also slight variations in joint angles among human gestures. While some of the differences are meaningful, it has been found that they do not detract from the intended movement and that there appears to be no apparent correlation between angle measure of rotation and degree difference. Miniscule differences would not affect the overall interpretation of a movement. That is, even though there are changes in the angle measurements, the movement performed by the robot is recognizable as the same motion performed by the human. If this was not the case, and the movement was not recognizable, then the changes in joint angles would have a drastic effect on the overall result. For example, if a motion performed by the NAO was intended to be a wave, but was instead unable to be recognized as a wave or was perceived to be a different motion because of the angle changes, then the system would not be able to successfully communicate with a human through common gestures. Significant angle differences in crucial joints of the NAO’s ex

ecution of a movement could possibly affect how the movement is interpreted, depending on the necessary motions and arm trajectories.

The bounds (safety parameters), that prevent the NAO from conducting a motion beyond its workspace limits and from colliding with itself ultimately alter the angle measurements sent to the robot, making it perform slightly different movements to prevent possible damage to itself. These alterations have the possibility of changing a movement to the point of being unrecognizable. For example, when testing the collision detector, joint angles were assigned in a motion where the NAO mimics a clapping, which involves the two hands colliding with each other. Instead of stopping short of each other, the hands maneuvered so that one hand was above the other hand. Such drastic changes affect not only the trajectory of the motions, but also the implied meaning of the gestures.

The Kinect sensor was successfully interfaced with the NAO robot to create a robot imitation platform. The robot was able to imitate human movements in real-time. The overall success of system has many implications for the future of human-robot interactions. The development of the system successfully integrates two available and popular



technologies, eliminating the need to create an entirely new system with similar capabilities. The use of a remote sensor allows for more variability in terms of spatial confines and usage areas, and adds the convenience to choose where the sensor gathers data and where the robot responds.

This project will ultimately be taken further to develop an autonomous system to treat a child with autism. A learning algorithm will be added to the Kinect side to facilitate autonomous recognition of and reaction to movements and gestures done by the child. A subject would first “teach” the robot gestures, such as waving hello, to add to a bank of known gestures. Then, the robot’s learning capabilities would be tested by performing motions similar, but not identical to, motions already taught to the NAO and observing whether the robot is able to correctly classify the movement as a certain type of gesture by referencing a pre-programmed library and respond appropriately. This will allow the robot to better respond to the child’s body language and as a supplement to responding to explicit motions.

The current system is a step towards a robot-centered rehabilitation approach for children with autism. With more capabilities, the system will not only advance human-robot interaction technology, but also will further the goal of creating a viable treatment method for autistic children.

ACKNOWLEDGMENTS. Shuvajit Das, Jacob Bumpus, Esubalew Bekele, Dr. Nilanjan Sarkar, Dr. Chris Vanags, School for Science and Math at Vanderbilt, and the Center for Science Outreach. The project described was supported by Award Number R25RR024261 from the National Center For Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Research Resources or the National Institutes of Health.

SUPPORTING INFORMATION.

Figure S1. Microsoft Xbox360 Kinect [6]

Figure S2. Aldebaran NAO humanoid robot [7]

Figure S3. Model of the robotic arm [12]

REFERENCES.

1. “New Data on Autism Spectrum Disorders,” CDC (2012).
2. Shriver E, “Autism spectrum disorders (ASDs),” (2011).
3. Liu C, *IEEE Trans Robotics*. 24, 883 (2008).
4. Liu C, Conn K *et al*, *Proc IEEE Int Conf Robot*. Roma, Italy (2007).
5. Lopes M, Melo FS *et al*, *Proc IEEE IROS*. San Diego, CA, USA (2007).
6. “Kinect for Windows,” Microsoft, 2012. [Online].
7. “NAO For Research,” Aldebaran Robotics. [Online].
8. Kober J, Peters P, *IEEE Int Conf Robot*. Kobe, Japan (2009).
9. Ende T, Haddadin P *et al*, *Proc IEEE IROS*. San Francisco, CA, USA (2011).
10. Sigalas M, Baltzakis H *et al*, *Proc IEEE IROS*. Taipei, Taiwan (2012).
11. Dorsey R and Howard AM, *Proc IEEE ICALT*. Athens, Greece (2011).
12. Kar A, Available: <http://home.iitk.ac.in/~akar/Projects.html>.
13. Sigalas M, Baltzakis H *et al*, *Proc IEEE IROS*. St. Louis, MO, USA (2009).
14. Bockemühl T, Troje NF *et al*, *Hum Move Sci*. 29, 73 (2012).
15. Robins B, Dautenhahn K, *et al*, *Proc. CWUAAT*. 225 (2004).
16. Dautenhahn K, Werry I, *Autism*. 1 (1980).



Melissa Guo is a student at Martin Luther King Jr. Magnet High School in Nashville, Tennessee, and enrolled in the School for Science and Math at Vanderbilt.