Presented at the DLSU Research Congress 2014
De La Salle University, Manila, Philippines
March 6-8, 2014

Developing Robot Motion Models and Gesture Recognition System for a Human-Robot Interaction System to assist in Play Therapy

Kathy Hong[1,*], Paul Jerome Lim[1], Dennis Earl Talplacido[1], Jerome Cyruss Macaspac[1] and Ana Marian Pedro[1]

[1] *Computer Technology Department, College of Computer Science, De La Salle University*

**kathy_hong@dlsu.edu.ph*

Abstract: Studies show that children with low-functioning Autism Spectrum Disorder (ASD) have problems with their motor skills and social development. Play therapy is a technique used to aid the development of children with special needs through play, as play helps children develop cognitive, social and perceptual skills. This study aims to assist therapists in the process through a follow-the-leader type game with an interactive robot system. This paper tackles the results of two main experiments: the data extraction of natural human motion as basis of robot motion model and the evaluation of Hidden Markov Model (HMM) gesture recognition algorithm to enable human-robot interaction. Natural movement is defined as a robot trying to achieve human-like motions as close as possible. This process can be done by collecting a total of fifteen sample video of human actions, tracking specific joints and manipulating the robot's actuator such that the velocity, acceleration, and torque are controlled based on the action provided. The primary goal is to identify whether the robot can reach arm accelerations of $5m/s^2$ to $20m/s^2$ and $-5m/s^2$ to $-20m/s^2$. The arm should also reach elbow angles ranging 35^o to 50^o during the transition of the robot's arm toward its body and 86^o to 116^o during the transition away from the body. The values are taken from the conducted experiments to produce the desired output with natural human movement. A direct attempt to reproduce the same movement patterns from the data collected from the test subjects produced a lot of jitter as the robot tries to move as fast as possible such as reaching the peak acceleration of $20m/s^2$ and $-20m/s^2$. Aside from producing natural movement, the system also needs to detect and recognize gestures that are given by the child. Gesture recognition requires machine learning algorithm such as HMM to be able to train gestures efficiently. To evaluate the HMM gesture recognition algorithm, a total of 65 gesture data per gesture were collected to use for training and testing. The HMM algorithm yielded low accuracy rate on determining the Lateral Raise gesture wherein only one out of five Lateral Raise were correctly classified while Hand Wave has four out of five. The result could be influenced by the small training set or the complexity of the gestures which made it hard to attain more than 50% accuracy rating.

Key Words: machine learning; gesture recognition; natural robot motion; human-robot interaction; play therapy

1. INTRODUCTION

Children diagnosed with developmental disorders such as Attention Deficit Disorder (ADD), Attention Deficit Hyperactive Disorder (ADHD) and Autism Spectrum Disorder (ASD), need specialists and therapists to help hone and sharpen their basic motor and social skills. Traditional therapies suggested for patients with developmental disorders, such as ADD/ADHD and autism, can be costly, inefficient and stressful both

for the family and the patients themselves (Parker-Pope, 2010). Currently, social robots and therapeutic robots are used in many therapies to assist autistic children in their learning. These robots are disguised as toys to welcome interaction from children.

Some of the most significant social robots or therapeutic robots made and/or marketed are Keepon (Bever, 2007), Paro (2004), Nao (Burns, 2012) and Popchilla (Hamilton, n.d.). Keepon and Popchilla are robots that are mainly designed for autistic children. Keepon can dance along with any music. It would shake, twist and crouch on the beat of the sound. Popchilla can speak, move its ears and tails, and change the color of its eyes to express its feelings. On the other hand, Nao is a programmable robot that can also be designed to teach autistic children like normal humans and have many features that the children enjoy like memory games. While Paro is a seal robot that is used for elderly care to assist senior citizens in coping up with loss, loneliness and depression. It can express its feelings by blinking its eyes and moving its head by mimicking a real seal. Aside from that, the robot can also respond to names given to it.

This research aims to develop a robot platform disguised as a common stuffed toy to help children diagnosed with these disorders by acknowledging their needs through play. A follow-the-leader game is identified for the basis of the interaction flow design based on consultation with an occupational therapist. With the use of an RGB-D camera sensor, the robot gathers feedback from the user, which is the child, then formulates and executes commands based from the data gathered. Taking advantage of the RGB-D camera sensor, parents or therapists can monitor the performance and development of their child or patient. Through this research, children with ADD/ADHD or ASD can develop their motor and social skills through play.

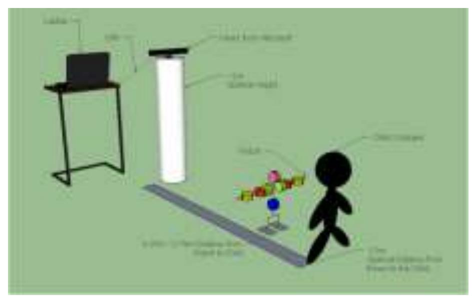

Fig. 1. System Diagram

In order to accomplish the objectives of the project, robot motion models and gesture recognition algorithms are evaluated. The project aims to develop a robot motion model that will be based on natural human movements to respond to gestures accurately recognized by the system. This paper presents the results of benchmarking to determine the performance of the robot motion and performance of Hidden Markov Model (HMM) gesture recognition algorithm to evaluate the classification of user gestures which are presented in Section 3.

1.1 Robot Motion Models

Controlling robot movement such that the robot moves as natural and as human-like as possible requires one to be able to quantify the characteristics present in a human movement. Based on data collected from observing human movement, the robot must be able to follow the movement as close as possible given that it has its own limitations in terms of degrees of freedom, speed, and acceleration. (Harada et al., 2006).

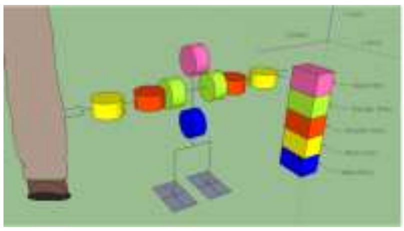

Fig. 2. Robot Model

The robot consists of motors that are able to support its own weight with additional load from the robot skin. The chosen actuator for the robot is the Dynamixel AX-12. It can output 1.50 Nm of torque and runs at 9-12V that produces 59 rpm without the load. Fig 2 shows the mechanical frame of the robot. The robot has two joints in its limbs with a total of 3 degrees of freedom; the head along with the waist each have 1 degrees of freedom. The robot's possible movements are determined and limited by its moving parts. The movements that the robot can perform is determined by the Robot Movement Set which is the possible responses of the robot to the child. Two types of experiments are performed in order to quantify some natural movement characteristics. The motion models are derived from the results of the experiments.

1.2 Gesture Recognition Algorithms

Gesture recognition is an essential element of the system which assists the interaction between the user and the robot. Through the use of the RGB-D sensor, researchers can obtain both RGB and depth data which can detect a person in an environment and perform full body tracking by tracking the skeletal joint positions of a person in real time. Microsoft has provided a programming toolkit to use with the Kinect RGB-D sensor. Windows Software Development Kit is capable of tracking the skeletal 3D model of the human target and obtain data from the joint positions (Zhang, et al., 2012).

Gesture recognition is an important aspect in human-robot interaction. There are many available approaches for gesture recognition such as statistical modeling, computer vision and pattern recognition, image processing. The most used approach is the statistical modeling such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs). According to a survey, HMM is one of the most used tools for gesture recognition among all other tool (Mitra et al., 2007).

Hidden Markov Models (HMM) (Bünger, 2013; Mitra et al., 2007; Wilson et al., 1999) is a very popular and efficient in modeling spatial-temporal information. It is claimed to be one of the best approach for gesture recognition. It works by training a model with sequences. Classification of the input sequence is done by computing the likelihood to each gesture class models. The classified gesture is then obtained by maximum likelihood where in the similarities of the input sequence and the model is above some threshold. In this way, HMM is capable to adapt to any gestures and sequences given optimal observed parameters for gestures. HMM uses Forward–Backward algorithm for evaluation, Baum–Welch algorithm for training or estimation of models and Viterbi algorithm for decoding. HMM has a weakness which assumes that current observations are statistically independent of the previous observations (van Kasteren et al, 2008).

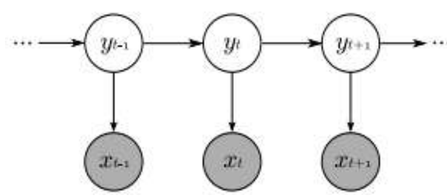

Fig. 3. Hidden Markov Model (HMM) Graphical Representation. Non-shaded nodes as hidden variables and shaded nodes as observable variables

In an experiment conducted by Kelly et al. on 2009, HMM and CRF models were evaluated and compared in recognizing sign language gestures. Based from the data gathered, HMM is claimed to have better performance than CRF in terms of identification of movements and classification of gestures. It was also observed that even though the assumption of independence is a weakness of HMM, a threshold HMM model trained on appropriate parameters can beat CRF models in sign language detection.

2. METHODOLOGY

2.1 Data Extraction of Natural Human Motion as a Basis for Robot Motion Model

Two experiments were performed in order to characterize movements by human subjects and the robot. The first experiment aims to record accelerations of the subjects' arms while they move. The experiment is conducted by asking the subject to perform a hand wave gesture while holding an accelerometer. Additional information such as the angles of the subjects' elbows are taken from the video playback. Markers are placed behind the subjects (Fig 5) in order to determine the distances travelled by the subjects' arm from their body. The two sets of markers are placed horizontally and vertically with 10 cm and 25 cm intervals respectively (Fig 5). The video camera is placed 240cm away from the markers and 120cm above the ground. The purpose of video recording is to compare the acceleration values from the Wiimote, a 14.81cm x 3.63cm x 3.07cm, 136.08g (Aol Tech, 2013) accelerometer sensor used in the experiment, and the actual movements of the subject during the experiment. The elbow angle values serves as limitations that will be placed in the robot. The subject is asked to gesture the hand wave for six seconds and the recording is repeated three times per subject. The acceleration values detected by the accelerometer are sent to the laptop to be recorded and interpreted.

The second experiment is to survey the movement of the robot arm's acceleration in performing a hand wave and its elbow angle. The robot is programmed using a proprietary software from ROBOTIS called the RoboPlus Motion. The RoboPlus Motion is a software to edit motion files saved on the robot's controller where the task code and speed data necessary for robot movements are found (ROBOTIS, 2010). The task code is a set of actuator position values that are programmed in order to achieve the desired robot movement. The same procedure is done with the first experiment although the distance between the video camera is 79cm far and 33cm high. It was also taken into consideration that the weight of the accelerometer will be added to the weight of the arm of the robot. The added load will increase the torque requirement of the shoulder joint from 0.1825Nm to 0.2828Nm. The new torque requirement of the shoulder joint is still well within the torque capacity of the AX-12 motors. The extra load provided by the accelerometer and the battery does not change the performance of the robot provided that the battery is fully charged.

2.2 Hidden Markov Model Training and Recognition Experiment

Hidden Markov Model Toolkit (HTK) is a software toolkit mainly designed for building HMM-based speech recognition. HTK can also be used on gesture recognition as it is also general-purpose wherein HMM can model any time series data. Gesture and Activity Recognition Toolkit (GART) is a user interface toolkit that was designed around HTK. GART helps facilitate gesture-based application.

For the experiment, GART was used and modified to help facilitate the gesture recognition module. The data collection used an RGB-D sensor Kinect for Xbox 360 to track the movement of the subject's right hand. Six subjects were requested to perform *Hand Wave* and *Lateral Raise* (Fig. 3) with their right arm for approximately eleven times which made a total of sixty five *Hand Wave* and sixty five *Lateral Raise* gesture training data. The training data is composed of x-, y-, and z-axes points of the hand movement which were obtained using Robot Operating System (ROS). ROS, an open source framework for writing robotic software, is being used as the developmental framework for the experiment. It includes tools and packages that enables developers to record and playback gathered data within ROS.

Fig. 4. (a) *Hi* and (b) *Lateral Raise* Right Hand Gestures

The training set is composed of 60 gesture data per gesture while the test set is composed of 5 gesture data. The training set served as the input to the GART software toolkit for training and building model. After training, the software is ready for recognition. Each test gesture data is processed by evaluating the maximum likelihood of the input data using the trained model. The most similar gesture is then considered to be the recognized gesture.

3. RESULTS AND DISCUSSION

3.1 Subject Motion Experiment Results

(a) (b)

Fig. 5. Hand Wave Motion Data Experiment Setup of Person 1

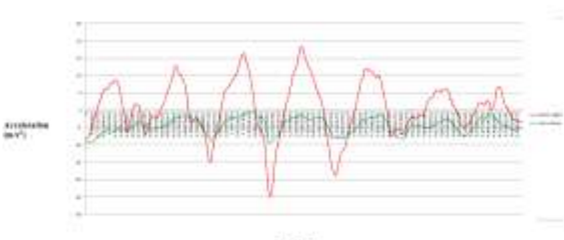

Fig. 6. Person 1 Arm Acceleration Graph (Range: $25 m/s^2$ to $-25 m/s^2$)

Table 1. Person 1 Elbow Angles

Low points of “left to right”	Value in Degrees	Peak points of “left to right”	Value in Degrees
start point	46º	1st peak	86º
1st low point	30º	2nd peak	88º
2nd low point	35º	3rd peak	100º
3rd low point	36º	4th peak	96º
4th low point	40º	5th peak	91º
5th low point	38º	6th peak	94º

(a) (b)

Fig. 7. Hand Wave Motion Data Experiment Setup of Person 2

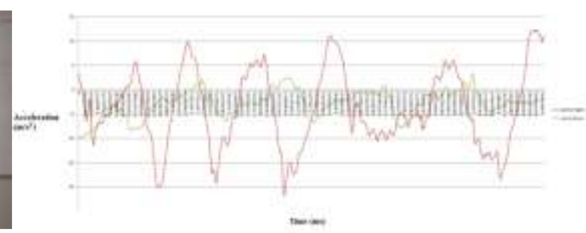

Fig. 8. Person 2 Arm Acceleration Graph (Range: $+10 m/s^2$ to $-25 m/s^2$)

Table 2. Person 2 Elbow Angles

Low points of “left to right”	Value in Degrees	Peak points of “left to right”	Value in Degrees
start point	50º	1st peak	116º
1st low point	38º	2nd peak	103º
2nd low point	43º	3rd peak	103º
3rd low point	36º	4th peak	105º
4th low point	38º	5th peak	105º

5th low point	38º	6th peak	110º

The two graphs on Fig 6 and Fig 8 are the data gathered from the accelerometer. The graphs show the range of the acceleration for the hand wave movement from Person 1 is $20m/s^2$ to $-25m/s^2$ and Person 2 is $10m/s^2$ to $-25m/s^2$. The data found in Table 1 and Table 2 are the angles made from the elbow of the person illustrated in Fig 5 (a) and Fig 7 (a) when it hits the low point of the graph and Fig 5 (b) and Fig 7 (b) for the peak of the graph. The graph reaches its low point when the angles produced by the person's arm moving upward is in the range of 35º to 50 º. The graph reaches its peak point when the angles produced by the person's arm moving downward is in the range of 86 º to 116 º. This determines the maximum and minimum angles done by the elbow of the person doing the hand wave. There is a difference between the number of peaks per trial since the trial is recorded for six seconds and each test subject performs the hand wave inconsistently.

Fig. 9. Hand Wave Motion Data Experiment Setup of the Robot (Range: $+10m/s^2$ to $-25m/s^2$)

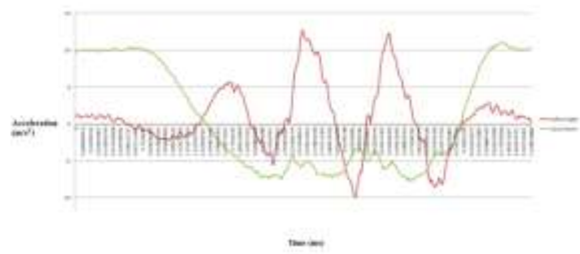

Fig. 10. Robot Arm Acceleration Graph (Range: $15m/s^2$ to $-10m/s^2$)

Two solutions to determine the robot movement were implemented. One is created based on elbow angles of one of the subjects and another is made by making the robot do the action while reducing the jitter effect on its body. Based on the data obtained from Fig 6 and Fig 8, the acceleration required by the robot to mimic the a hand wave of a person is approximately $5m/s^2$ to $20m/s^2$ and $-5m/s^2$ to $-20m/s^2$. The joints that will be controlled the robot are the elbow joint and the shoulder joint. The angle of the arm and forearm of the robot which is connected by the elbow joint on its low point will be 40° and on its peak will be 100°.

3.2 Hidden Markov Model Performance Result

Table 3. Hidden Markov Model Training Confusion Matrix

	Lateral Raise	Hand Wave
Lateral Raise	9	9
Hand Wave	1	17

Based from the output result of the program, only 18 from each gesture data were classified out of 60 gesture data which makes out only 30% of the input data. The training module classified half of the *Lateral Raise* gesture correctly while the other half was incorrectly classified as *Hand Wave*. Furthermore, majority from the *Hand Wave* were correctly classified, only one was misclassified.

Table 4. Hidden Markov Model Gesture Recognition Result

Gesture	Number of training data	Number of test data	Accuracy Rate
Hand Wave	60	5	80%
Lateral Raise	60	5	20%

For the gesture recognition module, *Lateral Raise* gesture data were mostly incorrectly classified as *Hand Wave* - only one out of 5 *Lateral Raise* was correctly classified. As for the *Hand Wave* gesture data, 4 out of 5 were correctly classified by the program. In total, the accuracy rate of the HMM gesture recognition is 50% which is relatively low from what was expected. Based from the training result, it can be pointed out that most of the *Lateral Raise* input data were incorrectly classified as a *Hand Wave* because the model was trained with more *Hand Wave* than *Lateral Raise*.

It is also important to take note that the data collection settings made for this experiment was subscription to the left hand and right hand only. This settings was made to avoid noise to be encrypted in the gathered gesture data. More subscribed joints will yield to noisier gesture data output.

4. CONCLUSIONS

Data for the robot motion models vary greatly between different subjects such as the frequency of their hand waves per experiment. Some data are unusable due to the fact that the subject used only wrist movements during the hand wave. There were problems in directly using the data from human subjects due to the difference in the proportion, scale, and weight distribution of humans and the robot. The data obtained will still need further processing in order to account for the said discrepancies between robot and human body structure. Future studies may focus on applying better actuator control and additional robot joints to see if these would result in more natural movements. Improvement on the robot structure such as increasing the degrees of freedom in different joints may be added.

Hidden Markov Model has its weakness of the independence assumption which may be one of the factor for its relatively low accuracy rate. One factor can also be the capacity of the training set or the ratio of the training set to test set. In machine learning, the system should be also well trained by injecting more training data but at the same time, the system should not be over-trained as well. Conditional Random Fields (CRF) is one of the available alternative gesture recognition algorithms. CRF eliminates the independence assumption, allowing re-evaluation of previous observations. On the other hand, CRF can result to low accuracy rate caused by inevitable variability of natural human gestures on every individuals (Elmezain et al., 2011).

5. ACKNOWLEDGMENTS

The authors would like to thank Ms. Jocelyn Cu for helping and assisting the proponents on the HMM machine learning algorithm and gesture recognition.

6. REFERENCES

(2004). Seal-Type Robot "PARO" to Be Marketed with Best Healing Effect in the World. AIST. Retrieved from http://www.aist.go.jp/aist_e /latest_research/2004/20041208_2/20041208_2. html

Aol Tech (2013). Nintendo Wii Remote Controller. Retrieved from: http://www.engadget.com/ products/nintendo/wii/remote-controller/specs/

Bever, C. (2007). Robots with rhythm could rock your world. *New Scientist*. Retrieved from http://www.newscientist.com/article/dn11434-robots-with-rhythm-could-rock-your-world.html

Bünger, M. (2013). Evaluation of Skeleton Trackers and Gesture Recognition for Human-robot Interaction. Denmark: Aalborg University.

Burns, J. (2012). Robots in the classroom help autistic children learn. *BBC News*. Retrieved from http://www.bbc.co.uk/news/education-20252593

Elmezain, M., Al-Hamadi, A., Sadek, S., & Michaelis, B. (2011). Robust Methods for Hand gesture Spotting and Recognition Using Hidden Markov Models and Conditional Random Fields. *IEEE*, 131-136.

Hamilton, A. (n.d.). Can Robot Help Autistic Kids Connect? *Co.Exist*. Retrieved from http://www.fastcoexist.com/1680618/can-robots-help-autistic-kids-connect

Harada, K., Hauser, K., Bretl, T., & Latombe, J. C. (2006, October). Natural motion generation for humanoid robots. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 833-839). IEEE.

Kelly, D., McDonald, J., & Markham, C. (2009). Evaluation of Threshold Model HMMs and Conditional Random Fields for Recognition of Spatiotemporal Gestures in Sign Language. *2009 IEEE 12th International Conference on Computer Vision Workshops*, ICCV Workshops, 490-497.

Mitra, S., & Acharya, T. (2007). Gesture Recognition: A Survey. *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, 37 (3), 311-324.

Parker-Pope, T. (2010). *The Cost of Autism*. The New York Times.

Ren, L., Patrick, A., Efros, A.A, Hodgins, J., Rehg, & J.M. (2005). A Data Driven Approach to Quantifying Natural Human Motion. *Robotics Institute*. Paper 288.

ROBOTIS (2010). RoboPlus Motion. Retrieved from http://support.robotis.com/en/software/roboplus/roboplus_motion_main.htm

vanKasteren, T.L., Noulas, A., & Krose, B.J. (2008). Conditional Random Fields versus Hidden Markov Models for activity recognition in temporal sensor data. In Proceedings of the 14th Annual Conference of the Advanced School for Computing and Imaging (ASCI'08).

Wilson, A., & Bobick, A. (1999). Parametric Hidden Markov Models for Gesture Recognition. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 21 (9), 884-900.

Zhang, H., Du, W. & Li, H. (2012). Kinect Gesture Recognition for Interactive System. Retrieved from http://cs229.stanford.edu/proj2012/ZhangDuLiKinectGestureRecognitionforInteractiveSystem.pdf