

**DETECTING ENGAGEMENT LEVELS FOR AUTISM
INTERVENTION THERAPY USING RGB-D CAMERA**

A Thesis
Presented to
The Academic Faculty

by

Bi Ge

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2016

COPYRIGHT © BI GE 2016

**DETECTING ENGAGEMENT LEVELS FOR AUTISM
INTERVENTION THERAPY USING RGB-D CAMERA**

Approved by:

Dr. Ayanna M. Howard, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Linda Wills
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Jun Ueda
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: April 29th, 2016

ACKNOWLEDGEMENTS

I would like to thank Prof. Ayanna Howard for her guidance in completing my thesis. I thank my mentor Dr. Hae Won Park for her advise and mentoring. I am grateful to my committee members: Prof. Linda Wills and Prof. Jun Ueda for agreeing to be on my committee. I would like to thank my fellow HumAnS lab mates, students, staff and professors of Georgia Tech for helping me complete this work. I also thank all the staff at Kid's Creek Therapy center, children, their parents and instructors for participating in this study. I thank my parents and family, without whom I would not be here. I thank my roommate Fu Hung. Last but not least, I thank Yihan for her support and wake-up calls.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF SYMBOLS AND ABBREVIATIONS	viii
SUMMARY	x
<u>CHAPTER</u>	
1 Introduction	1
2 Background	3
Autism Spectrum Disorder	3
Autism Intervention and Applied Behavior Analysis	3
Prompts and Its Categories	5
Prompt Fading	8
Robotics in Autism Research	10
Kinect for Autism Study	11
Detecting Engagement/Disengagement	11
3 Approach	14
Introduction to Kinect Camera	18
Methods for Data Pre-processing	18
Stream Annotation and Segmentation	18
Feature Definitions and Extraction	19
Summary	25

4	Experimental Setup	27
5	Results	30
6	Conclusion and Future Work	34
	REFERENCES	36

LIST OF TABLES

	Page
Table 1: Sample of CSV file containing joint coordinates captured by Kinect.	18
Table 2: Demographic data about participants.	28
Table 3: Accuracies for classifiers and feature sets combinations.	31
Table 4: Accuracy for feature set and classifier combination, trained with child interacting with therapist and tested on child interacting with robot.	33

LIST OF FIGURES

	Page
Figure 1: Example of an autism intervention, RGB image (left) and segmented image (right). The instructor (left) is issuing a gestural prompt.	4
Figure 2: Example of therapy scene and the skeletons recognized by Kinect.	5
Figure 3: Screenshot of the game running on the tablet used as therapy task. The start menu (top left), default game view (top right), the user found a matched pair (bottom left) and the user failed to find a matched pair (bottom right).	15
Figure 4: Therapy scene captured by Kinect. RGB image (top left), recognized skeletons (top right) and reconstructed 3D scene (bottom).	16
Figure 5: Flow diagram of optimal robot system for autism intervention.	16
Figure 6: Flow diagram of the system deployed in this work.	17
Figure 7: Photo of the tablet used throughout the sessions.	17
Figure 8: Distance of head joint to task in meters vs. no. of frames. Engagement clips (top) and disengagement clips (bottom).	22
Figure 9: Typical experimental setup where the instructor (left) and patient (right) are sitting together interacting on the task (game running on the tablet).	28
Figure 10: Experimental results from "hold-one-child-out" cross validation.	31
Figure 11: RGB (top left), reconstructed (top right) and skeleton view of session where the child interacts with the robot.	32
Figure 12: Accuracies for classifier and feature set combination, models trained with child interacting with therapist and tested on child interacting with robot.	33

LIST OF SYMBOLS AND ABBREVIATIONS

t_{start}		First timestamp in a CSV file
t_{remain}	All the timestamps in a CSV file except for the first one	
$J_{joint\ name}$		A joint on the human skeleton
θ		Leaning angle
P	Plane constructed using $J_{mid\ spine}$, J_{neck} and J_{head}	
D		Planar distance
\vec{c}_i		Vector representing child's joint i
\vec{t}_i		Vector representing therapist's joint i
\vec{b}		Vector representing center of sphere
\vec{T}		Vector representing the task (tablet)
$depth_{unit}$		Depth obtained from image
x_c, y_c		Index to a pixel on the depth image
f_{cx}, f_{cy}		Focal lengths in x and y directions
d_x, d_y	Tangential distortions in x and y directions	
c_{cx}, c_{cy}		Index into principal pixel
γ		Margin in SVM
w, b	Definition of a hyper-plane, i.e. $wx + b$	
y_i		i th label
$\phi(x^i)$		Kernel version i th input vector
x_1, y_1	Input feature vector and the corresponding label	
ASD		Autism spectrum disorder

ABA	Applied behavior analysis
PDD-NOS	Pervasive developmental disorder-not otherwise specified
PDD	Pervasive developmental disorders
SDK	Software development kit
EEG	Electroencephalogram
3D	3-dimensional
CSV	Comma separated value
SVM	Support vector machine
KNN	K-nearest-neighbor classifier
IRB	Institutional review board
HMM	Hidden Markov Model

SUMMARY

Our motivation for this work is to develop an autonomous robot system that is able to perform autism intervention therapy. Autism spectrum disorder (ASD) is a common type of neurodevelopmental disorder that affects millions of people in the United States alone. The best way of treating ASD and help people with ASD learn new skills is through applied behavior analysis (ABA, i.e. autism intervention therapy). Because of the fact that people with ASD feel less stressful in a predictable and simple environment compared to interacting with other people and autism intervention therapy provided by professional therapists are generally expensive and inaccessible, it would be beneficial to build robots that can perform intervention therapy with children without a therapist/instructor present.

In this research, we focus on the task of detecting engagement/disengagement levels of a child in a therapy session as a first step in designing a therapy robot. In this work, we mainly utilize an RGB-D camera, namely the Microsoft Kinect 2.0, to extract kinematic joint data from the therapy session. We also set up a child study with the Kid's Creek therapy center to recruit children with ASD and record their interactions with a therapist while working on a touch-screen based game on a tablet. After carefully selecting features derived from skeletons' movements and poses, we showed that our system could produce an accuracy of 97% when detecting engagements and disengagements using cross-validation assessment.

CHAPTER 1

INTRODUCTION

It has been shown by the Centers for Disease Control and Prevention that approximately 1 in 68 children in the United States are diagnosed with autism spectrum disorder (ASD) [1]. Research has found that for children diagnosed with ASD, the younger they enter an early ASD intervention program, the larger gains they may have in developmental skills [2]. It is also worth noting that as the duration of ASD intervention therapy increases, the effect, including acquisition of new skills and behaviors, also increases without diminishing returns [3]. There has been a trend in developing robots as therapy tools since robots have been identified as improving engagement as well as encouraging novel social behaviors for people with ASD [4]. Another reason behind deploying robots in autism intervention therapy is that therapy services provided by professional therapists are often expensive and inaccessible [5].

Aiming at increasing the availability of intervention services to children with ASD, a number of different alternative technologies have been proposed and evaluated for their performance in a therapy setting. One such technology consists of a social robot functioning as a therapist. Some examples of social robots used in autism intervention therapy include a humanoid used as both therapist and interactive toy designed to help children with ASD learn and practice social skills [6] and a mobile robot designed to stimulate reciprocal interaction such as imitative play [7].

When used in a traditional therapy setting, prompts serve as an intervention in which instructions and cues are issued by a professional therapist when a child appears distracted or to help the child gain/eliminate desired/undesired behaviors [8]. For example, prompts are usually used by a therapist to reengage a child when they are working on a task together and the child's attention shifts away from the task. Prompts are important

throughout the process of intervention therapy because children with ASD generally respond differently than typically-developing children. As a result, therapists tend to utilize prompts as extra stimulus.

One major advantage of incorporating robots into autism intervention therapy is due to the elevation in mood often observed in children with ASD interacting with a robot. Researchers have shown that when interacting with a therapist and a robot at the same time, both typically-developing children and children with ASD spend more time looking at the robot [8]. In addition, consistent, repeatable and standardized stimuli provided by the robot can help to build a standardized prompt system [9].

In order to develop such a low-cost robot platform that can perform autism intervention therapy autonomously, we look at the first step required to provide prompts: the ability to detect engagement/disengagement and to provide prompts at the right time.

CHAPTER 2

BACKGROUND

1.0 Autism Spectrum Disorder

Autism spectrum disorder (ASD) is recognized as a type of neurodevelopmental disorder that can have a broad range of severity in symptoms, affecting behavior, communication with others and other social interactions [10]. According to CDC, 1 in every 68 children is diagnosed with ASD in the United States alone [1]. The word “spectrum” in ASD suggests the variations in symptoms within and between each diagnostic category including autistic disorder, Asperger’s disorder and pervasive developmental disorder-not otherwise specified (PDD-NOS) [10].

Some clinical characteristics of ASD include impairment in social communications, impaired verbal communication, repetitive behavior and even cognitive and motor impairment [10]. Individuals with minimum cognitive impairments are recognized as “high functioning” [10].

2.0 Autism Intervention and Applied Behavior Analysis

In the past, there have been different philosophies in how Applied Behavior Analysis (ASD) should be applied and different intervention protocols have been derived from them. Some interventions include behavioral, developmental and cognitive-behavioral interventions [2]. Even though the strategies from different interventions may

differ, there are mainly two aspects which they all agree on: the age at which a child should enter an intervention program and the intensity of that program [2].

Among all the intervention types, applied behavior analysis (ABA) is probably the most well-known. There has been an unparalleled amount of research in favor of applying ABA in autism intervention based on its performance [11]. ABA is based on the belief that most social behaviors acquired by humans are learned throughout a history of interaction between this individual and his(her) environment and environmental cues and stimuli can influence how a person diagnosed with ASD behaves [12]. Furthermore, by applying positive and negative reinforcements, existing patterns can be altered and new ones can be

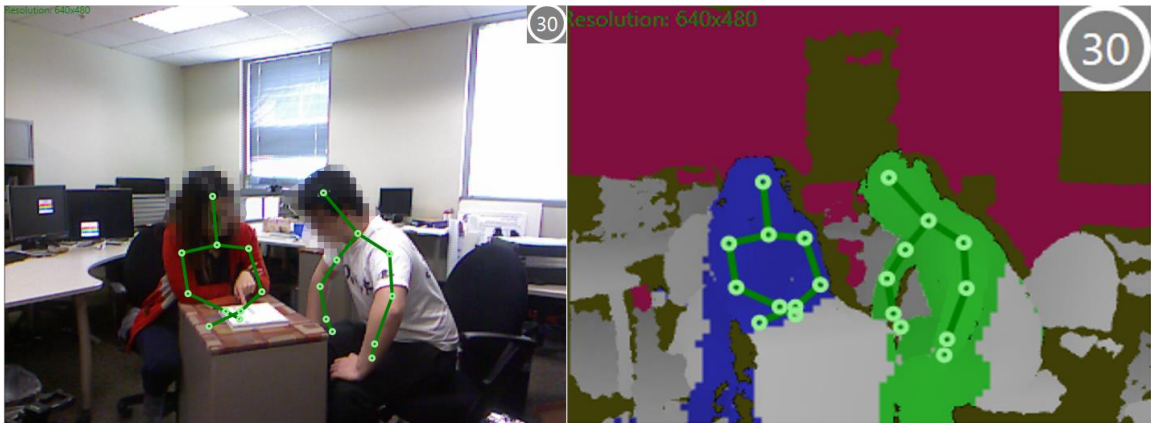


Figure 1. Example of an autism intervention, RGB image (left) and segmented image (right). The instructor (left) is issuing a gestural prompt.

established.

Figure 1 is an example of how intervention therapies are conducted. The instructor (on the left) is issuing a gesture prompt by pointing to the book (task object). The segmented image and recognized skeletons are shown on the right. Figure 2 is an image of an instructor and a child working together in a home setting with their skeletons shown.

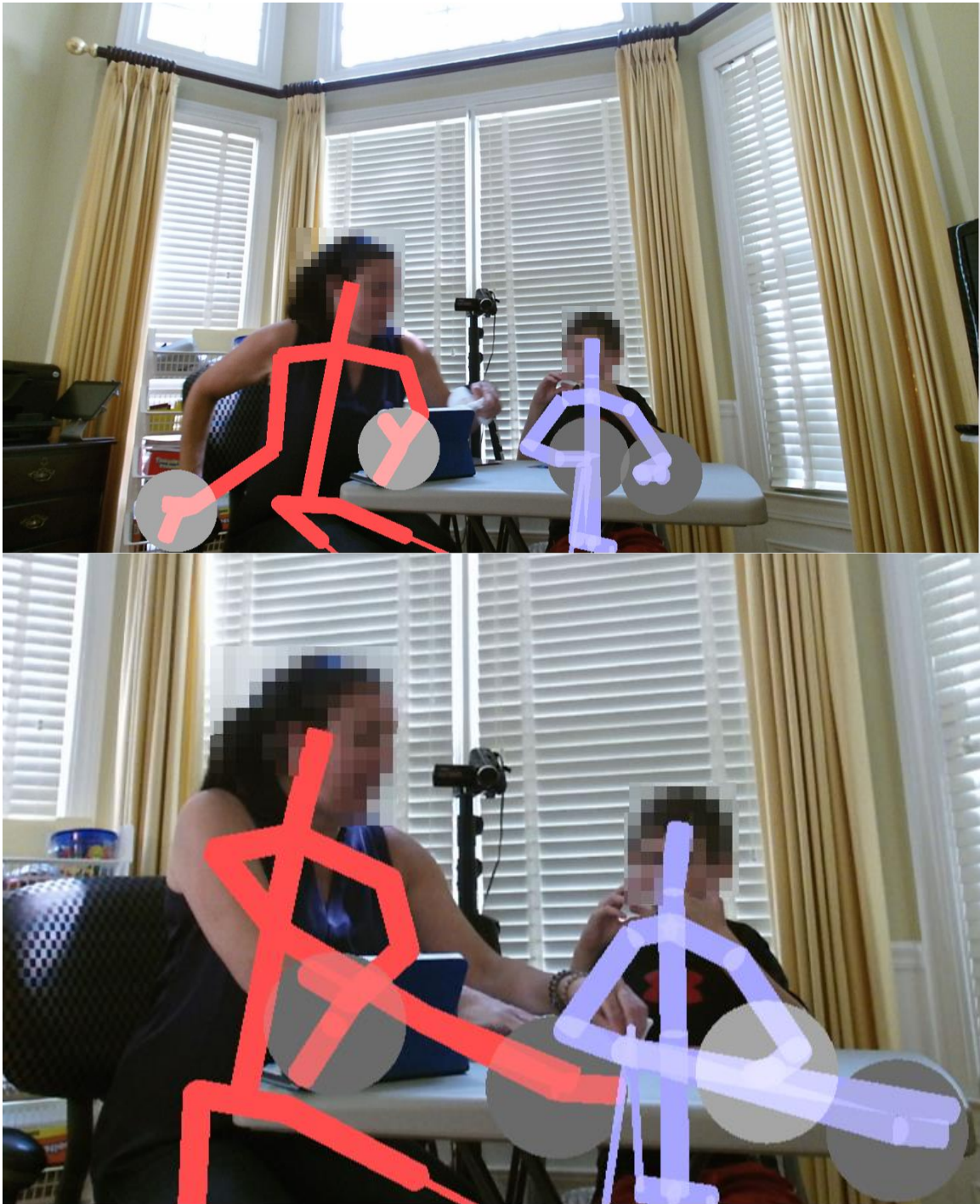


Figure 2. Example of therapy scene and the skeletons recognized by Kinect.

3.0 Prompts and Its Categories

Environmental stimuli, called prompts, are often defined as cues and hints that serve as extra stimuli to help people with ASD acquire the desired behaviors or remove undesired ones at the correct timing and correct environment [8]. For example, when asking a child “How old are you?”, to help this child answer this question correctly when asked this question outside the therapy/intervention setting, a parent/therapist will present a “verbal prompt” by saying “Say, ‘I’m 12 years old’”. Another example might be when teaching a child with ASD how to tie a shoe properly, the therapist might hold the child’s fingers to demonstrate the knot.

There are different kinds of prompts used in ABA, namely verbal prompts, modeling, manual prompts, gestural prompts, photographs and textual prompts [8].

Verbal prompts:

Verbal prompts are instructions or words issued to help and engage a child in a particular response [8] and are often used with other types of prompts [8].

Modeling:

Modeling refers to the intervention approach of demonstrating a response (either in person by a therapist or in video recorded from a therapist or peer) to a child [8]. The goal for modeling prompts is for the learner to mimic and imitate the presented skill/task. Some studies have shown that modeling can be the best performing prompt method when the learner and the model are similar [8]. For example, 8-year-old children with ASD may learn skills better and faster when those tasks and skills are demonstrated by other 8-year-old children.

Reference [8] also mentioned the instances where some people diagnosed with ASD can only learn to respond to a scenario exactly taught in modeling, i.e. the skill

demonstrated by modeling does not generalize well on some children as well as some adults with ASD.

Manual Prompts:

Manual prompts are often defined as physical contacts from an instructor/therapist [8]. As mentioned before, manual prompts can be used to teach a task that requires movement of the body (make a bed or tie a shoe). Manual prompts are usually used in conjunction with other forms of prompts and research has also shown that elevated engagement levels encouraged by manual prompts can sustain even after the therapist is no longer in sight [8].

Gestural Prompts:

Unlike manual prompts, which requires the therapist to physically touch the patient, gestural prompts only involves pointing, nodding or using body languages to prompt a child [8]. Like manual prompts, gestural prompts are also used in addition to other forms of prompting.

Photographs and drawings:

Daily routines such as meal preparation and laundry tasks have been taught to people with ASD using photographs or drawings [8]. Prompts involved with photographs have been shown to improve on-task time as well as reduce undesired behaviors [8].

Textual Prompts:

Textual prompts are cues and prompts written in words in forms of checklists and task lists [8]. Textual prompts can be easily combined with photographs to teach assembly tasks as well as daily chores as these tasks usually follow a step by step pattern [8].

4.0 Prompt Fading

Although prompts can help people with ASD acquire new skills and remove undesired behaviors, our ultimate goal is that people with ASD can perform acquired skills in the correct situation without the presence of prompts [8]. Thus, it is important to apply prompt-fading protocols so that patients can practice their skills without the help of prompts.

Least-to-Most Prompts:

As the name suggests, least-to-most prompts (also called increasing assistance) requires a therapist to only use prompts or increase the intensity of prompts when the student fails to respond to the natural stimuli (the one that controls the behavior and is present even without any prompt) within a time frame (usually 10 to 15 seconds) [8] and the instructor should increase his/her assistance until the child makes a correct response. One advantage of least-to-most prompts is that the child has an opportunity of completing the task without any prompts [8].

Most-to-Least Prompts:

In most-to-least prompts (also referred to as decreasing assistance), the child always receives prompts during the beginning of a session [8]. Only after a number of successful trials, the degree and intensity of prompts will be gradually decreased and eventually no

prompts will be present [8]. Most-to-least prompts are preferred because it is easy to implement in practice and can achieve stable rates of correct responses [8].

Delayed Prompts:

Delayed prompts involve first presenting the child with the natural stimuli that is present in the real world and controls the behavior and waiting for a period of time before presenting the prompts [8]. Delayed prompts have been shown to be effective and efficient in transferring stimulus from prompts to environmental cues [8] thus is often used in teaching new skills.

Graduated Guidance:

Graduated guidance is often related to most-to-least fading in manual prompts. For example, in the beginning the therapist may teach a task by hand-over-hand prompts, then switches to less forceful guidance, then fade the prompts from hand to wrist, elbow, shoulder and eventually no physical contact [8].

Stimulus Fading:

By exaggerating some physical dimensions of a natural stimulus, stimulus fading can help patients make the correct response [8]. It is important to note that the exaggerated aspect of the stimuli should be the aspect that eventually controls the behavior.

Stimulus Shaping:

Stimulus shaping refers to the shaping of certain physical aspects of the stimuli [8]. For example, in [8], people with ASD were taught the concept of decimal points in money.

They were first presented expressions such as “\$2 and 70” and gradually the word “and” shrinks to a decimal point.

5.0 Robotics in Autism Research

According to [13], people diagnosed with ASD appear more comfortable in a predictable environment, for example interacting with a computer. By building a robot that is simple and predictable in appearance, the intervention therapy should become easier to approach. It also has been noted by [14] that the focus of people with ASD tend to be on objects isolated from the surrounding, which makes computers ideal as a screen can attract people with ASD and prevent them from moving their attention to any external stimuli. Reference [13] also mentions the benefit of having a physical robot instead of computer software or virtual characters is that real robots can provide multi-modal prompts (e.g. physical and manual) to better help children with ASD learn new skills. There has been a number of research efforts on utilizing robots from simple mobile robots to androids and humanoids for autism intervention.

Reference [15] explored the possibilities of using a mobile robot as an imitation agent with children with ASD during intervention. It has been reported that low-functioning children generally have deficits in sharing attention and thus have difficulties in understanding social signals [15]. The benefits of developing a mobile robot platform that behaves in a predictable fashion is that mobile robots can generate interplay situations and become the focus of sharing attention themselves [15].

Reference [16] built an interactive robot that is small, soft and simple to help engage not only people diagnosed with ASD, but also people with other communication difficulties such as pervasive developmental disorders (PDD). In addition, they also found that not only could their simple and predictable robot engage children with ASD, it also served as a pivot for triadic interactions between children and others [16].

The Aurora project conducted by [17] encourages social skills by imitation and turn-taking games with both mobile and humanoid robots. During the longitudinal study, a number of behavioral metrics were measured to gauge how much improvement each participant gained over the study. Furthermore, [17] also showed how a robot can act as a mediator for joint attention during interactive play.

6.0 Kinect for Autism Study

The Microsoft Kinect 2.0 is a platform which was originally developed for entertainment purposes. It combines an RGB camera, an infrared depth sensor as well as an array of microphones. The software development kit (SDK) provided by Microsoft contains libraries for face recognition and skeleton extraction based solely on the image/video taken by the camera. Therefore, it is suitable for research related to human-robot interaction as it frees the scientists from writing programs or finding open-source libraries for detecting face/body movements themselves.

There have been other efforts focused on utilizing the Kinect for autism research. In [18], a Kinect game was proposed to facilitate memory improvement and social skill acquisition. The problem with this proposal, as mentioned by the author, is that it was difficult to assess the improvements of children's behavior.

In [19], a motion-based "full-body" game was developed to help children with ASD learn attentional skills. Moreover, this study showed how a touchless game based solely on body movements could engage children with ASD and prompt attention skills [19].

7.0 Detecting Engagement/Disengagement

There has been a number of prior research efforts focused on developing algorithms to detect user disengagement or engagement states. Some features used by past research efforts include: body posture [20], gestures [20], facial expression [21], eye gaze [20] [21],

electroencephalography (EEG) [22], contextual information [21] and spatial relationship between the robot and human [23].

In [20] and [21], the feature sets included eye gaze directions, gestures and head directions, which were extracted from human observations. In such a scenario, a robotic system would need to employ modeling methods that can run autonomously, without any human intervention. Reference [21] also talks about using contextual information obtained from the log file of a storytelling app on a tablet. This kind of information though may not always be available to the robot, since robot interventions should occur in real-time, during the therapy session, rather than after the child has completed the task.

With respect to EEG applications such as in [22], even though EEG is considered noninvasive, placing electrodes on some children with ASD might be intolerable. In addition, using EEG devices limits the naturalness of the session and thus behaviors learned in the session may not be transferrable to the child's natural environment.

The spatial model in [23] used the relative location of robot and human for a receptionist robot to detect engagement levels. However, in the case of a therapy robot for children with ASD, relative location information gives very little information about the engagement level of the child since children are typically engaged within a local zone of proximity to the therapist during the therapy session.

Beyond the ones mentioned, there are a number of other challenges associated with the process of selecting features that enable the modeling of engagement levels in children with ASD. Using eye gaze to analyze engagement, such as in [24] and in the ASD study by [25], requires the tracked face to be directly aligned towards the sensor. In our previous studies [26], we have seen that in most therapy sessions, the face is not always orientated toward the sensor, and, in fact, is just as likely to be orientated toward the task, the therapist, or, in some cases, towards the floor/table.

Using voice and verbal recognition has also been shown as a feasible option, such as in [27] where acoustic features from speech were used to model and detect engagement

levels in daily conversations. However, depending on the therapy, a child with ASD may or may not talk during the session and certain children with ASD are nonverbal.

Other work, such as in [28] monitors user input during tablet interaction in order to assess engagement. However, children with ASD may not always be interacting with a tablet during a therapy session, thus features for modeling engagement should be extracted from interactions outside the tablet.

CHAPTER 3

APPROACH

In order to build a system which is able to perform autism intervention therapy autonomously, we will look at the first step of detecting child engagement and disengagement levels. We know that one key aspect of prompting and autism therapy is targeted at attention skill, i.e. to help a patient stay focused on a task. Also, because current engagement detection algorithms rely on features that are inapplicable to people with ASD, we propose the use of Kinect and kinematic movements extracted from Kinect as features towards modeling the engagement level of a child during an autism intervention therapy session. Besides the fact that the Kinect can recognize the skeleton accurately without adding any attachments to the child's body, which can be intolerable, it also provides objective measures instead of features extracted by human annotators as in some previous research efforts.

Autism intervention therapy usually consists of a therapist, a student (child) and a task. Some examples of the task include a piece of paper, a book, interactive toy and a touch-screen tablet. While the student is working on the task, the therapist will issue prompts whenever the child appears to be distracted and stops concentrating on the task.

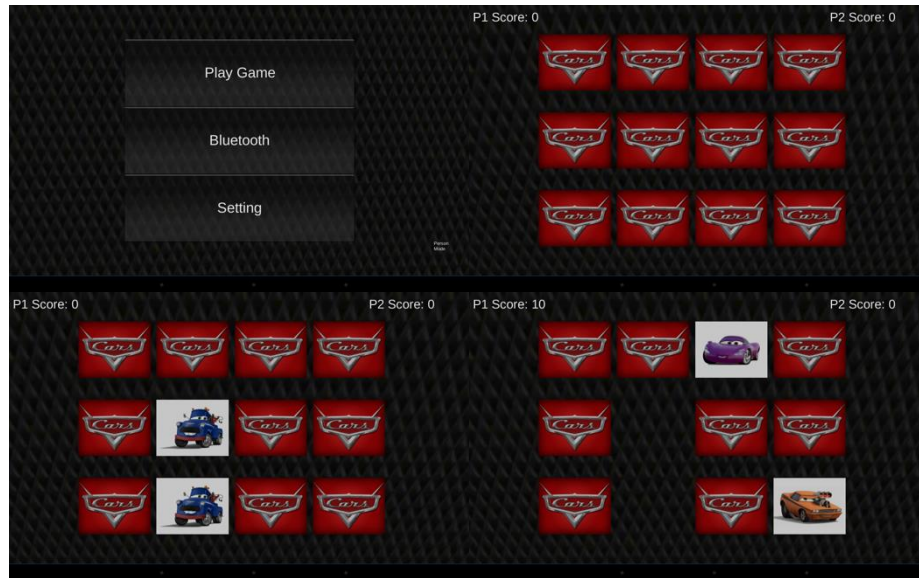


Figure 3. Screenshot of the game running on the tablet used as therapy task. The start menu (top left), default game view (top right), the user found a matched pair (bottom left) and the user failed to find a matched pair (bottom right).

For our experiment, a turn-taking game on a Samsung Android tablet was employed. The goal is to have two users try to match hidden cards and whoever matches more cards wins. In the room where the therapy is conducted, a Kinect mounted on a tripod is located in front of the table where the therapist and the child sit in order to capture as much as the movement from the interaction as possible. There is also another RGB camera in the room mounted to the side of the table that is used to record an extra viewing angle as well as to serve as a backup. Figure 3 contains screenshots of the game used in this study. Figure 4 shows the environment of the therapy center where the studies are conducted and how a typical session is captured using RGB image, skeletons and reconstructed 3D scene. Figure 7 shows the tablet from the user's point of view when interacting with it.

In our work, we define two states consisting of engagement/engaged and disengagement/disengaged. Engagement suggests focused and concentration on the task while disengagement is defined as the opposite state.

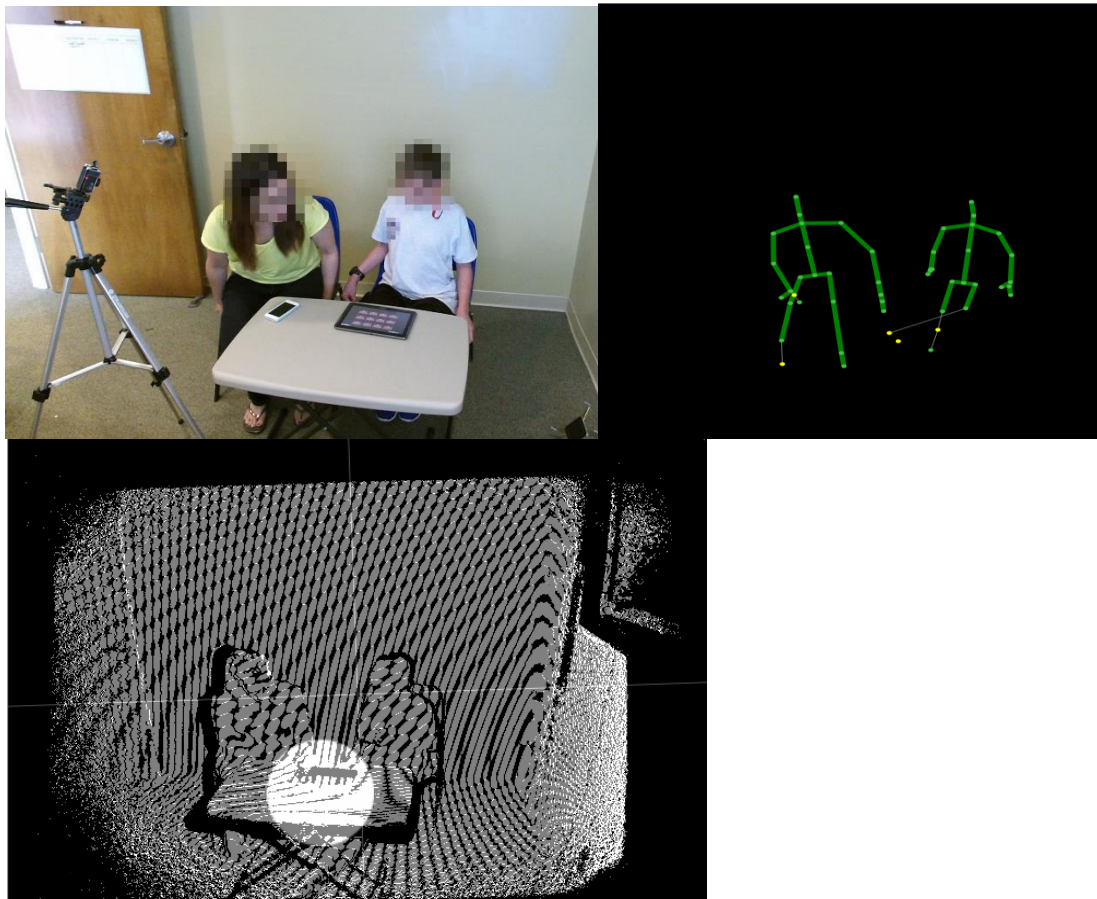


Figure 4. Therapy scene captured by Kinect. RGB image (top left), recognized skeletons (top right) and reconstructed 3D scene (bottom).

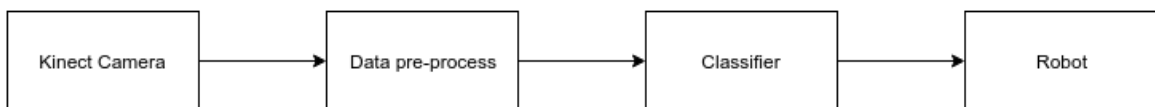


Figure 5. Flow diagram of optimal robot system for autism intervention.

In order for a robotic platform to perform autism intervention autonomously and in real-time, it needs to have a flow diagram similar to the one shown in Figure 5. However, for testing and validating proposes, the system in this work does not have to predict if the child is disengaged in real time. Instead, since we will be following the standard cross-validation method for validating and testing our ideas (as discussed later), our system will

only need to perform a classification on a specific window of interest. Thus the flow chart will look like the one shown below, and explained in more details later:



Figure 6. Flow diagram of the system deployed in this work.

Because our goal is only to validate our approach and in this work, we will not apply our system on a robot, we only need the labels generated by the classifier so that we can gauge the performance of our proposed system. Thus, we segment our recorded videos into clips containing interesting streams, extract features from those clips, feed the features into our classifiers and measure the performance in terms of accuracy of generated labels against group truth labels annotated by human observers.



Figure 7. Photo of the tablet used throughout the sessions.

1.0 Kinect Camera

First the real world coordinates of each joint on the skeletons recognized by the Kinect camera are stored in a comma separated value (CSV) file along with the timestamp of the frame where the skeleton is extracted. Table 1 contains an example of the CSV file captured during a therapy session.

Table 1. Sample of CSV file containing joint coordinates captured by Kinect.

Timestamp	Skeleton No.	Joint No.	x (m)	y (m)	z (m)
12:03:39	3	0	0.32079	-0.2746	2.16135
12:03:39	3	1	0.26085	0.01162	2.08869
12:03:39	3	2	0.20064	0.29004	2.00293
12:03:39	3	3	0.1927	0.41802	1.94864
12:03:39	3	4	0.06413	0.23925	2.07017
12:03:39	3	5	0.03706	0.04817	2.11797

2.0 Data Pre-processing

Before we can extract various features from the joint coordinates, we need to pre-process and clean up the CSV files. As both the therapist and child are sitting during the therapy trials, their lower bodies will not be visible to the camera as they are blocked by the table. Thus, the coordinates for joints on the lower body (feet, ankles etc.) will not reflect their correct positions and we will simply remove them from the CSV files both for saving storage and boosting the processing speed in the future steps.

Another issue with the original CSV files is the timestamp. The timestamps are generated based on the time at which the Kinect SDK is running on. Therefore, all the timestamps will be in “wall-clock time”. For obvious reasons, we want to convert all the timestamps to be relative to the start time of each session. Thus we store the first timestamp (t_{start}) in a file and replace all remaining timestamps (t_{remain}) with their difference $t_{remain} - t_{start}$.

3.0 Stream Segmentation

After the CSV files are cleaned up, they are segmented into smaller clips. Our goal is to chop a long stream into smaller pieces for the future steps. In addition, to produce more training and test instances, we utilize a moving window of 2 seconds and an overlap of 1 second. As time increases, the sliding window also travels. Moreover, because we are only interested in engagement and disengagement sections of the therapy, we segment the clips into those with only the label “engagement” or “disengagement”.

4.0 Feature Extraction

After we obtain the segmented clips, we are ready to extract various features related to kinematic movements that might contribute to classifying engagement and disengagement states. For this work, we classify these features as leaning angle, planar distance to therapist, mean joint to joint distance, distance of joints traveled within task ball, mean joint coordinates, mean joint distance to task and mean joint to joint distance.

Features

a. Leaning angle: The leaning angle is the angle between the vertical y-axis in the camera’s view and the vector constructed from the midpoint of the spine and neck of the child. Leaning angle is chosen to represent the scenario correlated with an individual concentrating on a task. In such cases, it has been observed that individuals tend to lean towards the object of interest associated with achieving a task. For example, in our experimental case, the tablet device becomes the object of interest. The leaning angle θ for the spine joint vector $\vec{J}_{mid\ spine}$ and neck jointing vector \vec{J}_{neck} is calculated as:

$$\theta = \cos^{-1} \left(\frac{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot (\vec{J}_{mid\ spine} - \vec{J}_{neck})}{\left\| \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\| \left\| \vec{J}_{mid\ spine} - \vec{J}_{neck} \right\|} \right) \quad (1)$$

b. Planar distance to therapist: The planar distance to therapist feature is one measure used to calculate the distance measured between two people interacting with a task's common object of interest. A plane is constructed by using the middle point of the spine, neck and head of the therapist. Distances between this plane and the child's skeleton joints are then calculated. For the therapist mid-spine joint vector $\vec{J}_{mid\ spine}$, neck \vec{J}_{neck} and head \vec{J}_{head} , the plane P and planar distance D can be derived using the following equations:

$$P = (\vec{J}_{mid\ spine} - \vec{J}_{neck}) \times (\vec{J}_{mid\ spine} - \vec{J}_{head}) \quad (2)$$

$$D = \frac{\|\vec{c}_i \cdot P\|}{\|\vec{c}_i\|} \text{ for child joint vector } \vec{c}_i \quad (3)$$

c. Mean joint to joint distance: The mean joint to joint distance records the average distance between each joint of the therapist and each joint of the child. This feature reflects the relative pose and distance between the therapist and child during an interactive session. For each child's joint i and its joint vector \vec{c}_i as well as therapist's joint vector \vec{t}_i at the f th frame of F frames, where F is the number of recorded Kinect frames associated with a movement profile, the mean joint to joint distance d_i is determined as:

$$d_i = \frac{1}{F} \sum_{f=0}^F \|\vec{c}_i - \vec{t}_i\| \quad (4)$$

d. Distance traveled within task ball: This feature measures the distance traveled by each joint inside a sphere whose center is located at the task's object of interest and has radius r . For a sphere with radius r and centered at \vec{b} , if joint vector \vec{j}_i^f at the f th frame satisfies:

$$\|\vec{b} - \vec{j}_i\| < r, \text{ then } d_t = \sum_{f=1}^F \|\vec{j}_i - \vec{j}_{i-1}\| \quad (5)$$

e. Mean joint coordinates: The mean joint coordinates, associated with either an engagement or disengagement state, reflects the absolute pose of the child during a therapy session measured in the 3D world. It is obvious that in order to make this feature meaningful across various intervention sessions, it has to be normalized to eliminate overfitting to a specific session setup (for example how far the child sits away from the camera should not affect the performance of the system). As such, for each joint vector \vec{j}_i and F frames, the mean joint coordinates c_i is calculated as:

$$d_i = \frac{1}{F} \sum_{f=0}^F \|\vec{j}_i\| \quad (6)$$

f. Mean joint distance to task: The mean joint distance to task calculates the average distance between each joint and the location of the task object of interest. This feature reflects how far away a child is from the task during a therapy session. The intuition of selecting this feature comes from the observation that when engaged, a user's head tends to be close to the task. Figure 9 contains plots of head distance to task across 5 engagement clips and disengagement clips. Even though the trends does not look similar, the range for head to task distance in disengagement is generally more concentrated near (0.8 to 0.9 meter) while the range for engagement clips is mostly centered around 0.5 meter. For each child joint vector \vec{c}_i at the f th of F frames, the mean joint distance to task T is:

$$d_i = \frac{1}{F} \sum_{f=0}^F \|\vec{c}_i - \vec{T}\| \quad (7)$$

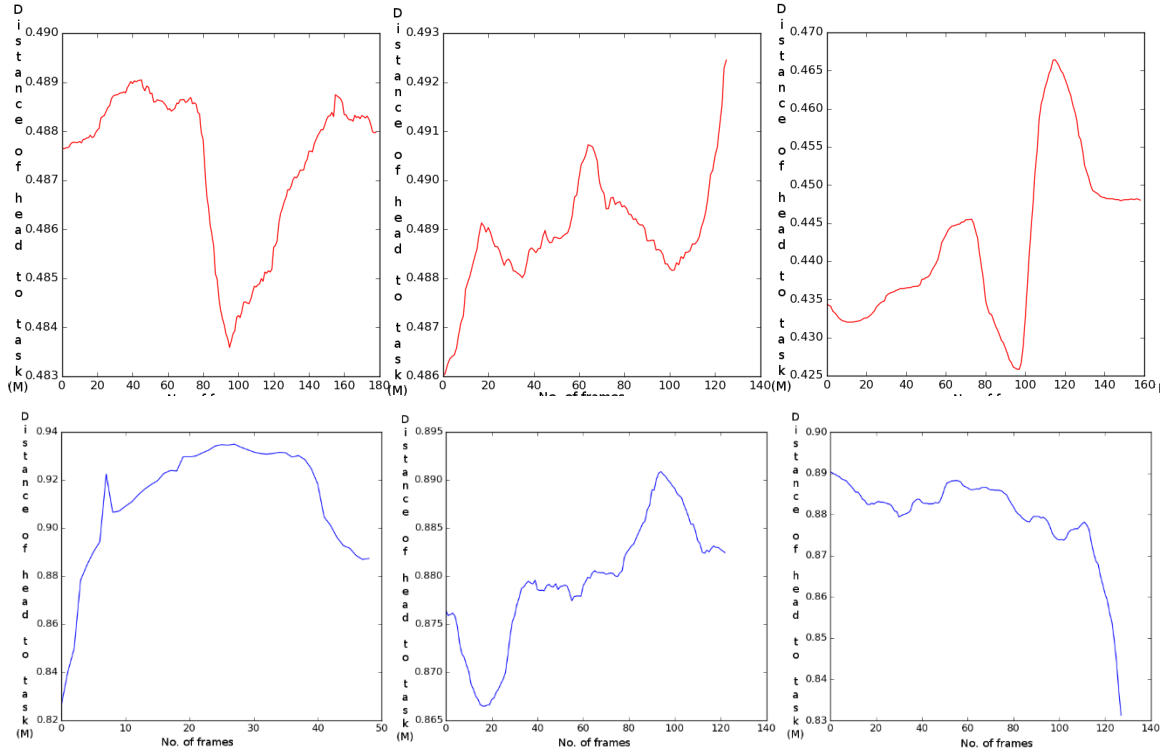


Figure 8. Distance of head joint to task in meters vs. no. of frames. Engagement clips (top) and disengagement clips (bottom).

Once determined, these relevant features can then be utilized to identify engagement state. It is worth noting that some of these features require the presence of two skeletons during the therapy interaction, i.e. a child and a therapist/caregiver. However, the robot would not be autonomous if it requires the presence of a professional therapist. Therefore, to classify engagement state, we will mainly focus only on those features that are derived solely from the child’s skeletal data.

Depth image to 3D reconstruction

In addition, when locating the task’s real world coordinates relative to the Kinect camera, we need to reconstruct the 3-dimensional scene captured by the depth/infrared camera. The depth data is stored as a CSV file containing ushort data types representing millimeters in C# from the .NET Microsoft SDK. First we want to convert the depth data from millimeters ($depth_{millimeter}$) to meters ($depth_{meter}$) by

$$depth_{meter} = \frac{depth_{millimeter}}{1000} \quad (8)$$

In [29], the distortion in Kinect’s depth camera is modelled by both radial and tangential distortions. The projection matrix is given by [29] as:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} f_{cx} & 0 \\ 0 & f_{cy} \end{bmatrix} \begin{bmatrix} d_x \\ d_y \end{bmatrix} + \begin{bmatrix} c_{cx} \\ c_{cy} \end{bmatrix} \quad (9)$$

where x_c and y_c are coordinates on the depth image, f_{cx} and f_{cy} are focal lengths of the camera, d_x and d_y are the tangential distortions and c_{cx} and c_{cy} are the principal point/pixel. The intrinsic parameters (focal lengths and principal point) used in this work came from [30].

Classifiers

After the features have been extracted, we move on to the two-class classification problem defined earlier, where the goal is to evaluate our features and gauge the performance of a few classifiers on our problem. For this work, we evaluate the performance of four classification methods that can be used to distinguish between the two different states, namely support vector machines (SVM), Random Forest, AdaBoost and K-Nearest-Neighbor (KNN).

Support Vector Machine

Support vector machine (SVM) is known for its robustness and ability to perform well when given noisy data. SVM separates binary data by creating a hyper-plane that is at maximal distance from two labels (i.e. maximal margin hyper-plane) [31]. Furthermore, since the optimal separating plane might not always be linear, the idea of a “kernel” is introduced. Kernel works by projecting data instances into higher dimensions by utilizing a “dot product” with the feature vectors. When the maximal margin hyper-plane is found,

it is mapped back to the original feature space [31]. SVM is solved as an optimization problem [31]:

$$\gamma = \min_i y_i \{ \langle w, \phi(x^i) \rangle - b \} \quad (10)$$

where γ is the margin, (w, b) defines the hyper-plane, y_i denotes the i th label and $\phi(x^i)$ is the kernel version i th input vector.

Random Forest

Random Forest (RF) is an ensemble method consisting of a collection of decision trees that classifies an instance of data based on the label voted by the majority of trees [32]. Each decision tree is trained with a random sample set of features. Thus, Random Forest is a classification method that utilize both bagging (collection of trees) and unstable learners (sampled feature set) [32]. Furthermore, each tree in a RF is unpruned in order to obtain low biased learners [32]. In addition, the random sampling of the feature set results in low correlation between trees. Thus RF is a classifier with both low-bias and low-variance [32]. The training procedure for RF with n trees is as follows [32]:

1. For each tree, randomly sample a set of features with replacement,
2. Grow a tree on the sampled features by finding the best splitting feature until no feature is left (i.e. no pruning)
3. Repeat 1 and 2 until we obtain n number of trees

AdaBoost

Being an iterative algorithm, AdaBoost is trained to approximate the Bayes classifier by incorporating many weak classifiers [33]. AdaBoost first tries to build a classifier with equal weights for all the training samples and uses it to classify all the training data. If a data point is misclassified, its weight will increase and the increased weight will be used to train the second classifier [33]. After obtaining a group of classifiers,

each classifier is given a score and the final label is simply the linear combination of outputs from each classifier and their corresponding scores [33].

KNN

K-Nearest-Neighbor (KNN) is a nonparametric classification method that for training data $\{x_1, x_2, x_3 \dots\}$ and their corresponding labels $\{y_1, y_2, y_3 \dots\}$, we predict the label for unseen data q by finding its k nearest neighbors $\{y_i, y_{i+1} \dots, y_k\}$ and simply choose $mode(y_i, y_{i+1} \dots, y_k)$ for q 's label [34].

Hold-one-child-out

In this work, these methods were trained and tested using “hold-one-child-out method” using Scikit-learn [36]. In “hold-one-child-out” method, we pick data recorded with one child as the test data, and train the classifier with data from all other children. In this process, the following procedure was followed for each of the k folds: A model using one of the respective training algorithms (SVM, Random Forest, AdaBoost and KNN) was trained using clips from all the children except the test one. The resulting model was then validated for classification accuracy using the remaining hold out part of the data as the test set. This process was repeated for each set of children and the final accuracy value is computed as the average of accuracy values for each round.

5.0 Summary

So far, we discuss how our system is constructed, namely what major components are essential for an autonomous robot designed for autism intervention therapy, how kinematic data generated by the Kinect camera is stored and processed, how a long stream of session is segmented into smaller clips, what features are used and how they are calculated, how the task is reconstructed and located in the 3D world, how the cross validation is conducted and which classifiers are used to detect engagement/disengagement states.

In the next chapter (Chapter 4), we will talk about how an autism therapy session is set up to acquire data for this study and present demographic data about our participants. Chapter 5 presents our experimental results based on our classifiers and features. Chapter 6 elaborates on our conclusion as well as some possible future works.

CHAPTER 4

EXPERIMENTAL SETUP

The goal of the classifier is to accurately detect engagement/disengagement states of a child in order for a therapy robot to provide prompts at the right time. As such, given our set of possible body movement features and classifiers, our goal was to determine the accuracy rates for each classifier and feature set (i.e. model) in order to select the set with the greatest ability to discriminate between states.

For this experiment, a pilot study was conducted at the Kid’s Creek Therapy Center. The parents of each participant signed the IRB (Institutional Review Board) approved consent form allowing their child to engage in the testing sessions. Children diagnosed with developmental disabilities were recruited for this experiment with 3 boys, $mean(age) = 12.3$ and $\sigma(age) = 1.5$ (Table 2). The child study consisted of sessions where, in each session, the experimenter and the child played a turn-taking game on the tablet [i.e. the task] (Figure 10). During interaction, the experimenter asked a series of questions to distract the child during the child’s turn such as “Do you remember my name?”. Three sessions of approximately 28 minutes in length were recorded and processed. During the experiment, the real world coordinates of all joints of the human upper body skeleton were recorded as well as color video and audio streams from the Kinect camera. Since the camera was fixed on a tripod across all sessions, the classifier trained using data from one session was able to be applied to another session without normalization. For training, we did not select the features including planar distance to therapist and mean joint to joint distance, since these features required two skeletons to be present. As the goal of a therapy robot is to allow children to receive intervention without a therapist closely present, we determined that our model should only be built from skeletal data based on the child’s movement profile.

Lastly, when calculating the feature vectors associated with the child and task, we also needed to know the location of the task's object of interest (i.e. the tablet). To obtain this information, we calculated the real world coordinates of the tablet by first reconstructing the 3D scene using the depth image obtained by the Kinect camera and manually selecting the 3D point corresponding to the tablet in the image scene.

Table 2. Demographic data about participants.

Participant	Diagnosis	Gender	Age
1	ASD	Male	12
2	Down Syndrome	Male	14
3	ASD	Male	11



Figure 9. Typical experimental setup where the instructor (left) and patient (right) are sitting together interacting on the task (game running on the tablet).

Once collected, the stream of data from the pilot study was annotated by a human annotator with timestamps indicating the start and end of both engagement and disengagement states. Timestamps were annotated based on the identified behaviors in the videos. For example, some typical disengagement behaviors included standing up and walking away from the tablet and talking to others about things unrelated to the session. Disengagement states were also associated to those instances of time when the experimenter asked the series of questions designed to expressly distract the child.

Once the timestamps associated with the start and end times for the different states were obtained, clips were then segmented into smaller ones, each lasting for 2 seconds with a 1 second overlap. This was done in order to provide us with a sufficient number of training and testing instances to validate the model.

CHAPTER 5

RESULTS

Data from the various child interaction sessions recorded during our pilot study were used to evaluate the performance of the different classifiers and feature sets using the “hold-one-child-out” cross validations as discussed previously.

As shown in Table 3, the best performing body movement feature was identified as Mean Joint Coordinates. This can be due to the fact that the child’s joints tend to be in fixed locations when he/she is engaged/disengaged from the task. It is surprising that even though the Kinect and table/task are always mounted at the same locations across all the session recorded for this study, classifiers trained using joint coordinates generalize so well even without normalization. However, it is not practical to ask users of our system to always position the task and Kinect camera at the same exact places. Thus, the robot should not rely on only Mean Joint Coordinates. The second best performing feature is Mean Joint Distance to Task and it simply acts as a normalized version of Mean Joint Coordinates. The reason why Distance Traveled within Task Ball works poorly can be that because of the relative small radius for the sphere (0.5m), most joints never enter the sphere thus the feature vector we obtained contained a lot of zeros which may cause the classifier to overfit.

Figure 10 expands on the corresponding table results. Based on this assessment, using AdaBoost with Mean Joint Coordinates achieved the best single-feature performance at 96% accuracy while AdaBoost with all the features gives an accuracy of 97% appears to be the best performing combination.

Table 3. Accuracies for classifiers and feature sets combinations.

Features	SVM	Random Forest	AdaBoost	KNN
Mean Joint Coordinates	88%	96%	96%	93%
Mean Joint Distance to Task	70%	65%	93%	82%
Distance Traveled within Task Ball	70%	60%	66%	67%
All Three Features	88%	96%	97%	93%

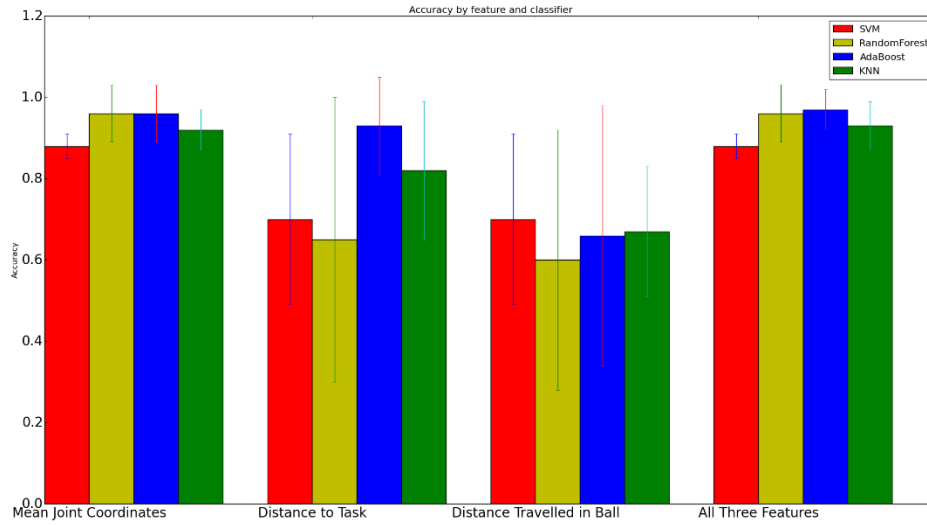


Figure 10. Experimental results from "hold-one-child-out" cross validation.

We also tested our model on two sessions where the child only interacts with the robot during the game (i.e. no therapist present, see Figure 11). These streams are segmented exactly like the ones in which the child interacts with the therapist: each clip only contains one label: engagement or disengagement. Then the clips are processed and features are extracted. The 3D scene is also reconstructed in order to locate the task.

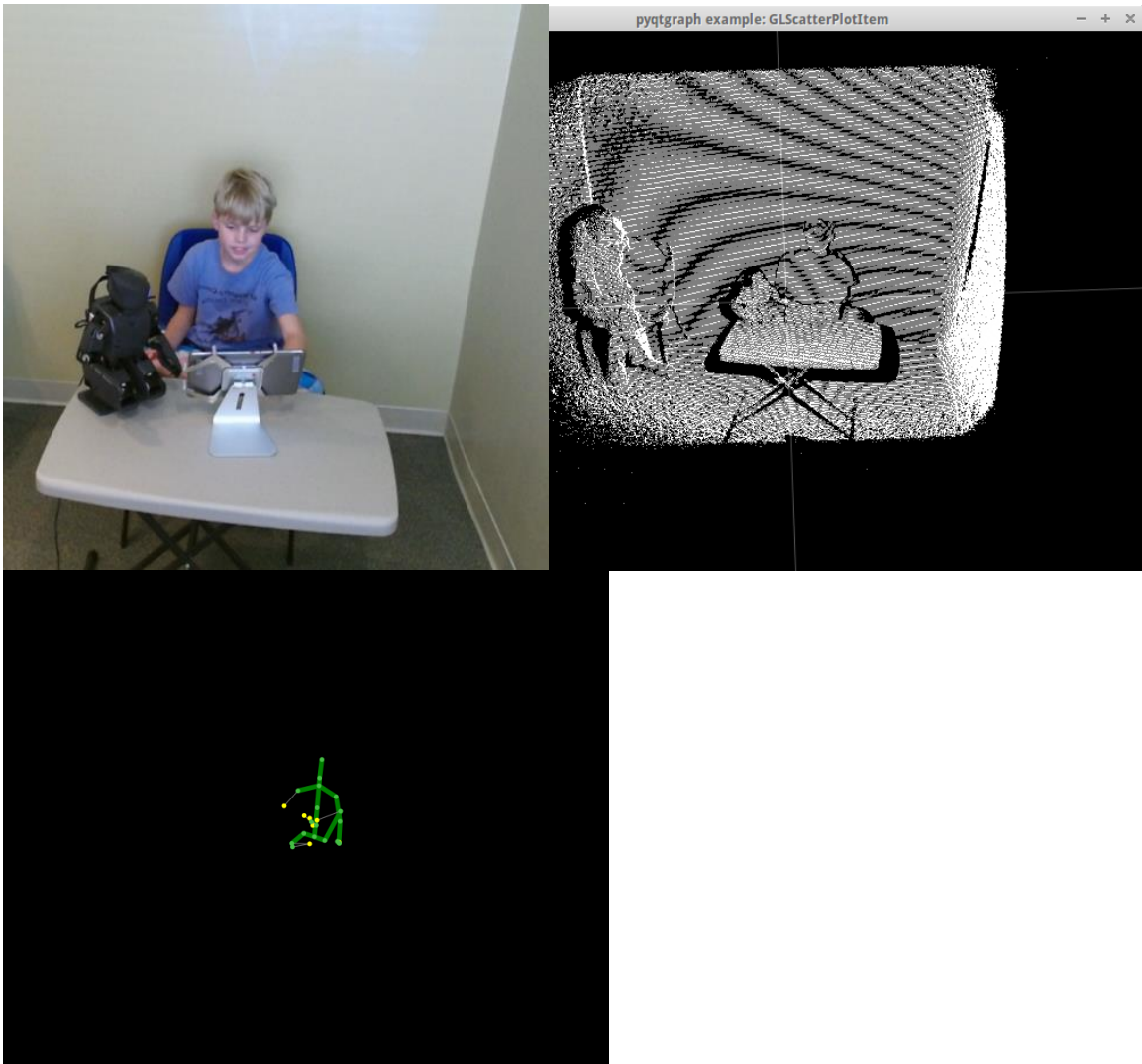


Figure 11. RGB (top left), reconstructed (top right) and skeleton view of session where the child interacts with the robot.

The result is presented in Table 4. The best performing combination is Random Forest (RF) with Mean Joint Distance to Task at 57% accuracy, illustrated in Figure 12. One of the reasons why the performance is degraded compared with the previous results from sessions where the child interacts with the therapists is the small number of training instances (30 instances compared with 120).

Table 4. Accuracy for feature set and classifier combination, trained with child interacting with therapist and tested on child interacting with robot.

Features	SVM	Random Forest	AdaBoost	KNN
Mean Joint Coordinates	56%	51%	51%	46%
Mean Joint Distance to Task	56%	57%	46%	51%
Distance Traveled within Task Ball	56%	56%	56%	44%
All Three Features	56%	56%	54%	40%

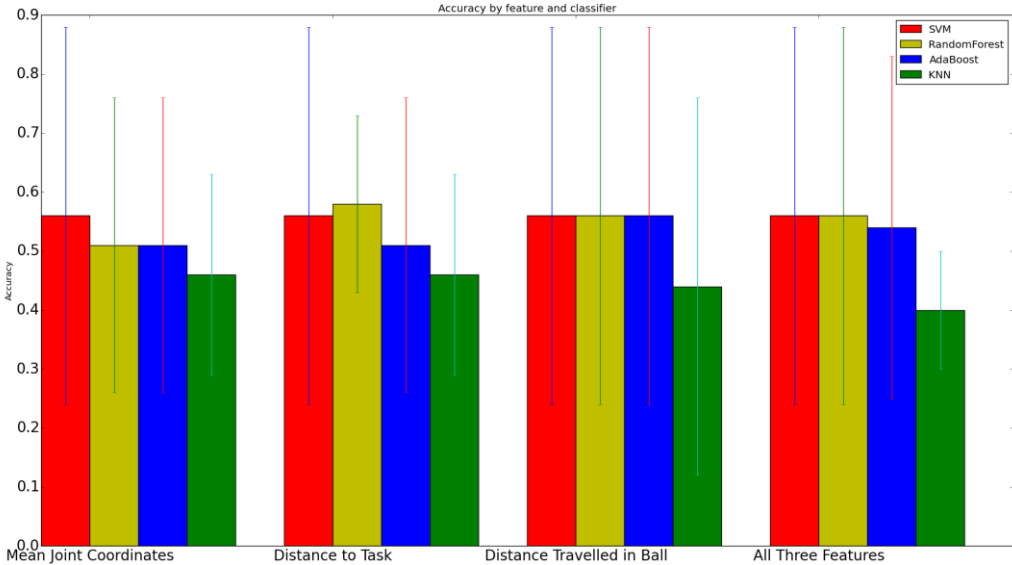


Figure 12. Accuracies for classifier and feature set combination, models trained with child interacting with therapist and tested on child interacting with robot.

CHAPTER 6

CONCLUSION AND FUTURE WORK

1.0 Conclusion

So far, we reviewed the design of a method to be used as a first step in performing autism intervention therapy independently. We limit the intervention that our system can perform to detecting and prompting disengagements in a timely manner, propose a data-processing and machine learning pipeline and evaluate its performance using data obtained with children with autism spectrum disorder (ASD). The pipeline starts with the Microsoft Kinect RGB-D camera that recognizes the kinematic movements of the child during an intervention therapy and ends with a classifier detecting the engagement/disengagement states. The reason why we choose Kinect as our source of features is that most previous research on engagement level detection are either based on invasive technologies (children with ASD generally feel uncomfortable when there is an attachment positioned on their bodies) or behaviors/events that may not be applicable to all the people on the autism spectrum (e.g. features related to speech or voice). We show that by carefully selecting features acquired with non-invasive technology (a camera), we can still achieve relative good performance of 97% accuracy on human-annotated clips containing only engagement or disengagement states.

2.0 Future work

1.0 Longitudinal Study and Comparison with Therapist

One possible future work of this project is to deploy our system into additional therapy sessions and collect data from a longitudinal study to evaluate its performance compared to the performance of a therapist. The motivation of this work is to develop

autonomous systems that can help people with ASD learn social and cognitive skills and the goal of developing a system like this is for individuals with autism to learn and work around it. After gathering data from a longitudinal therapy study, drawbacks and deficiencies of our current design can be identified and solutions developed in order for future patients to gain optimal results from robotic-based intervention therapy.

2.0 Novel Features

Another aspect of future work is to design and evaluate other novel features not present in this work. All the features in this project are derived from kinematic skeletal data. If the child is interacted with the robot, contextual data gathered by either sensors embedded within the robot or events triggered by the interaction could be utilized.

We can also observe the fact that the action of losing concentration generally unfolds itself over time. It might be beneficial to switch to a generative model such as Hidden Markov Model (HMM) to model how disengagement evolves over time.

3.0 Expand to detect other behaviors

Our system is currently limited to detect engagement and disengagement states of the patient in order to issue prompts accordingly. We would like to expand the behaviors that the robot can detect to undesirable behaviors and teach social and cognitive skills to people with ASD. For example, the robot can provide prompts for a child who makes a mistake during a math exercise.

REFERENCES

- [1] D. L. Christensen, et al., "Prevalence and Characteristics of Autism Spectrum Disorder Among 4-Year-Old Children in the Autism and Developmental Disabilities Monitoring Network," *Journal of Developmental & Behavioral Pediatrics*, vol. 37, no. 1, pp. 1–8, 2016.
- [2] C. M. Corsello, "Early intervention in autism," *Infants & Young Children*, vol. 18, no. 2, pp. 74–85, 2005.
- [3] D. Granpeesheh, et al., "The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders," *Research in Autism Spectrum Disorders*, vol. 3, no. 4, pp. 1014–1022, 2009.
- [4] B. Scassellati, et al., "Robots for use in autism research," *Annual review of biomedical engineering*, vol. 14, pp. 275–294, 2012.
- [5] D. L. Sharpe and D. L. Baker, "Financial issues associated with having a child with autism," *Journal of Family and Economic Issues*, vol. 28, no. 2, pp. 247–264, 2007.
- [6] B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?," *Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, 2005.
- [7] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism," *Autonomous Robots*, vol. 24, no. 2, pp. 147–157, 2008.
- [8] G. S. MacDuff, P. J. Krantz, and L. E. McClannahan, "Prompts and prompt-fading strategies for people with autism," *Making a difference: Behavioral intervention for autism*, pp. 37–50, 2001.
- [9] S. Eikeseth, T. Smith, E. Jahr, and S. Eldevik, "Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7 a comparison controlled study," *Behavior modification*, vol. 31, no. 3, pp. 264–278, 2007.

- [10] E. B. Caronna, J. M. Milunsky, and H. Tager-Flusberg, "Autism spectrum disorders: clinical and research frontiers," *Archives of Disease in Childhood*, vol. 93, no. 6, pp. 518–523, 2008.
- [11] B. Rosenwasser and S. Axelrod, "The contributions of applied behavior analysis to the education of people with autism," *Behavior modification*, vol. 25, no. 5, pp. 671–677, 2001.
- [12] F. C. Mace, "The significance and future of functional analysis methodologies," *Journal of applied behavior analysis*, vol. 27, no. 2, pp. 385–392, 1994.
- [13] B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?," *Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, 2005.
- [14] D. Murray, "Autism and information technology: therapy with computers," *Autism and learning: a guide to good practice*, pp. 100–117, 1997.
- [15] A. Duquette, et al., "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism," *Autonomous Robots*, vol. 24, no. 2, pp. 147–157, 2008.
- [16] H. Kozima, et al., "Interactive robots for communication-care: A case-study in autism therapy," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, 2005, pp. 341–346.
- [17] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn, "Robot-mediated joint attention in children with autism: A case study in robot-human interaction," *Interaction studies*, vol. 5, no. 2, pp. 161–198, 2004.
- [18] E. Boutsika, "Kinect in education: A proposal for children with autism," *Procedia Computer Science*, vol. 27, pp. 123–129, 2014.
- [19] L. Bartoli, et al., "Exploring motion-based touchless games for autistic children's learning," in *Proceedings of the 12th International Conference on Interaction Design and Children*, 2013, pp. 102–111.

- [20] I. Leite, et al., “Comparing models of disengagement in individual and group interactions,” in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 99–105.
- [21] G. Castellano, et al., “Detecting user engagement with a robot companion using task and social interaction-based features,” in *Proceedings of the 2009 international conference on Multimodal interfaces*, 2009, pp. 119–126.
- [22] C. Berka, et al., “EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks,” *Aviation, space, and environmental medicine*, vol. 78, no. Supplement 1, pp. B231–B244, 2007.
- [23] M. P. Michalowski, et al., “A spatial model of engagement for a social robot,” in *Advanced Motion Control, 2006. 9th IEEE International Workshop on*, 2006, pp. 762–767.
- [24] Y. I. Nakano and R. Ishii, “Estimating user’s engagement from eye-gaze behaviors in human-agent conversations,” in *Proceedings of the 15th international conference on Intelligent user interfaces*, 2010, pp. 139–148.
- [25] E. Bal, et al., “Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state,” *Journal of autism and developmental disorders*, vol. 40, no. 3, pp. 358–370, 2010.
- [26] H. W. Park ,A, Howard. “Engaging Children in Social Behavior: Interaction with a Robot Playmate Through Tablet-Based Apps.” *Rehabilitation Eng. and Technology Society of North America (RESNA) Annual Conference*, Indianapolis, IN, June 2014.
- [27] C. Yu, P. M. Aoki, and A. Woodruff, “Detecting user engagement in everyday conversations,” arXiv preprint cs/0410027, 2004.
- [28] H. W. Park, R. A. Coogle, and A. Howard, “Using a shared tablet workspace for interactive demonstrations during human-robot learning scenarios,” in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 2713–2719.
- [29] C. Raposo, J. P. Barreto, and U. Nunes, “Fast and accurate calibration of a kinect sensor,” in *3D Vision-3DV 2013, 2013 International Conference on*, 2013, pp. 342–349.

- [30] Ian Ormesher, Kinect V2 Depth Camera Calibration,” *Three Constants*, 09-Nov-2014. .
- [31] T. S. Furey, et al., “Support vector machine classification and validation of cancer tissue samples using microarray expression data,” *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [32] R. Díaz-Uriarte and S. A. De Andres, “Gene selection and classification of microarray data using random forest,” *BMC bioinformatics*, vol. 7, no. 1, p. 1, 2006.
- [33] J. Zhu, H. Zou, S. Rosset, and T. Hastie, “Multi-class adaboost,” *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
- [35] V. Garcia, et al., “Fast k nearest neighbor search using GPU,” in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on*, 2008, pp. 1–6.
- [36] F. Pedregosa, et al., “Scikit-learn: Machine learning in Python,” *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.