Automated Detection and Classification of Positive vs. Negative Robot Interactions With Children With Autism Using Distance-Based Features

David Feil-Seifer Maja J Matarić Interaction Laboratory University of Southern California University Park, Los Angeles, California, USA 90089 dfseifer|mataric@usc.edu

ABSTRACT

Recent feasibility studies involving children with autism spectrum disorders (ASD) interacting with socially assistive robots have shown that some children have positive reactions to robots, while others may have negative reactions. It is unlikely that children with ASD will enjoy any robot 100% of the time. It is therefore important to develop methods for detecting negative child behaviors in order to minimize distress and facilitate effective human-robot interaction. Our past work has shown that negative reactions can be readily identified and classified by a human observer from overhead video data alone, and that an automated position tracker combined with human-determined heuristics can differentiate between the two classes of reactions. This paper describes and validates an improved, non-heuristic method for determining if a child is interacting positively or negatively with a robot, based on Gaussian mixture models (GMM) and a naive-Bayes classifier of overhead camera observations. The approach achieves a 91.4% accuracy rate in classifying robot interaction, parent interaction, avoidance, and hiding against the wall behaviors and demonstrates that these classes are sufficient for distinguishing between positive and negative reactions of the child to the robot.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: Human information processing; J.4 [Social and Behavioral Sciences]: Psychology

General Terms

Algorithms, Experimentation

Keywords

ASD, Human-Robot Interaction, Socially Assistive Robotics, behavior modeling

Copyright 2011 ACM 978-1-4503-0561-7/11/03 ...\$10.00.

1. INTRODUCTION

Socially assistive robots (SAR) have been shown to have promise as potential behavioral assessment and therapy tools for autism spectrum disorder (ASD), because children with ASD express an interest in interacting socially with such machines [17, 19]. The long-term goal of this and related endeavors is to develop robot systems that can aid in the diagnosis and treatment of ASD, providing methods complementary to and not competitive with human-delivered care. Specifically, this work is part of a larger effort to develop an autonomous socially assistive robot systems capable of social skill training delivery to children with ASD.

We aim to develop tools that engage children on a broad range of the autism spectrum, including lower-functioning children with less developed communication abilities. As such, we focus on facilitating scenarios wherein the child and robot can interact however the child chooses, with no specific task or game rules or constraints. However, autonomous robot operation in such free-form social settings presents a number of challenges, including autonomous on-line understanding the child's behavior and timely, appropriate realtime robot responses. In addition, the unconstrained nature of the interaction means that a priori categorizations of the child's behavior, especially in the light of the known heterogeneity of the ASD population, is not realistic. We therefore aim to develop methods that can automatically classify the child's behavior during the interaction, and then use that information to enable the robot to respond in a timely and appropriate fashion.

Our prior work has shown that children have varied reactions to a socially interactive robot, some positive, some negative [7]. This is not surprising, as children with ASD are not likely to enjoy any robot 100% of the time. Some past work has reported more uniformly positive child responses to robots [4, 11] but may not have involved the same spectrum of severity of ASD diagnoses. We found that there were specific morphological and behavioral features of the robot that some children, especially those with more severe diagnoses, identified as distracting or annoying. This led us to explore methods for autonomously detecting negative behaviors in order to minimize distress and respond properly, in order to facilitate effective human-robot interaction.

The children's response to the robot could be seen in the distances they observed relative to the robot and other features of the environment. Children who had a positive session with the robot spent a good deal of time directly in-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.

teracting with the robot, standing in front of the robot at a distance of less than 1.5m. Children that did not react well to the robot, on the other hand, spent a good deal of time moving away from the robot, and stayed close to a wall or to a parent. These behaviors are quite informative regarding the child's state and acceptance of the robot even in the absence of other modes of communication (e.g., eye gaze, gesture, speech) that may be lacking in some children with ASD.

This paper demonstrates a method for using these distancebased behaviors as a means of classifying and interpreting the child's responses in the context of HRI for ASD. Our prior work used a set of distance features, such as the child's distance from the robot, walls of the room, and the parent, to classify the child's distance state into a small set, including *in front of robot, near parent*, and *near wall*. These distance states were then used to automatically differentiate between children who enjoyed the robot and children who did not [8]. However, because the features were heuristic, that approach would be difficult to generalize and apply to contexts with large numbers of features and states.

In this paper, we describe an improved approach that addresses the weaknesses of our prior work. Specifically, we present and validate a method that can automatically annotate interaction data based on a small labeled set of training data. We use Gaussian Mixture Models (GMMs), fitted by expectation-maximization (EM), with a model order selected automatically. A naive-Bayes classifier is then employed to annotate the data based on the small humanlabeled training data set. The approach achieves a 91.4% accuracy rate in classifying robot interaction, parent interaction, avoidance, and hiding against the wall behaviors. These behaviors are the necessary behaviors for classifying positive and negative reactions to the robot in our approach.

The next section details the experimental design for the data collection that provided the training data for this work. We describe how primitive features are detected, how they are clustered, and how the recognized states are used to group individual data recordings into categories.

2. EXPERIMENT DESIGN

We conducted a feasibility study with participant families which had children with ASD. This study provided the data used in this work. The study consisted of a free-play scenario involving a robot, a child, and a parent. The recruited children were all diagnosed with ASD. In this section, we describe the robot system, the conducted experiment, and the results that formed the foundation for the work described in this paper.

The goal of the study was to observe the children's reaction to the robot in a free-play setting and to evaluate the feasibility of using an autonomous robot for interaction with children with ASD. The parent was present for two reasons: 1) to minimize any child distress in the unfamiliar context; and 2) to pave the way for using the robot as a catalyst for social interaction with other people, as our goal is not to isolate the child in human-machine interactions only.

2.1 Recruitment

The participating children were recruited from Autism SpeaksâĂŹ Autism Genetic Resource Exchange (AGRE) [9]. The AGRE program provides bio-materials and phenotype and genotype information of families with two or more chil-



Figure 1: The humanoid robot used in the experiment

dren (multiplex) with ASD to the scientific community. Participants were eligible for inclusion in the study if they had been diagnosed with autism by AGRE researchers using a combination of the Autism Diagnostic Observation Schedule (ADOS) [13], and Autism Diagnostic Inventory (ADI) [14]. Additional inclusion criteria required that the children be between the ages of five and ten, and have a minimum verbal ability. This was determined either if the child scored above 2.0 years of age on the communication sub-scale of the Vineland Adaptive Behavior Scale [21], or were evaluated using either modules two or three of the ADOS.

All participants lived in the greater Los Angeles area. Fliers were sent to eligible families. Interested families responded by phone and email. Of the 65 families who were sent a flier, 8 responded (12.3%), and 8 children from 5 families participated in the study. Because the AGRE database consists of multiplex families, three of the participating families had siblings recruited for the study.

2.2 Robot System Design

We equipped the experiment space with an overhead camera to detect the positions of the child, the parent, the robot, and other obstacles in the room. Given *a priori* information about the layout of the room, this sensor alone was enough for the robot to maneuver autonomously in the experimental space and to use the detected positions over time to sense movement-based actions of the participants to trigger robot behavior.

We equipped the robot with two infrared (IR) emitters that emitted light visible to the overhead camera, but invisible to the naked eye so as not to be salient for the study participants. Different-sized spots were used so that the system could observe both the position and orientation of the robot.

The system used background subtraction to recognize the experiment participants (child and parent). Given that the



Figure 2: A view from the overhead camera with annotation about the pose of the robot and of the participant

camera is hard-mounted to the ceiling and the experiment room is windowless, the background of the image (the floor) is mostly static (a small amount of image static occurs as a result of normal camera operations, but can easily be filtered out). Therefore, the parent and child can both be found as foreground blobs in the image. To identify the child uniquely from the parent, we asked the parent to wear a brightlycolored t-shirt. This allowed us to reliably distinguish the parent from the child, and to thereby reliably detect and track the child's behavior. In practice, we found that the children were also willing to wear such a t-shirt. In later experiments, we had parents wear a t-shirt in one bright color, and children in another, in order to identify distinct individuals without needing foreground identification.

Each video frame, collected at 15 frames per second, serves as a single sample for detecting the locations of the interaction participants and deciding on an appropriate action. Each frame required less than 40% of the frame interval to compute leaving additional time for behavior classification, supporting the real-time needs of the system. After each frame, we are able to determine the position and orientation of the robot, the position of the child, the position of the parent, and any other obstacles that may be in the room. These positions were the data used to model the behaviors taking place during the experiment.

2.3 Experiment Protocol

Participants took part in three five-minute sessions, two with a robot and one with a non-mobile toy. The aim of the experiment was to examine the child's interaction with: an autonomous robot behaving contingently with respect to the child; a robot behaving randomly; and a toy control condition. This paper describes results obtained from data from the two robot conditions, contingent and random. A free-play scenario was deemed most suitable, in order to enable the social interaction to be as natural as possible. As a result, the children were given little specific instruction on what to do in the interaction. The parents were told that there was a chair available for sitting down, but if the children wanted to involve them in their interaction with the robot, that they could move around the room and participate.

To help alleviate any potential effects due to novelty of the robot or environment, we employed an introductory "feetwet" period, where the child, robot, parent, and experimenter were in the room together. The experimenter demonstrated each interaction behavior of the robot. This introduction ensured that the robot's behavior would not be surprising to the child, and allowed for comforting the child if he become upset with the robot. The experimenter answered any questions from the child or parent. The actual experiment proceeded only after the experimenter judged that the child understood all of the robot's behaviors, and that both the parent and child were comfortable with the robot. The "feet-wet" period typically lasted less than five minutes. At its termination, the experimenter left the child and parent in the room alone with the robot.

The contingent robot behavior was programmed to encourage social interaction. When the child approached the robot, the robot nodded its head and made an encouraging vocalization. When the child moved away, the robot acted disappointed by moving its head down and making a sadsounding vocalization. When the child pressed the button on the robot, or vocalized toward the robot, the robot blew bubbles and turned in place. When the child was behind the robot, the robot ignored the child, thereby giving the child the opportunity to hide/separate from the robot. As much as possible, the robot oriented itself to face the child, and approached the child when s/he was far away (> 1m). In all cases, the robot ignored the parent, except to avoid him/her as an obstacle.

The random robot was programmed to execute the same behaviors as the contingent robot, but at random intervals (between 17 and 23 seconds) rather than in response to the child's behavior. The robot executed the wave, head nod, head shake, spin, bubble blowing, and approach behaviors at random intervals, and otherwise executed the random walk behavior around the environment. The random robot had the same safety features and collision free navigation behaviors as the contingent robot.

2.4 Experiment Results and Comparison of Reactions to Other Robots

A total of 100 minutes of experiment time was recorded over all sessions with all participants, 54 of those involving human-robot interaction and the rest involving interaction with a non-robotic toy that was not used for this work. There is a discrepancy between the number of children in the study and the number of session minutes recorded. This is because not every child was able to continue participating in the whole study. Two sessions were left out of this analysis, and several were shortened. In all, thirteen sessions were used. In this section, we detail other work involving SAR for children with ASD, and how the children's behavior compare to the behavior which we observed during this study.

Robins et al., [18] used the KASPAR platform to encourage social interaction through a drumming game. Their system could either demonstrate a rhythmic drumming pattern for a child to imitate, or imitate a child's demonstrated drumming pattern. The results showed that a child would participate in a drumming game, even imitating the pauses between drumming sets generated by the robot. Although there were individual differences in how children reacted to the unprompted pauses, they correctly adapted to continue to play the game. Kim et al., showed similar results [1]. In contrast, our work did not use a definitive game, or a specific task to complete, so, instead, our participants created their own games. We observed children trying to get the robot to play a particular game (e.g., imitation of movements and sounds, hide and seek), whether or not the robot was capable of playing that game. Some children reacted with anxiety to the robot when it did not seem to be doing anything and was just watching them. The children who were successful in triggering the robot's behaviors would try to do whatever they did to repeatedly trigger the desired robot behavior.

Like [12], we intend to use our robot to encourage dyadic interactions between the child and robot in order to facilitate triadic interaction among the child, robot, and another person. The Keepon platform in that work used a minimallyexpressive robot to interact with children through dance, directing eye gaze, and simple expressive behaviors. In response, children with ASD would typically try to dance with the robot, meet its gaze, and talk to another person. Since our robot did not dance or attempt to direct eye gaze, we did not observe imitation of those behaviors (though we have observed spontaneous imitation of the same humanoid robot in another of our studies); we did observe children try to talk to the robot to get it to do behaviors that they had seen, including demonstrating arm movements for the anthropomorphic robot to imitate, and asking the robot to blow bubbles in order to play with them. We also observed some children attempt to engage the parent, either by explaining how they got the robot to do something, or by pointing out things that the robot was doing.

After a preliminary data coding, in which sessions were evaluated by an expert anthropologist specializing in autism, several observations were made regarding how children reacted to the robot. This expert stated that some children (n = 4, 7 sessions) had a positive impression of the robot and made several attempts to engage the robot socially. In particular, these children played with the robot when it blew bubbles and spoke to the robot in order to encourage it to socially interact with them. Some children beckoned the robot to follow them around the room. In contrast, our expert judged that some children (n = 4, 6 sessions) had a negative reaction to the robot. These negative reactions ranged from avoiding the robot, to backing up against the walls of the (rather cramped) experiment space, to staying close to the parent without interacting with the robot. All of the children who had negative reactions to the robot requested to end the study early. We classified a reaction to the robot as positive if the child was able to complete both robot sessions. We classified a reaction to the robot as negative if the child was unable to complete one or both of the robot sessions.

The children's actions could be classified into several behaviors based on what could be observed from the overhead camera. These behaviors, and their classifications, include:

- Avoiding the robot: child is moving so as to consistently increase her/his distance to the robot.
- Interaction with robot or playing with bubbles: child is moving or still and is remaining proximal to the

robot. Generally, the child was in front, or to the side, of the robot.

- Staying still: child is not moving and is not near the wall or the parent.
- Near parent: child is touching the parent or is next to the parent while not moving. This is not to suggest that being near the parent is the same as interacting with the parent. By and large, when a child was in this state, it was because the child was upset and going to the parent for comfort.
- Against wall: child is touching the wall while not moving. Most of the time, when the child was near the wall, the child was facing the robot.
- Null/none of the above: a catchall state used if none of the above conditions are met. Less than 5% of session time fit into this category.

A single human evaluator (the author) coded the data for the above behaviors. While this coding scheme and single evaluator leave room for subjectivity, the authors feel that the behaviors are sufficiently explicit (requiring only observing moving/not moving and proximity to walls and agents) that subjectivity of coding is a minor concern. An expected but important trend emerged, indicating that children who had negative reactions to the robot avoided the robot, stood against the wall, or interacted with the parent far more than the children who had positive reactions to the robot. Quantitatively, the children with a positive reaction spent more than 80% (std dev 8.9%) of the session time interacting with the robot or playing with bubbles, while the children who had a negative reaction to the robot spent less than 20%(std dev 13.4%) of the session time in those states (see Figure 6). In particular, the children who were in visible distress spent the session time with the parent or against the wall. This suggests that some distance states are correlated with a negative response to the robot, and are worth modeling.

The larger aim of this work is to detect such negative and positive reactions in order for the robot to autonomously appropriately adapt its behavior in response. For example, if the child is having a negative reaction, the robot could respond with reassuring behavior or could back away; analogously, if the child is having a positive reaction, the robot could encourage it with supportive behaviors. To enable such automated response, we wish to automatically classify child behavior data into the two categories of interest. Specifically, in this scenario, if the robot can classify the avoiding, wall, and parent states from the interaction and bubbles states, it can then detect if the child is having a positive or negative reaction. The next section describes a data-driven method for automatic data classification of behaviors in this scenario.

3. BEHAVIOR CLASSIFICATION

Various data-driven approaches to activity modeling based on position tracking have been developed [16, 22, 10, 23]. Most of these assume that events are related to absolute locations in the observed environment and utilize image coordinates as features in the model [22, 23]. This assumption works when a fixed camera observes a mostly static scene. When attempting to model social interaction, which involves



Figure 3: Percentage of session time spent in each interaction state (Group A is blue, Group B is red); group A spent a far larger amount of session time interacting with the robot than group B

movement of multiple agents, proxemic information (the interpersonal distance and/or orientation between agents) is often used as a supplemental or replacement feature [10]. Activity modeling is also used for recognition of anomalous behavior [2], gesture classification [16, 5], and detection of target behaviors, such as drowning [6]. In contrast to datadriven approaches, some work uses semantic labels of actions described *a priori* [15, 3], usually for modeling infrequent events or assigning linguistic descriptions to modeled actions.

Two lines of past work [10, 22] are particularly related to the approach described in this paper. Xiang and Gong [22] developed a data-driven approach to action selection that used a Gaussian mixture model (GMM) to identify clusters in a 7-dimensional feature space. The authors used Bayes Information Criteria (BIC) [20] to select the number of recognized actions in an unsupervised manner. Unlike our work, the features used were all directly tied to either the object being tracked (size of blob, size of bounding box) or the static properties of the environment (position in the image, change in position in image coordinates). Kelley et al. [10] described an approach that relied on interpersonal distances and velocities, but unlike in our work, the actions were selected in a supervised manner.

In contrast to the above prior work, we focus on identifying social behaviors, and on using distance information as features for an activity model. This work only assumes one environment-specific parameter, the distance of the child to the wall of the room.

To examine the feasibility of using overhead data to model a child's behavior, we conducted a simple test using distance heuristics to determine what state the child was in. For each frame: if the child was within 1.25 meters of the parent, the system observed the child as being near the parent; if the child was within 0.3 meters of the wall, the system would record an observation of being near the wall; finally if the child was behind the robot at any distance (greater than 135° or less than -135° from the front of the robot) the system would record a behind the robot observation.

We grouped the recorded sessions into two groups. Group A (n = 7) consisted of the sessions with children who liked the robot and spent a significant amount of time interacting and playing with it. Group B (n = 6) consisted of the sessions with children who did not like the robot and spent a significant amount of time avoiding it and/or seeking comfort from their parent. When we compared the percentage of session time that the automatic system annotated that the child was either near the wall, near the parent, or behind the robot for Group A and Group B, we observed that all sessions in Group A had these observations less than 40% (mean 30%, std dev 0.07) of the time, while Group B had these observations greater than 50% (mean 71.9%, std dev 0.15) of the time, a clear discrimination between the two groups. With these results, a classifier could easily be constructed based on this percentage: greater than 50%time spent in the negative behaviors would indicate a session wherein the child was not trying to interact socially with the robot, while less than 50% would indicate that the child was attempting to interact socially with the robot. However, when we compared the classified distance states to a human annotation, we achieved an accuracy of 84.98%; these heuristics only correctly identified the wall state 73% of the time and the parent state 85.03% of the time. Since these states are the most critical for identifying if the child is in distress, the accuracy of these states would need to improve to meet our goals.

This example showed that automatic data coding could be used to discriminate between children who are attempting to interact socially with a robot and children who are not. However, this method alone would not be sufficient to annotation. These results supported further investigation into the use of the overhead camera data to make determinations about what occurs during an experiment session.

Prior work using similar methods typically used the image coordinates of detected and tracked agents as part of a feature vector. However, such features assume that the relevant information can be expressed as a Cartesian position. That assumption holds for fixed-camera activitymodeling systems [22]. However, for social interaction activity modeling, fixed positions are not nearly as useful as the distances between the social actors in the scene. Therefore, we intended to use an 8-dimensional feature vector: $v = \langle d_c^r, d_c^p, d_c^w, \psi_c^r, v_c, v_c^r, v_c^w, v\psi_c^r \rangle$ where d_c^r is the distance between the child and the robot, d_c^p is the distance between the child and the parent, d_c^w is the distance between the child and the nearest wall, ψ_c^r is the orientation of the child to the robot, v_c is the absolute velocity of the child, v_c^r is the velocity of the child relative to the robot, v_c^w is the velocity of the child relative to the wall, and $v\psi_c^r$ is the change in orientation of the child to the robot (all velocities are measured over 1s). These features can easily be extracted by the overhead system described in Section 2.2. The recorded data set provided n = 48245 samples.

3.1 Model Formulation

Heuristic classification was effective for the above example, but would not scale to a full 8-dimensional feature vector where classification would include interaction between dimensions. Instead, we applied a data-driven approach to model formulation and classification: we used Gaussian Mixture Models (GMMs) to model the data.

The data were grouped into 3216 sequential tiles of 30



Figure 4: Distance states for children that had a positive reaction to the robot (Group A, left), and those that had a negative reaction to the robot (Group B, right); Group A spent less time near the walls, the parent, or behind the robot



Figure 5: The optimal model order k is the min over k of $BIC(\theta_k)$

observations (2s). A training set of 10% of the tiles was created; the other 90% was set aside for testing. The model was created by fitting a GMM using expectation-maximization (EM) to the training tiles. The initial seeding was random, and repeated 25 times in order to combat local minima. Bayes Information Criterion (BIC) [20] can determine the optimal model order (number of clusters). This method uses the log likelihood of the fitted GMM as part of a cost function of the hypothesized model order. The optimal model order (θ) is determined autonomously:

$$\theta = argmin_k(loglik(\theta_k) + BIC(n, q, k)) \tag{1}$$

where θ_k is a mixture of k Gaussians, n is the number of data points, q is the dimensionality of the data, and k is the model order. Consult [22] for a full derivation of the BIC formula. We used Matlab for the GMM implementation. Depending on the training data, the optimal model order was between 23 and 25 clusters (see Figure 5).

3.2 Classification

The classifier was constructed using the human-labeled training data where the closest cluster to the feature vector is an observation. Given the annotated behavior for each tile and the clustered observation, we can record p(o|c), where o is an observation and c is the annotated behavior for the training data. We can then classify new data by clustering the feature vectors for each frame to the model. For each tile, the classified behavior c is $argmax_c \ p(c|o)$.

4. **RESULTS**

We classified the test data set using a 5-fold validation of the data. For this exercise, 20% of the data were used for training, 80% for testing. For simplicity, we only discuss the results for the avoidance, interaction, parent, and wall behaviors since they account for 90% of session time. Overall we achieved 88.5% classification accuracy. The confusion between behaviors is as follows:

	avoidance	interaction	parent	wall
avoidance	34.7648	1.1052	3.8680	1.2587
interaction	55.8282	97.7024	25.5973	16.3636
parent	8.2822	1.0276	70.5347	3.2168
wall	1.1247	0.1648	0	79.1608

While the correct recognition of interaction behavior was largely correct, the avoidance recognition behavior was hardly better than chance (34%), and parent and wall correct recognition rates (70.5% and 79%, respectively) were somewhat worse than the correct recognition of a heuristic filter (85% and 73%, respectively). To see if model order was a possible culprit, we doubled the fitted model order and obtained the following results:

	avoidance	interaction	parent	wall
avoidance	52.7619	0.7993	1.4000	2.5850
interaction	34.8571	97.5295	7.6000	11.5646
parent	9.9048	1.5077	90.8000	3.6735
wall	2.4762	0.1635	0.2000	82.1769

For a total of 3216 tiles, 2938 were correctly identified, while 278 were incorrectly identified, for an overall correct recognition rate of 91.4%. The above confusion matrices do not have an equal allocation of each class, a much greater number of samples were from the interaction group than in the others groups, accounting for the disparity between the confusion matrices and the overall correct recognition result. These results are encouraging, especially the correct recognition of parent and wall behavior (90% and 82%, respectively), which were markedly better than the heuristic filter (85% and 73%, respectively). Since these behaviors were the main indicators of a child's negative reaction, their correct classification significantly improves the ability to discriminate between positive and negative child reactions, the stated goal of this work.

Figure 6 shows the result of the classification for Group A vs. Group B. A similar trend is observed from the humanrated annotation, where the children from group A (who had a positive reaction to the robot) spend a greater amount of time in the interaction state (78%) as compared to the



Figure 6: Classified percentage of session time spent in each interaction state (Group A is blue, Group B is red); Group A spent a far larger amount of session time interacting with the robot than group B

parent (3%), avoidance (0%), or wall (11%) states while the children from Groups B spent far less time in the interaction state (36%) and more time in the other states (2.6%, 20%, and 38%). This shows that the annotation technique can be used to differentiate between children with distinct reactions to the robot.

5. DISCUSSION AND FUTURE WORK

The goal of this work was to automatically distinguish between positive and negative reactions of children with ASD to a robot. We developed an unsupervised classifier using distance features as primitives, and classified behavior trends from sequences of actions that relate to social behavior of children with ASD in an experimental free-play setting. A secondary goal of the work was to apply this classification in such a way that it could be used in a real-time action/activity transcription system to be an input into a robot's autonomous action selection.

A child's response to a robot may be due to the robot's morphology or to its behavior. In such cases, changing the offending appearance feature, if possible, or the undesirable behavior, may be the simplest solution. Preferences of robot morphologies and behaviors by children with ASD is not a well-studied problem to date, given the novelty of the field itself, and the expense involved in developing different robot bodies capable of comparable behaviors. However, this is an important area of study; our own work has begun to address comparisons of the responses of children with ASD to mobile vs. humanoid robots when compared to a function-matched toy [7]. Future work will continue to add insights to this question.

Regardless of the robot's appearance and behavior, all children, and especially children with ASD, will not respond well 100% of the time. It is thus very important that an autonomous system be able to identify and respond to a negative reaction from a child in order to facilitate the humanmachine interaction. Some features of our robot that resulted in negative child reactions could be readily improved. The primary complaint was a high-pitched noise that the mobile base of the robot made when trying to hold its position. We eliminated the noise by disabling the motors at low speeds. Some children complained that the robot's torso was too loud when it moved and that it moved too suddenly. We are addressing this in two ways. First, we are redesigning the movements of the robot to be more smooth and continuous. Second, we have designed a mobile non-anthropomorphic robot that otherwise has the same capabilities of the robot with the torso. An ongoing study will compare children's' responses to these robots.

The overhead camera system discussed in this work was able to extract the relevant features from the recorded data. We were able to extract the positions of all the experiment participants (robot, child, and parent) as well as the orientation of the robot, and were able to obtain motion information for an effective set of distance information.

We have shown that the GMM-based method for state clustering can efficiently and effectively cluster the 8-dimensional feature space. These states are easily labeled by using annotated training data and could be used for partial behavior transcription. Potential concerns include over-generalization that can happen with human labeling, and over-specialization given the heterogeneity of the participant population. However, the BIC-based method for unsupervised model order selection did not provide a model order which was effective for modeling the data.

Our real-time goals were completely met by this strategy. Model formulation was accomplished quickly, and relevant feature detection coupled with state classification can be easily accomplished in the frame time interval (15 frames/s). In addition, the classification goals were met; we were able to differentiate between Groups A and B with high accuracy.

Continuing work on this project involves the use of motionbased features in addition to the distance-related features in order to cluster the data based on actions as well as distance. We will use this information to direct the robot to change its behavior if the child's actions warrant such a change. Our immediate application for this work is to detect interactions between the child and the robot and use that as a cue to attempt to include other participants in the interaction, and to detect when a child is in distress so that the robot could reduce emitted noise or potentially uncomfortable behavior to better meet the child's needs. Continuing work will also examine robot morphology in more detail by comparing a child's reaction to multiple robots (i.e., humanoid, non-humanoid).

Acknowledgments

The authors thank Prof. Shri Narayanan, Prof. Marian Williams, Matthew P. Black, Aaron St. Clair, Dr. Elisa Flores, and Emily Mower for their help with data collection, and Prof. Peter Mundy and Prof. Michele Kipke for their invaluable help in various aspects of this work.

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Prof. Clara M. Lajonchere (PI). We thank Ryan Butler for his assistance with recruitment. This work was supported by the National Science Foundation (CNS-0709296, IIS-0803565), the Nancy Lurie Marks Family Foundation, the LA Basin Clinical and Translational Science Institute, and the Dan Marino Foundation.

6. REFERENCES

- Robotic tools for prosodic training for children with asd: A case study. In *International Meeting for Autism Research (IMFAR)*, London, UK, 2008.
- [2] O. Boiman and M. Irani. Detecting irregularities in images and in video. International Journal of Computer Vision, 74(1):17–31, 2007.
- [3] M. Chan, A. Hoogs, J. Schmiederer, and M. Petersen. Detecting rare events in video using semantic primitives with HMM. In *Proceedings of the 17th International Conference on Pattern Recognition* (*ICPR*), volume 4, pages 150–154, Cambridge, England, UK, 2004.
- [4] K. Dautenhahn, I. Werry, J. Rae, P. Dickerson,
 P. Stribling, and B. Ogden. Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot. In
 K. Dautenhahn, A. Bond, L. Canamero, and
 B. Edmonds, editors, Socially Intelligent Agents: Creating Relationships with Computers and Robots, pages 117–124. Kluwer Academic Publishers, 2002.
- [5] A. Efros, A. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *Proceedings of the International Conference on Computer Vision*, pages 726–733, Nice, France, 2003.
- [6] H. Eng, K. Toh, A. Kam, J. Wang, and W. Yau. An automatic drowning detection surveillance system for challenging outdoor pool environments. In *Proceedings* of the International Conference on Computer Vision, pages 532–539, Nice, France, 2003.
- [7] D. J. Feil-Seifer and M. J. Matarić. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In 11th International Symposium on Experimental Robotics 2008, volume 54, pages 201–210, Athens, Greece, Jul 2008. Springer.
- [8] D. J. Feil-Seifer and M. J. Matarić. Using proxemics to evaluate human-robot interaction. In *Proceedings of* the International Conference on Human-Robot Interaction, pages 143–144, Osaka, Japan, Mar 2010.
- [9] D. Geschwind, J. Sowinski, C. Lord, P. Iverson, L. Ducat, and S. Spence. The autism genetic resource exchange: A resource for the study of autism and related neuropsychiatric conditions. *American Journal* of Human Genetics, 69:461–463, 2001.
- [10] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis. Understanding human intentions via hidden markov models in autonomous mobile robots. In *Proceedings of the Conference on Human Robot Interaction*, pages 367–374, Amsterdam, Netherlands, March 2008.
- [11] H. Kozima, C. Nakagawa, and Y. Yasuda. Children-robot interaction: a pilot study in autism therapy. *Prog Brain Res*, 164:385–400, 2007.
- [12] H. Kozima, Y. Yasuda, and C. Nakagawa. Social interaction facilitated by a minimally-designed robot: Finding from longitudinal therapeutic practices for

autistic children. In *Proceedings of the International* Conference on Robot and Human Interactive Communication, pages 599–604, August 2007.

- [13] C. Lord, S. Risi, L. Lambrecht, E. H. C. Jr., B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3):205–223, Jun 2000.
- [14] C. Lord, M. Rutter, and A. L. Couteur. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5):659–685, Oct 1994.
- [15] S. Park and J. Aggarwal. Semantic-level understanding of human actions and interactions using event hierarchy. In *Conference on Computer Vision* and Pattern Recognition, pages 12–12, 2004.
- [16] D. Ramanan and D. Forsyth. Automatic annotation of everyday movements. In S. Thrun, L. Saul, and B. Schölkopf, editors, Advances in Neural Information Processing Systems 16, Cambridge, MA, 2004. MIT Press.
- [17] B. Robins, K. Dautenhahn, R. 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, 4(2):105–120, 2005.
- [18] B. Robins, K. Dautenhahn, R. te Boekhorst, and C. Nehaniv. Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinesics. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*, pages 17–24, Amsterdam, The Netherlands, March 2008. ACM New York, NY, USA.
- [19] B. Scassellati. Quantitative metrics of social response for autism diagnosis. In *IEEE International Workshop* on Robots and Human Interactive Communication (ROMAN), pages 585–590, Nashville, TN, Aug 2005.
- [20] G. Schwarz. Estimating the dimension of a model. The annals of statistics, 6(2):461–464, 1978.
- [21] S. Sparrow, D. Balla, and D. Cicchetti. Vineland adaptive behavior scales. Test Corp of Amer, 1984.
- [22] T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67(1):21–51, 2006.
- [23] H. Zhong, J. Shi, and M. Visontai. Detecting Unusual Activity in Video. In Proceedings of the Conference on Computer Vision and Pattern Recognition, volume 2, pages 819–826, Washington, DC, USA, 2004. IEEE Computer Society.