2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Understanding a Child's play for Robot Interaction by Sequencing Play Primitives using Hidden Markov Models

Hae Won Park, Ayanna M. Howard, *Senior Member, IEEE*

Abstract— **In this paper, we discuss a methodology to build a system for a robot playmate that extracts and sequences low-level play primitives during a robot-child interaction scenario. The motivation is to provide a robot with basic knowledge of how to manipulate toys in an equivalent manner as a human does - as a first step in engaging children in cooperative play. Our approach involves the extraction of play primitives based on observation of motion gradient vectors computed from the image sequence. Hidden Markov Models (HMMs) are then used to recognize 14 different play primitives during play. Experimental results from a data set of 100 play scenarios including child subjects demonstrate 86.88% accuracy recognizing and sequencing the play primitives.**

I. Introduction

WHY do children need playmates? Interactive play in childhood is closely linked to children's social, physical, and cognitive development [1]. However, due to many social factors, children are often left alone, spending hours of time watching television, playing video games, and computers, which threatens to undermine the process of play, with grim implications for the intellectual and emotional health of children [2, 3].

Simple toys, such as those depicted in Fig. 1, can accelerate a child's imagination as they build their own scenes, knock them down, and start over. Along with the toys, playmates are also an important source for building collaboration and cooperation, controlling impulses, reducing aggression, and having better overall emotional and social adjustments [4, 5]. Children with development delays can benefit from a robotic toy, which can yield better attention [6, 7]. Robots also have shown the potential in assisting physically challenged children [8], and in engaging children in imitation base play [9]. Although robots are shown to be of use in these various children-robot interaction scenarios, robots, in these venues, are positioned more as tools rather than partners or playmates. Long-term interaction and the effectiveness of robot usage in interactive play therefore has not reached its full potential.

The effect of playing has shown to have a lasting effect due to the dynamic nature of interacting with the world [10]. With respect to playing with others, a shared interest arises between playmates to make the play continuously entertaining, thus engaging the mind, and creating opportunities for extended play over longer durations. Based on this theory, in order to transition robots from toys to playmates the first challenge to be addressed is enabling the robot to become an *effective* playmate. We believe that by observing others play, the robot can effectively learn acceptable play behaviors. This type of research is closely aligned with research focused on learning manipulation tasks. Unfortunately, most research that addresses learning manipulation tasks from human-robot interaction tries to address the problem associated with general learning [11-15]. In addition, gesture recognition, another component of understanding play behavior, is often done by means of colored gloves [16] or data gloves as described in [17].

In contrast, the intent of the approach proposed within is the use of dynamic pattern recognition methods, using only visual information without further aids. There are many earlier gesture recognition works that use vision-based pattern recognition technique. Yamato *et al.* [18] uses discrete Hidden Markov Models (HMMs) to recognize image sequences of six different tennis strokes among three subjects. Starner *et al.* [19] describe an extensible system which uses one color camera to track hands in real time and interprets American sign language (ASL). They use HMMs to recognize a full sentence and demonstrate the feasibility of recognizing a series of complicated gesture. Darrell and Pentland [20] uses dynamic time warping to match the interpolated responses of several learned image templates. The work of Calinon *et al.* [13] considered learning trajectories and constraints from demonstrations for robotic tasks. Their approach aims at extracting the relevant characteristics of the gesture that needs to be reproduced. Motion data is encoded using a mixture of Gaussian/Bernoulli distributions (GMM/BMM) which provides a measure of the spatio-temporal correlations across the different modalities. Among the various pattern recognition techniques that exist, we have chosen HMMs due to their reliability and simplicity in modeling sequential patterns, as shown in these related research efforts.

Our previous work has shown that through observation, a robot can successfully sequence low-level motion behaviors into a proper order to form a similar play action [21]. This work extends on that concept by developing a library of play actions based on low-level motion behaviors, such that a general methodology can be established for extracting play primitives without a-priori knowledge. In Section II, we describe further the concept of play primitives. Section III presents the detailed approach for play primitive recognition using Hidden Markov Models. Preliminary training and experimental results are presented in Section IV and Section V provides concluding remarks.

This work was supported in part by the National Science Foundation under Award Number IIS-0705130. Seed funding to begin the robotic playmate effort was provided by the Center for Robotics and Intelligent Machines (RIM@GT)

Hae Won Park and Ayanna Howard are with the School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA, USA (e-mail: haewon.park@gatech.edu).

II. Child Play Primitives

Baranek *et al.* describe in their work a subset of toy manipulation that can be used towards screening a child's developmental stage [22]. The list contains the behaviors shown in the first five levels of object play: *grasp, rub, shake, bang, mouth, roll, pull apart, stack, scoop, pour, cover, and join*. Based on this list, we conduct a research study with public sources from the web to identify the kinds of basic motions that form these manipulations when children interact with various types of toys. With regards to a robotic playmate, these basic motions are what we further define as *play primitives*.

Video sequences of children playing with various toys were gathered from 25 video sources from YouTube. Play scenarios in these data set consisted of building blocks, stacking cups, inserting blocks into bins, hammering tables, and etc. (Fig. 1). The videos consisted of 32,676 valid frames, where only frames that contained an image of a child actually interacting with a toy were classified as valid. The play primitive identified in each valid frame was then determined and used to calculate the ratio of primitives initiated by the child.

Observations made during this study are as depicted in Fig. 2. As can be seen, many of the play scenarios begin with picking up a toy in an upward direction. Younger children tend to shake and drop the toy more frequently. Older Children, over three years old, are more accurate in manipulating the toy in specific directions. The seven most distinct primitives (94.21%) found from our study are renditions of the behaviors in Baranek *et al.*'s list, and thus provide strength to the defined play primitives in this paper. For instance, the *Stack* behavior from the list is a sequence of the basic motions (or primitives) of *moving up-left/right-down*, and the *Roll* behavior can be a repeated sequence of the basic motion of *horizontal shaking*. *Bang* is another repeated behavior of the *vertical shaking* primitive. Other primitives that were observed less frequently are spinning, pressing, and hugging. In addition to these primitives, three kinds of final resting state of the manipulated toy were observed: insert (44%), stack (32%), and drop (24%). The disappearing of the play object characterizes the inserting action while stacking and dropping are distinguished by whether the toy rests on top of another toy.

The play primitives we implement throughout the rest of this paper are based on the statistics we learned from this study. Fig. 3 shows the final fourteen play primitives used in our database.

III. Technical Approach

To endow a robot with the ability to learn acceptable play behavior, we use a three-step process consisting of 1) Preprocessing, 2) Motion Gradient Extraction, and 3) Play Primitive Recognition and Learning. Starting with the original image sequence of a play scenario, the preprocessing detects and tracks the toys in a scene. The motion gradient extraction calculates a vector for each frame that includes direction and size information. Hidden Markov Models (HMMs) [23] are then used to recognize and classify play primitives, represented by the vector sequence computed by the motion gradient extraction unit. The output of the system is the sequence of the primitives (Fig. 4).

Fig. 1. Examples of toys children interact in their everyday play.

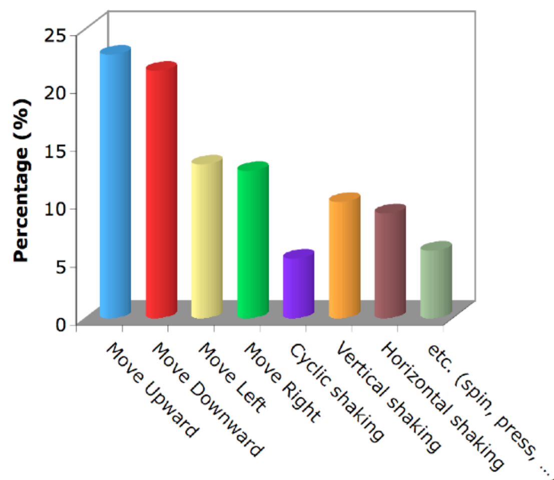


Fig. 2. Percentage of each play primitives observed through child play with toys.

A. Preprocessing

When play is initiated, the system first observes the entire play scenario. Since the main focus of this research is to understand how a child interacts with toys, we track toys in the image scene versus the child's body parts. And since most children's toys use saturated colors to keep visual attention, color is used as the key parameter for toy detection. As described in the previous work [21], histogram back-projection and filtering techniques are used for the

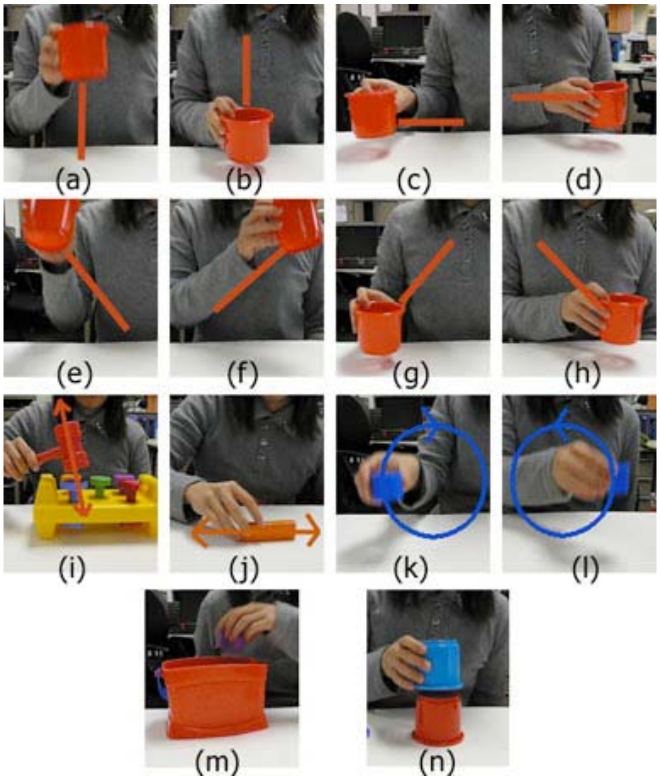


Fig. 3. The fourteen most frequently repeated gestures are defined as Play Primitives. (a)Up (b)Down (c)Left (d)Right (e)Up-Left (f)Up-Right (g)Down-Left (h)Down-Right (i)Vertical-shaking (j)Horizontal-shaking (k)Circular-shaking(clockwise) (l)Circular-shaking(counter-clockwise) (m)Insert (n)Stack/Drop

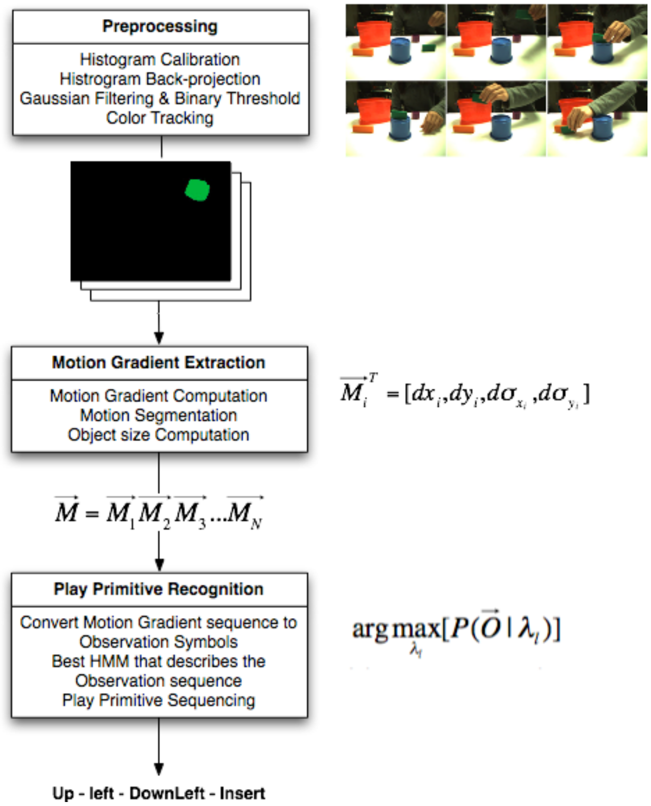


Fig. 4. Structure of the overall system. Preprocessed images are passed on to the Motion Gradient Extraction module where each frame is converted into a 4-dimensional motion vector. In the Play Primitive Recognition unit, the motion vectors are quantized into observation symbols and the best HMM that describes the observation sequence is selected.

detection process. Back-projection is a way of representing how likely each pixel fits the distribution of pixels in a histogram model [24, 25]. The calibration step is utilized in order to cope with variously changing illumination conditions, which involves redefining the histograms. Once computed, it is used to assign a probability value to each image pixel in subsequent video frames. As a new frame arrives, the hue and saturation value for each pixel is determined. From that, each color histogram is used to assign a probability to the pixel. In the case of our toy hue-saturation histogram model, if C is the color of the pixel and T is the probability that a pixel is a toy, then this probability map gives us $p(C|T)$, the probability of drawing that color if the pixel actually is a toy. Combining with the total probability of encountering a toy-colored object in a scene $p(T)$ and the total probability of encountering the range of toy colors $p(C)$, we can compute $p(T|C)$ according to Bayes' theorem,

$$p(T \mid C) = \frac{p(T)}{p(C)} p(C \mid T) \tag{1}$$

This process allows us to identify, with high probability, all toy objects within an image scene. Among the multiple toys detected from this process, we label the first one upon which an action is taken as the *play* object. This toy is then tracked and recorded until it comes to a complete stop. Fig. 5 depicts the preprocessing step.

B. Motion Gradient Extraction

From the preprocessed image, normalized motion gradients (dx_i, dy_i) are extracted using two adjacent frames imaged during the play sequence. The gradients are calculated using the following formula,

$$m_{x_i} = \frac{\sum_{x,y \in D_i(x,y)} x}{N_i} \qquad m_{y_i} = \frac{\sum_{x,y \in D_i(x,y)} y}{N_i}$$

$$dx_i = \frac{(m_{x_i} - m_{x_{i-1}})}{\sqrt{(m_{x_i} - m_{x_{i-1}})^2 + (m_{y_i} - m_{y_{i-1}})^2}} \tag{2}$$

$$dy_i = \frac{(m_{y_i} - m_{y_{i-1}})}{\sqrt{(m_{x_i} - m_{x_{i-1}})^2 + (m_{y_i} - m_{y_{i-1}})^2}}$$

where (m_{x_i}, m_{y_i}) is the mean pixel value of the region $D_i(x,y)$ which represents the region of the detected object

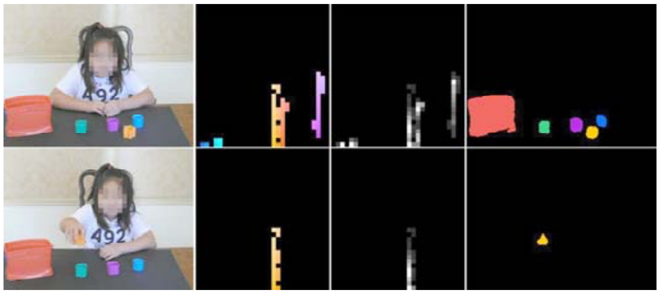


Fig. 5. Preprocessing of the play scene. (Top) Initial play scene, H-S histogram color map, H-S histogram probability map, and the toys detected. (Bottom) The play object is being tracked.

in the *i*-th frame. N_i is the number of pixels in region $D_i(x,y)$.

These features indicate the direction of movement of the play object. The identified sequences of directional features are the key characteristics used as input into the HMM for play primitive recognition and learning, as discussed in Section III. *C*.

Another useful feature used in play primitive recognition is the ratio of object spread, which can be calculated from the change in object size defined as,

$$d\sigma_{x_i} = \frac{\sum_{x,y\in D_i(x,y)} (x - m_{x_i})^2 \Big/ N_i}{\sum_{x,y\in D_{i-1}(x,y)} (x - m_{x_{i-1}})^2 \Big/ N_{i-1}}$$
$$d\sigma_{y_i} = \frac{\sum_{x,y\in D_i(x,y)} (y - m_{y_i})^2 \Big/ N_i}{\sum_{x,y\in D_{i-1}(x,y)} (y - m_{y_{i-1}})^2 \Big/ N_{i-1}} \tag{3}$$

Here, $(d\sigma_{x_i}, d\sigma_{y_i})$ is a ratio of the variance between the *i*-th and *(i-1)*-th frames. The size information can be used to distinguish between the play object's final resting state: insert or stack/drop. As the study discussed in Section II shows, since these three states are dominant in child play, we make the assumption that the toy, with high probability, will be inserted, stacked, or dropped during a play scenario. The inserted state is defined by disappearing of the object after a downward primitive action towards another toy. The stacked and dropped condition are recognized by the same HMM, but distinguished afterwards by whether the play object is on top of another toy or not.

From the four features introduced above, a 4-dimensional motion vector,

$$\overrightarrow{M_i}^T = [dx_i, dy_i, d\sigma_{x_i}, d\sigma_{y_i}] \tag{4}$$

is derived for every frame resulting in a vector sequence $\vec{M} = \overrightarrow{M_1}\overrightarrow{M_2}\overrightarrow{M_3}...\overrightarrow{M_N}$.

C. Play Primitives Recognition and Learning

In this research, Hidden Markov Models (HMMs) are used to recognize and sequence the fourteen play primitives shown in Fig. 3. HMMs have advantages in modeling sequential patterns such as speech. Our approach models a behavior as a sequence of play primitives. In effect, we model a play behavior with temporally sequenced play primitives, which is analogous to the representation of a word using a sequence of phonemes in speech recognition. It is therefore understandable why these techniques developed for speech recognition would perform well in our approach. The motion vectors $\overrightarrow{M_1}, \overrightarrow{M_2}, \overrightarrow{M_3}...$ are converted into discrete observation symbols $\overrightarrow{O_1}, \overrightarrow{O_2}, \overrightarrow{O_3}...$ before input into the HMMs.

1. Hidden Markov Models (HMMs)

The HMMs are doubly stochastic processes, which are an extension of discrete Markov chains, which cannot be directly observed. The special case of a discrete HMM is represented by three matrices, $\lambda = (A, B, \pi)$. The matrix $A = \{a_{ij}\}$ specifies the state transition probability from state i to j. $B = \{b_{jk}\}$ represents the probability of generating symbol k from state j, and π indicates the initial state probability distribution matrix.

There are three basic problems that must be solved for the real application of the HMM [23]: evaluation, decoding, and training (optimizing). The solutions to each of these problems are found using variations on the Forward- Backward algorithm, the Viterbi algorithm, and the Baum-Welch algorithm.

In this paper, fourteen different HMMs were trained, one for each play primitive. The Left-right model, also known as the Bakis HMM, is used to model the non-cyclical motions such as the first eight primitives and the two final state primitives. The repeated gestures such as the four shaking primitives are modeled using the cyclic HMM as shown in Fig. 6.

2. Vector Quantization

The extracted motion gradient features are quantized into discrete symbols to apply to the HMMs. Taking into account the important characteristics of the play primitives, the feature space is divided into 18 clusters. The first classification process uses the directional information (dx, dy), which is quantized into eight regions using the minimum least squared error (Fig. 7 Top). The size ratio $(d\sigma_x, d\sigma_y)$ is a crucial data in distinguishing the final resting state primitives, therefore forms another category. A steady motion and a disappearance of the toy are defined as independent codes. Final 18 codes are depicted in Fig. 7.

The simple structure of this codebook enhances the versatility in recognizing newly introduced primitives, which could be added later on with different kinds of toys.

3. Play Primitive Recognition

The recognition result of each observation sequence is the HMM with the highest probability $\arg\max_{\lambda_l}[P(\vec{O} \mid \lambda_l)]$ which is computed by the Viterbi algorithm. However, the result with a probability under some threshold is discarded as an unknown. Final primitives recognized with sufficient probability are sequenced to form a complete play action.

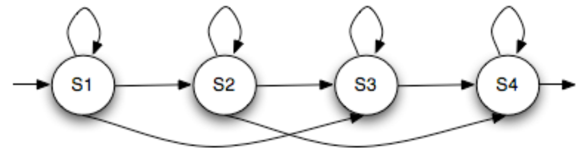
IV. Training and Experimental Results

In order to train the HMMs, we collected 20 play scenarios from three adult researchers. To verify the system, 30 scenarios were gathered from three adult subjects and 70 scenarios from three child subjects. Three different cameras were used in three different environments in order to verify the system's capability of tracking play primitives in low and high resolution images, and adaptability to illumination changes. (Fig. 8)

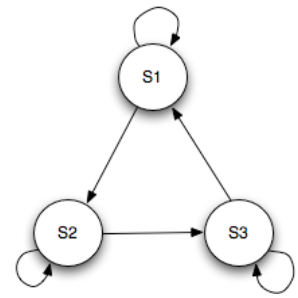
The scenarios consisted of inserting, stacking, rolling, hammering, shaking, and dropping with various speeds, some with simple trajectories and others having longer interaction with the toy. Play scenarios ranged from 2 to 30 seconds.

Using the HMM structure stated in Section III, we collected 5 to 15 play primitive data sets (Fig. 3) for each primitive extracted from the 20 training play scenarios. The vector sequence $\vec{M}$ received from the motion gradient extraction module was quantized into a sequence of discrete observation symbols $\vec{O}$ based on the customized VQ codebook. The data for each play primitive was simultaneously used as inputs to the Baum-Welch algorithm, which ran until it converged. Experiments show that it took an average of 4 iterations to achieve over 95% convergence.

The 30 test scenarios were conducted with three adult subjects in a same environment as the training data set, but different from the ones used to train the HMMs. Three child subjects each performed 20~25 scenarios, and were video taped at their homes. Scenarios varied in contents from a very simple pick up and insert operation to the one shown in Fig. 9. Test scenarios were conducted to examine the capability of the system in recognizing the play primitives and correctly sequencing them together. The average frame rate of the overall algorithm was 21.6 fps. Fig. 10 shows the recognition rate of each play primitives, and Fig. 11 is the confusion matrix. The average play primitive recognition rate for adults was 94.51% while sequencing was performed with 100% accuracy based on the play primitives that were recognized. The average play primitive recognition rate for child subjects was 83.61%, resulting in 86.88% overall performance. The lower rate is due to children becoming too excited and manipulating the toy outside the camera frame, operating two kinds of toys at once, and objects being

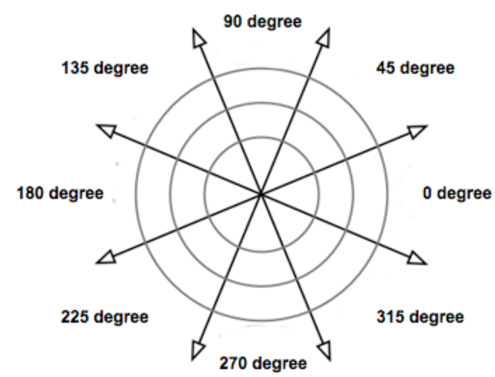


(a)



(b)

Fig. 6. (a) Left-Right HMM used to model sequential primitives and (b) Cyclic HMM used to model repeated shaking primitives



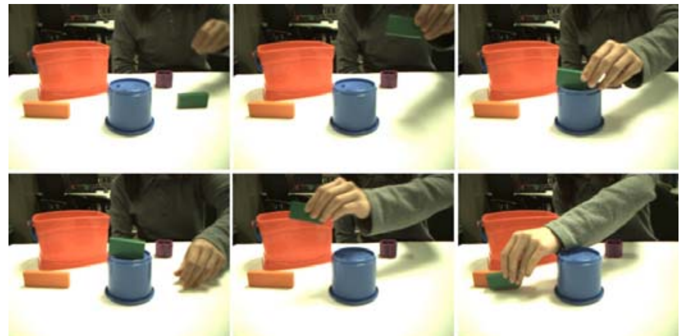
Symbol	Direction	Size ratio	Symbol	Direction	Size ratio
1	0°	> 0.83	10	0°	< 0.83
2	45°	> 0.83	11	45°	< 0.83
3	90°	> 0.83	12	90°	< 0.83
4	135°	> 0.83	13	135°	< 0.83
5	180°	> 0.83	14	180°	< 0.83
6	225°	> 0.83	15	225°	< 0.83
7	270°	> 0.83	16	270°	< 0.83
8	315°	> 0.83	17	315°	< 0.83
9	Steady		18	Out-of-sight	

Fig. 7. VQ Codebook. (Top) 8 directional regions (Bottom) Classification of each observation symbols

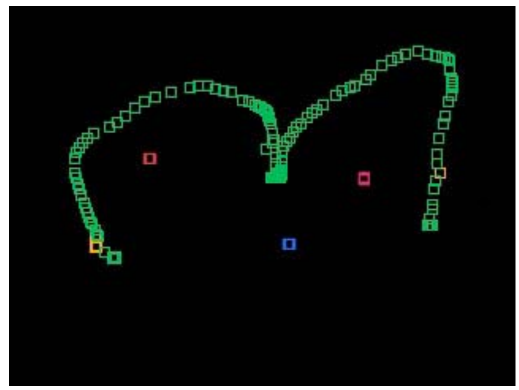
unidentified under severe light condition. The average recognition rate for the sequential play primitives using left-right model was 88.85% while repeated shaking primitives using a cyclic model resulted in 81.95% success rate. Even the difficult primitives such as the circular

Subjects	Test Environment	Resolution	Frame rate	Number of test scenarios
3 Adults	Lab	320x240	30fps	30 (10 each)
Child #1	Home	640x480	30fps	28
Child #2	Home	320x240	30fps	22
Child #3	Home	720x480	30fps	20

Fig. 8. Subjects in various illumination conditions, and different camera resolutions. (Top Left) Adult subject in controlled lab environment. (Others) Child subjects in their natural home environment.



(a)



(b)

Fig. 9. Example test scenario. (a) 26th, 66th, 110th, 129th, 186th, and 214th frame of the scenario (b) Trajectory of the play object. Sequence was correctly identified as <UP-LEFT-DOWNLEFT-DOWN-STACK-UP-UPLEFT-LEFT-DOWNLEFT-DOWN-DROP>

shaking motions were recognized with 78.48%, which is a satisfying result.

The training and experiment result show that most of the behaviors during a play can be understood by sequencing low-level play primitives. This fact is very encouraging in a sense that when applied to a robot playmate, the robot can observe and take turns with children while engaging in a daily basis and therapeutic play.

V. Conclusions and Future Work

This work presents the basic therapeutic concept of child play: engaging and interacting. Some articles emphasize that everyday of childhood should be a day of play [26-28]. In a therapeutic aspect, this paper presents a promising recognition system for a robot playmate that has huge potential in engaging a child in play using various types of toys.

The play primitive based behavior recognition system presented in this paper has several advantages. First, unlike most earlier works on gesture recognition which tries to recognize a specific motion in whole, such as recognizing "hello," "good-bye," and "rotate" [29], this system decomposes a large play behavior motion data to a temporally-sequenced play primitives which can be modeled with first order Markov process. Second, this allows us to extend the recognition rate to any play behavior, especially when dealing with unpredictable children. Finally, the recognition can be implemented in real-time using any low-resolution single camera. The real-time aspect has the potential for interactively learning new play primitives while in play.

We have first addressed here the most frequently observed play actions while a child is interacting with toys, and based on the study defined the fourteen play primitives implemented through out the research. The recognition of play primitives was modeled using Hidden Markov Models (HMMs) approach. The preprocessing step identifies toys in a play scene by back-projecting the 2-dimensional hue-saturation histogram to the input frame, and tracks the object in action. The calibration step is added to refine the histograms in order to adapt to changing illumination conditions. The Motion Gradient Extraction module extracts the four-dimensional motion vector that contains the directional and size information. The motion vector sequence is then converted into discrete observation symbols, which is used towards training HMMs and recognizing the play primitive. We have shown that the proper combination of these features can yield satisfying recognition rate.

Further work will be conducted with the robot platform that is focused on taking turns with its child partner. Based on the observation and prediction of a child interacting with a toy, the robot will choose its behavior and action.

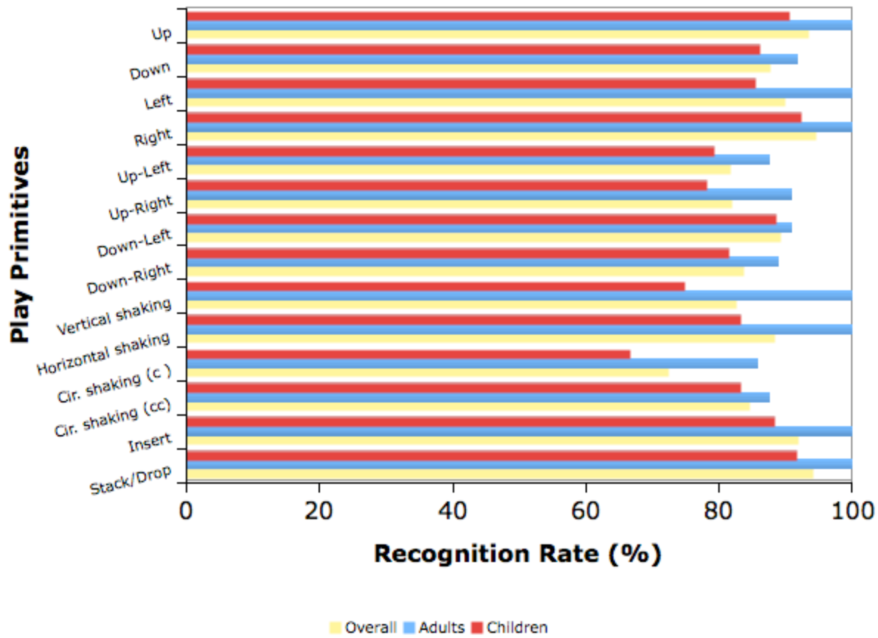


Fig. 10. Individual recognition rate for each Play Primitive. Average rate for adult subjects is 94.51 %, child subjects 83.61%, and overall 86.88%

		Actual													
		Up	Down	Left	Right	Up-Left	Up-Right	Down-Left	Down-Right	Vert. Shak.	Hor. Shak.	Cir. Shak. (c)	Cir. Shak. (cc)	Insert	Stack/Drop
Predicted	Up	91		1		5	4			2			1		
	Down		73					1	3	1				1	1
	Left	2	2	62		2									
	Right				57		1					1			
	Up-Left	2		2		49									
	Up-Right						35								
	Down-Left			1				49							
	Down-Right								3						
	Vert. Shak.	1								2		2	1		
	Hor. Shak.			1	1						12				
	Cir. Shak. (c)											1			
	Cir. Shak. (cc)				1								12		
	Insert		6					2	3					61	3
	Stack/Drop							1						2	59
	Unidentified	3	3	4	2	5	3	2		1	1			3	
	total	99	84	71	61	61	43	55	36	24	13	13	14	67	63

Fig. 11. Overall Play Primitive confusion matrix. Unidentified includes insufficient likelihood, children manipulating toys out of view, and etc.

REFERENCES

[1] J. Piaget, *Play, dreams and imitation in childhood*. New York: Norton, 1962.

[2] J. Funk, H. Baldacci, T. Pasold, and J. Baumgardner, "Violence exposure in real-life, video games, television, movies, and the internet: is there desensitization?" *Journal of Adolescence*, vol. 27, no. 1, pp. 23–39, 2004.

[3] L. Huesmann and L. Taylor, "The role of media violence in violent behavior," *Annual Review of Public Health*, 2006.

[4] L. McMahon, *The handbook of play therapy*. Routledge, 1992.

[5] J. Moor, *Playing, laughing and learning with children on the autism spectrum: a practical resource of play ideas for parents and carers*. Jessica Kingsley Publishers, 2008.

[6] A. Billard, B. Robins, J. Nadel, and K. Dautenhahn, "Building robota, a mini-humanoid robot for the rehabilitation of children with autism," *RESNA Assistive Technology Journal*, vol. 19, no. 1, pp. 37–49, 2006.

[7] B. Scassellati, "How social robots will help us to diagnose, treat, and understand autism," in *Proc. of the 12th International Symposium of Robotics Research (ISSR05)*, 2005.

[8] G. Kronreif, B. Prazak, S. Mina, M. Kornfeld, M. Meindl, and M. Furst, "Playrob - robot-assisted playing for children with severe physical disabilities," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, 2005, pp. 193–196.

[9] 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.

[10] K. Ginsburg, "The importance of play in promoting healthy child development and maintaining strong parent-child bonds." *Pediatrics*, vol. 119, no. 1, pp. 182–91, 2007.

[11] S. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE transactions on robotics and automation*, vol. 13, no. 1, pp. 81–95, 1997.

[12] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching: extracting reusable task knowledge from visual observation of human performance," *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 6, pp. 799–822, Dec 1994.

[13] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 2, pp. 286–298, April 2007.

[14] N. Pollard and J. Hodgins, "Generalizing demonstrated manipulation tasks," in *In Proceedings of the Workshop on the Algorithmic Foundations of Robotics (WAFR 02)*, 2002.

[15] C. Atkeson and S. Schaal, "Robot learning from demonstration," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML '97)*, 1997, pp. 12–20.

[16] Y. Iwai, K. Watanabe, Y. Yagi, and M. Yachida, "Gesture recognition using colored gloves," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 1, Aug 1996, pp. 662–666 vol.1.

[17] C. Lee and Y. Xu, "Online, interactive learning of gestures for human/robot interfaces," in *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, vol. 4, Apr 1996, pp. 2982–2987 vol.4.

[18] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, Jun 1992, pp. 379–385.

[19] T. Starner and A. Pentland, "Visual recognition of american sign language using hidden Markov models." in *International Workshop on Automatic Face and Gesture Recognition*, 1995, pp. 2982–2987.

[20] T. Darrell and A. Pentland, "Space-time gestures," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on*, June 1993, pp. 335–340.

[21] A. Howard, H. W. Park, and C. Kemp, "Extracting play primitives for a robot playmate by sequencing low-level motion behaviors," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, Aug. 2008, pp. 360–365.

[22] G. Baranek, C. Barnett, E. Adams, N. Wolcott, L. Watson, and E. Crais, "Object play in infants with autism: Methodological issues in retrospective video analysis," *American Journal of Occupational Therapy*, vol. 59, no. 1, pp. 20–30, 2005.

[23] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[24] Intel Corporation, *Open source computer vision library reference manual*, 2001.

[25] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[26] A. Pellegrini and P. Smith, "Physical activity play: The nature and function of a neglected aspect of play," *Child development*, vol. 69, no. 3, pp. 577–598, 1998.

[27] S. Oliver and E. Klugman, "Building a play research agenda: What do we know about play? What new questions do we need to ask?." *Exchange: The Early Childhood Leaders*, vol. 173, p. 4, 2007.

[28] R. Coplan, K. Rubin, and L. Findlay, "Social and nonsocial play," in *Play from birth to twelve: contexts, perspectives, and meanings*. New York: Garland, 2006, pp. 75–86.

[29] J. Schlenzig, E. Hunter, and R. Jain, "Recursive identification of gesture inputs using hidden markov models," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, Dec 1994, pp. 187–194.