Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3, 2008

Extracting Play Primitives for a Robot Playmate by Sequencing Low-Level Motion Behaviors

Ayanna M. Howard, *Senior Member, IEEE,* Hae Won Park, and Charles C. Kemp

***Abstract*— In this paper, we discuss a methodology to extract play primitives, defined as a sequence of low-level motion behaviors identified during a playing action, such as stacking or inserting a toy. Our premise is that if a robot could *interpret* the basic movements of a human's play, it will be able to interact with many different kinds of toys, in conjunction with its human playmate. As such, we present a method that combines motion behavior analysis and behavior sequencing, which capitalizes on the inherent characteristics found in the dynamics of play such as the limited domain of the objects and manipulation skills required. In this paper, we give details on the approach and present results from applying the methodology to a number of play scenarios.**

I. Introduction

INTERACTIVE play with physical toys has an important role in the development of cognitive, physical, and social development in children [1]. Robots as toys has shown to aid in early intervention for children with development delays [e.g. 2,3], assisting with physically challenged children [e.g. 4], and to engage children in imitation base play [e.g. 5]. 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. As such, in practice, it has been shown that most "commercially available robots seldom cross the 10-h barrier [i.e. 10 hours of combined human-robot interaction]'' typically found between humans and their life-long pet companions [6]. Long-term interaction and thus 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 [7]. 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. We believe that a step necessary to transition robots from toys to playmates involves providing them with manipulation capability so that they can physically interact with toys in conjunction with their human (child) partner. This is compatible with the play theory that exploration and manipulation are prerequisites to meaningful play experiences [8].

Based on this theory, the challenge that must first be addressed is whether a robot could become an effective playmate. We believe that by observing others play, a robot can effectively learn acceptable play behavior. This is similar in nature to learning from observation and/or demonstration. Most research though that addresses learning manipulation tasks from human-robot interaction derived from observation and/or demonstration tries to address the problem associated with a single or a set of arbitrary manipulation tasks [e.g. 9-13]. Additionally, observing a human perform a task does not imply that the robot will be able to perform the task, and, in fact, typically requires that the robot has the ability to determine how to move its body with respect to the human. In contrast, we propose that if the robot could *interpret* the basic movements of a human's play, it will be able to interact with many different kinds of toys, in conjunction with its human playmate.

As such, in this paper, we present a methodology designed to extract play primitives, defined as a sequence of low-level motion behaviors identified during a playing action, such as stacking or inserting a toy. This methodology is designed as the first steps to endow a robot with the ability to identify general play behavior associated with manipulation tasks related to play activities. The primary algorithm consists of two key components - Motion Behavior Analysis and Behavior Sequencing. We give details of the method and discuss results applied to toy manipulation in two play scenarios.

II. Defining Play Primitives

One of the benefits of playing with toys through manipulation is to stimulate the development of fine motor skills, which require control of small, specialized motions using the fingers and hands. These skills evolve over time starting with primitive gestures, such as grabbing [14]. By examining the interaction of children with toys, a number of common primitive gestures can be extracted, including acts of grasping, transporting, inserting, hammering, stacking, and pushing. With regards to a robotic playmate, these primitive gestures are what we further define as play

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).

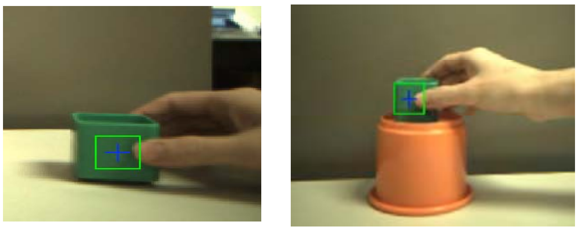
Ayanna Howard and Hae Won Park are with the School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA, USA (phone: 404-385-4824; e-mail: ayanna.howard@ece.gatech.edu).

Charlie Kemp is with the Wallace H. Coulter Department of Biomedical Engineering at Georgia Institute of Technology and Emory School of Medicine, Atlanta, GA, USA.

primitives.

In various studies conducted on the dynamics of play, a number of mutually exclusive categories of play have been derived [15, 16]. Of these, two are associated with physical manipulation - 1) functional play, which involves the conventional use of objects (Figure 1a) and 2) relational play, which is defined as the association of two or more objects together (Figure 1b). As such, we define two types of robot play primitives:

- Functional play primitives - hand/fingers manipulate a single toy object
- Relational play primitives - hand/fingers manipulate a single toy object such that it makes contact with another toy object.



(a) (b)

Figure 1. Examples of play: a) functional play - grabbing b) relational play – stacking.

Based on this classification, functional play primitives are defined with respect to a single play object, whereas relational play primitives are defined with respect to both a play object and a target object. These are key characteristics that allow a robotic system to distinguish between the two categories of play. We propose that the method needed to identify individual play primitives within each category involves first identifying low-level behaviors extracted during a human playing action.

A. Motion Behavior Analysis

A motion behavior is used to represent an interpretation of the basic movements of a human's play. It is not designed to compute specific motion vectors (such as specific arm joint trajectories), but rather to provide information about general movements of the toy object of interest. For example, if a human playmate lifts a toy up in the air (which would be defined as a functional play primitive), we would like the system to identify that a toy was grabbed and lifted. We are not concerned about its exact ending position in the z-plane. For a relational play instance (such as inserting), if a play action involves inserting a ring on a stack, we would like the system to identify that a toy was moved towards another toy object until contact was made. We are not concerned with the absolute position of the target object.

To enable this interpretive construct, we define a motion vector

$$\mathbf{M}_v = (d, v) \quad (1)$$

where d represents the direction of motion and v represents the velocity of motion. In addition, the possible values associated with d and v are discretized based on pre-defined linguistic classes, as depicted in Table I. As such, there is a finite number of motion vectors that exist for defining a low-level motion behavior. We define this finite set of possible motion vectors as the motion class κ_{motion}.

Table 1. Motion behavior definition structure

Motion Parameter	Play Primitive	Linguistic Values
Direction (d)	Functional	Left Right Up Down
	Relational	Positive: towards target Negative: away from target
Velocity (v)		Slow Fast

Computing Functional Direction

Functional direction represents the absolute direction of the play object with respect to a world coordinate system. The following direction vectors are used to classify this motion parameter:

$$LEFT = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, RIGHT = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, UP = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, DOWN = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Computing Relational Direction

Relational direction, compared to the functional direction, describes the motion of an object relative to the target. Distance between an object and a target is computed when a motion initiates, d_i, and terminates, d_t. These two distances are then compared and define the positive and negative relational direction.

- POSITIVE : object moving towards the target
$$d_t \leq d_i$$
- NEGATIVE: object moving away from the target
$$d_t > d_i$$

Computing Velocity

The velocity of the motion behavior, v, is measured as follows:

$$v = \frac{\Delta p}{\Delta t} \text{ (px/s)} \quad (5)$$

Δp is defined, with respect to an observation, as the distance between the location of an object when a motion initiates and terminates. Δt is measured by counting the frame numbers during a motion and dividing it by the average frame rate of the camera.

$$\Delta t = \frac{\#\ of\ frames}{fps}(\text{sec}) \quad (6)$$

Since the velocity required in this study need not be precise, it is reclassified as a speed: SLOW/FAST. If a motion is faster than the overall sequence speed average, it is defined as FAST, and as SLOW otherwise.

As an illustrative example, Table II shows the association between low-level motion behaviors and the resulting motion vectors.

Table II. Association between motion behaviors and vectors

Illustrative Description of Motion Behavior	Motion Vector
Human quickly lifts toy from table	(Up, Fast)
Human inserts toy into toy-bin	(Positive, Slow)
Human shakes toy to the right	(Right, Fast)

The goal of the motion behavior analysis process is to populate instances of the motion vector based on observation of a human playing action (such as depicted in Table II). This process is executed by computing a motion gradient during human play actions and fitting the motion gradient to the pre-defined motion class. Algorithmically, our motion behavior analysis process consists of the following steps:

1. Compute the motion gradient associated with observation of the human. Compute and normalize the motion gradients for N consecutive motion frames to minimize the effect of motion jitter (where Δx, Δy represents the change in pixel location of the imaged object).

$$G[i] = \begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix}, \ \sqrt{\Delta x_i^2 + \Delta y_i^2} = 1,\ 0 \le i \le N \quad (2)$$

2. Find the motion vector M_κ which minimizes,

$$M_\kappa = \begin{bmatrix} m_{\kappa,x} \\ m_{\kappa,y} \end{bmatrix} = Arg\min_M \sum_i \| M - G[i] \| \quad (3)$$

This minimization is equivalent to the least squares estimation of the motion model fit and a closed form solution to this problem can be expressed in terms of a linear matrix equation,

$$G^T M = \begin{bmatrix} \Delta x_0 & \Delta y_0 \\ \Delta x_1 & \Delta y_1 \\ \vdots & \vdots \\ \Delta x_n & \Delta y_n \end{bmatrix} \begin{bmatrix} m_{\kappa,x} \\ m_{\kappa,y} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (4)$$

3. Normalize and fit the motion vector M_κ to the pre-defined motion class to compute direction and velocity of the motion (as discussed below).

The best-fit motion vector (with populated values) defines the current instance of the motion behavior. As an example, test results of fitting a M_v=*(Right, Slow)* motion vector to an object moving in the right direction is shown in Figure 3, whereas Figure 4 shows a two motion sequence of an object moving right, then up. In these figures, N represents the number of frames used to calculate the motion gradient in Equation 2. The dots in the figures represent the position of the object following each N consecutive motion frames.

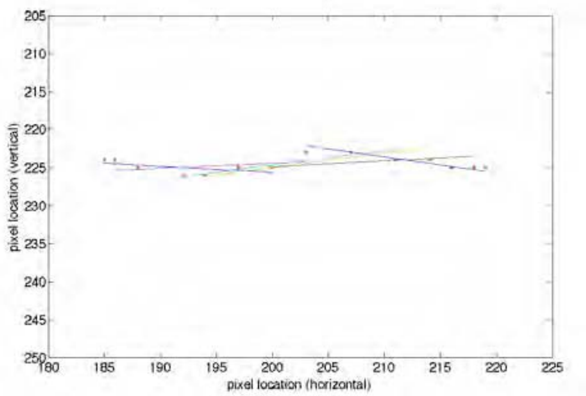


Figure 3. Motion model fitting. Above motion is labeled as κ_{RIGHT} (N=5)

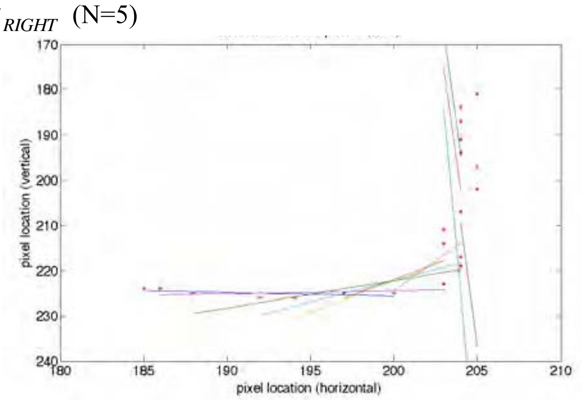


Figure 4. Motion Sequence. Above sequence is labeled as κ_{RIGHT} - κ_{UP} (N=5)

B. Behavior Sequencing

In order to extract play primitives for a robot playmate, behavior sequencing involves identifying, and labeling, the sequence of motion behaviors associated with a play scenario. Due to the limited nature of the play domain, we assume there is a finite number of common primitive gestures that need to be identified (e.g. grasping, transporting, inserting, hammering, stacking, pushing). Therefore, to perform this sequencing operation, the toy of interest is first identified and then tracked over subsequent motion frames. A set of individual motion behaviors is then determined and used to identify which of the corresponding play primitives is active.

1. Object Detection

Detecting toy objects in the scene consists of three primary steps - RGB to HSV color conversion, histogram back-projection, and segmentation. Since most children's toys use saturated colors to keep visual attention, we use color as the key feature for detecting an object. During a human playing action, a color input image is captured at 30 frames per second and converted into a one channel Hue image. This image is then back-projected with a pre-defined histogram to segment color. Each segmented group is then re-examined to eliminate outliers and unsymmetrical contours. Through this process, individual toy objects resident within the image can be identified. An example of this process is shown in Figure 5.

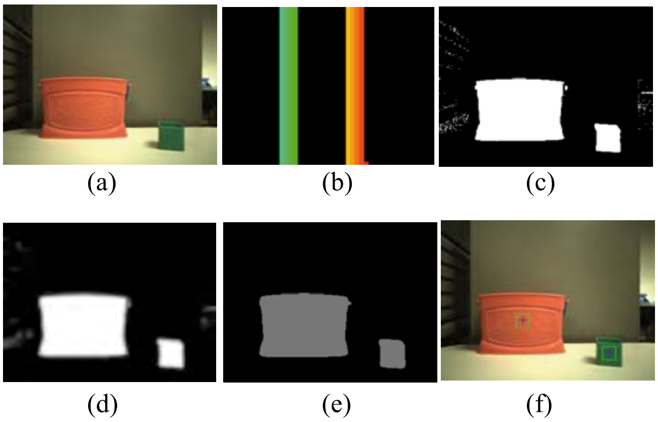


Figure 5. (a) Original Toy Scene Image (b) Back-projected Hue Histogram (c) Histogram Back-projected Image (d) Smoothed Image with Gaussian Filter (e) Binary Thresholded Image (f) Final Toy Objects Detected (center marked with green box)

2. Object Tracking

Among the multiple toys detected, we define the first one to take an action as the play object. The other toys are then marked as targets, and the motion of the reference toy is described relatively to them. Object tracking involves the repeated process of object detection, in which the back-projection histogram only references the color of the play object (Figure 6). This construct allows us keep track of the dominant play object, even when there might be different objects of the same color present in the play scene.

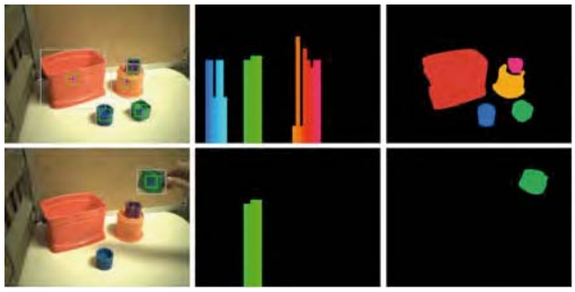


Figure 6. (Top) Identifying toy objects (Bottom) Tracking the identified play object

3. Play Primitive Extraction

Using the motion behavior analysis process, individual behaviors are identified and sequenced based on movement of the play object. The final resting destination of the play object is then used to identify the final play primitive. For testing results, we select two - 1) Insert: after a downward motion towards the target, the play object disappears, and 2) Stack: after a downward motion towards the target, the play object is placed on top of the target. In this same manner, other play primitives are pre-determined based on prior observation of the motion behaviors during human play dynamics.

Although this paper focuses on the methodology developed for extraction of the play primitives, the structure of the derived play primitive is designed for transmission to a robot platform for subsequent toy manipulation (Figure 7) [17]. In this scenario, the human has a toy that is identical to the robot's toy, so that the human and the robot can play side by side. The play primitives, once transmitted to the robot, are matched to the autonomous robot's behaviors in order to allow for subsequent side-by-side play actions.

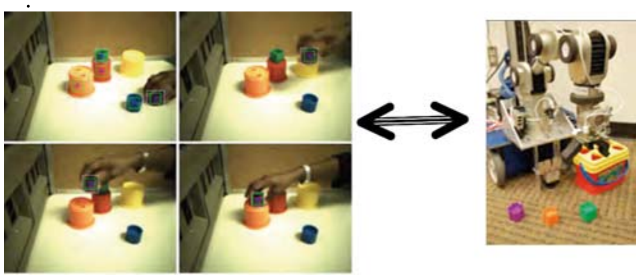


Figure 7. Robot platform used for subsequent toy manipulation.

III. Experimental Setup and Results

To test the capability of the system, we use a stereo camera system (Videre) to analyze the motion behavior. The supporting algorithms are run under the Fedora Linux operating system.

Four adult test subjects were asked to perform repeated play primitives (insert/stack) for five different play scenarios. The human was observed from a fixed camera with a side view. Multiple toy objects were randomly positioned in the play scene and the human was instructed to either 1) select any object and stack onto another or 2) select any object and insert into another. In three of the play scenarios, the human was instructed to perform these actions continuously (i.e. multiple times) and in two of the play scenarios the human was instructed to perform only one play action. The experiments were designed to test the capability of the system to identify both the motion behaviors as well as the correct play primitives, given the differences in motion behaviors for the different subjects. Figure 8 depicts a sequence associated with the inserting play primitive. Figure 9 shows snapshots of other play scenarios.

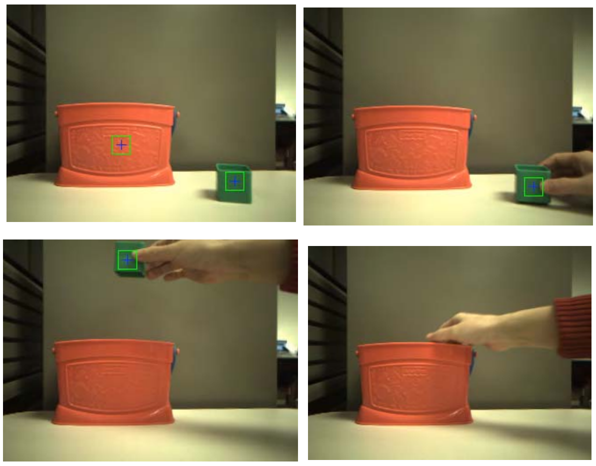


Figure 8. Inserting Sequence

Figure 9. Various Play Scenarios

Test results were documented using the construct shown in Table III, which represents a specific instance of an insertion play scenario. The testing input documents the actual color of the play object, the target object, and the velocity of operation. Although for the insertion and stacking operations, the speed of the motion is not significant, it helps to distinguish between other similar primitives (such as hammering versus repeated pushing). The testing output documents the results of applying our methodology to the play scenario. Each column represents the best-fit motion vector determined by the motion behavior analysis process and the sequence of motion vectors used to identify the play primitive.

Table III. Methodology Performance Results

Test Input for Insert Sequence			
Target Color	RED		
Object Color	GREEN		
Operation Type	INSERT		
Object Velocity	(0,42) px/s	(-64,0) px/s	(0,22) px/s

Test Output for Insert Sequence			
Absolute	UP	RIGHT	DOWN
Relative	NEG	POS	POS
Speed	SLOW	FAST	SLOW
Primitive	INSERT		

Given the various test scenarios, the performance of the system was categorized based on correct recognition of play and target objects, identification of the motion behaviors, and correct labeling of the play primitive. Table IV documents these results. The play/target object recognition rates are associated with the ability to correctly identify both the play object (upon grabbing a toy object) and the target object (upon releasing the play object). Motion behavior recognition results are associated with correctly determining the best-fit motion vector associated with human movement. Errors in this calculation were primarily due to miscalculation of speed (slow versus fast). The play primitive recognition rate was associated with correctly identifying the play primitive (insert/stack).

Table IV. Methodology Performance Results

Overall Recognition Rates	
Play Primitive	100.00%
Motion Behavior	98.17%
Play/Target Object	100.00%

IV. Conclusions and Future Work

As robots continue to hold a greater role in educational and therapeutic aspects with respect to children, the necessity for endowing them with interactive play capability poses interesting challenges. Through this research, we seek to show a methodology that will provide the robot the ability to identify general behaviors for playing with toys. Experiment results show that the methodology enables the robot to interpret the playing behavior of a human, with increased variety of toys. We believe that, although there will be differences between an adult versus child playmate (e.g. addition of 'uncorrelated' motion primitives, noise in low-level motion gradients, repeatability in play), the methodology proposed provides an infrastructure for providing instruction to a robotic playmate. Future work will be focused on expanded motions such as hammering, pushing, and aligning, as well as interacting with the human by taking turns within the same play scenario.

REFERENCES

[1] J. Piaget. Play, Dreams and Imitation in Childhood, London: Routledge and Kegan Paul Ltd., 1951.

[2] K. Dautenhahn and A. Billard, "Games children with autism can play with robota, a humanoid robotic doll," Proceedings of Cambridge Workshop on Universal Access and Assistive Technology, pages 179–190, 2002.

[3] B. Scassellati, "How social robots will help us to diagnose, treat, and understand autism," 12th International Symposium of Robotics Research, 2005.

[4] G. Kronreif, B. Prazak, S. Mina, M. Kornfeld, M. Meindl, and F. Furst, "Playrob - robot-assisted playing for children with severe physical disabilities," Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics, 2005.

[5] F. Michaud, "Assistive technologies and child-robot interaction," Proceedings of American Association for Artificial Intelligence Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, 2007.

[6] F. Tanaka, A. Cicourel, and Javier R. Movellan, "Socialization between Toddlers and Robots at an Early Childhood Education Center," Proceedings of the National Academy of Sciences of the USA (PNAS), Vol.104 No.46 p.17954-17958, November 2007.

[7] K. R. Ginsburg, "The Importance of Play in Promoting Healthy Child Development and Maintaining Strong Parent-Child Bonds," Pediatrics, Vol. 119, No. 1, January 2007.

[8] T. H. Escobedo, "Play at the Art Table: A Study of Children's Play Behaviors while Drawing," Annual Conference of the Association for the Study of Play, Austin, TX, April 25-27, 1996.

[9] S.B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE Transactions on Robotics and Automation*, Vol. 13, 1997.

[10] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning By Watching: Extracting Reusable Task Knowledge From Visual Observation Of Human Performance," *IEEE Transactions on Robotics and Automation*, Vol. 10, pgs. 799-822, 1994.

[11] S. Calinon, F. Guenter, F. and A. Billard, "On learning the statistical representation of a task and generalizing it to various contexts," *IEEE International Conference on Robotics and Automation* (ICRA), Orlando, FL., 2006.

[12] N. Pollard and J. Hodgins, "Generalizing Demonstrated Manipulation Tasks," *Workshop on the Algorithmic Foundations of Robotics* (WAFR '02), 2002.

[13] C.G. Atkeson and S. Schaal, "Robot Learning From Demonstration," *Proc. 14th International Conference on Machine Learning*, pgs. 12-20, 1997.

[14] Gale Encyclopedia of Psychology, "Fine Motor Skills," Oct 2005.

[15] P. Zelazo and R. Kearsley, " The emergence of functional play in infants: evidence for a major cognitive transition," *Journal of Applied Developmental Psychology*, 1, pgs. 95–117, 1980.

[16] A. Bigelow, K. MacLean, and J. Proctor, "The role of joint attention in the development of infants' play with objects," *Developmental Science*, 7(5), pgs. 518-526, 2004.

[17] Robotic Playmates Project Website, http://www.hsi.gatech.edu/hrl/project_toys.shtml