Object Feature Extraction and Tool Using for Robots with Recurrent Neural Network Model

Shun Nishide

I. INTRODUCTION

Construction of an efficient perception and action mechanism is the most crucial issue for creation of intelligent robots. Our objective is to apply the robot's own experience for developing the robot's perception and action mechanism based on human development. In this talk, I present my work on development of robot's perception mechanism based on affordance theory and motion generation using tools.

Concerning perception, affordance theory suggests that humans' cognition is based on perception of invariant features of the environment [1] [2]. Two types of invariant features exist: structural invariants and transformational invariants. Structural invariants are features that remain constant through motor interaction and transformational invariants are features that describe the change of the environment. The first half of the presentation is on using the robot's pushing motion to tabletop objects to extract features that represent structural invariants and transformational invariants.

The second half of the presentation is on motion generation with tools. Tool use is one of the largest distinctions of primates from other beings. The ability to use tools as if they are part of the body is called tool-body assimilation [3]. Tool-body assimilation experiments with a humanoid robot generating goal-oriented motions with and without a tool are presented.

II. MODEL COMPOSITION

Recurrent Neural Network with Parametric Bias (RNNPB) and Multiple Timescales Recurrent Neural Network (MTRNN) are used for the learning model of the system. A detailed explanation of the models are given in [4] and [5]. The two RNNs are both capable of learning multiple sequences in a single model. Trained sequences are compressed into parameters which are self-organized into a classification space based on the similarities of the sequence.

For training data acquisition, the robot performs pushing motions to target objects. During the motions, the robot acquires the motor angle sequence and image sequence from the camera. In the experiment, the background of the images are subtracted for extraction of object features.

Features representing transformational invariants are extracted during the training process of RNN, while those representing structural invariants are extracted after training of RNN. Features representing transformational invariants are extracted by a mutual optimization method of a feature extractor and RNN. Experiments are presented using both RNNPB and MTRNN. A hierarchical neural network is used as a feature extractor for RNNPB, while a feature extracting component is embedded within the RNN for MTRNN. Features representing structural invariants are extracted by applying the self-organized parameters of the sequences, obtained through training of RNN. The parameters are self-organized in a manner that similar sequences (motions) have a close value. The parameters are used as outputs of a hierarchical neural network which inputs static images of objects. The hierarchical neural network self-organizes prominent features (structural invariants) that link object shapes to the resulting motions in the middle layer. A diagram of the feature extraction model is shown in Fig. 1.

Tool-body assimilation is conducted using an RNN as the robot's body model, modifying it based on the tool the robot is holding. The model is trained by first training RNN using motion sequences acquired with the barehanded robot, and then training the modifying network with motion sequences acquired with the robot holding a tool. Identification of tool property is done by two methods.

1) Waving motion of the tool.
2) Estimation from tool shape.

The two methods are related to the two stages of human infant development where an infant first relies on dynamic touch to recognize the grasping tool, and tends to switch to identify the tool by the visual shape [6]. For the experiment with waving motion, the robot moves its arm to observe the environmental change for identifying the grasped tool.

S. Nishide is with the Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan nishide@kuis.kyoto-u.ac.jp
The second method estimating the tool property by tool shape is done using the visual image of the tool. After identification/estimation of the grasped tool property, the weights of the RNN (robot body model) are modified. A diagram of the tool-body assimilation model is shown in Fig. 2.

### III. Experiments

Experiments for feature extraction were conducted with tabletop objects, shown in Fig. 3. The robot pushed each object, obtaining the image sequence of roll, slide, fall over, and bounce, for training the model. A front image of the object was used for the input of feature extraction representing structural invariants. Training and analyzing the model has shown that features representing tallness (horizontal to vertical ratio), roundness, and sharpness of corners were extracted as structural invariants. For transformational invariants, feature sequences with peculiar characteristics for the four motions were extracted.

Experiments for tool-body assimilation was conducted with three tools: T-shaped, I-shaped, and L-shaped. The L-shaped tool was used as an unknown tool in the experiment and was not used for training. First the robot moved its arm (without tool) on the table with two target objects on it. The image sequence and robot arm motion obtained during the motions were trained using the RNN. Next, T-shaped and I-shaped tools were attached to the robot arm. The same motions were done with the target objects to obtain motion sequences. The motion sequences were used to calculate the relationship between the tool and modification of RNN. After training, a goal image was shown to the robot to achieve the goal state with the grasped tool. Fig. 4 shows an example of the motion generation experiment where the robot was required to pull the target objects with an untrained L-shaped tool.

### IV. Conclusion

As an approach for creating an intelligent robot based on its own experience, an RNN based method is introduced in the talk. For perception mechanism, we focused on structural and transformational invariants from affordance theory. For motion generation, tool-body assimilation model is introduced as an example. Experiments using these models with remaining issues for future work are also presented.

### References


